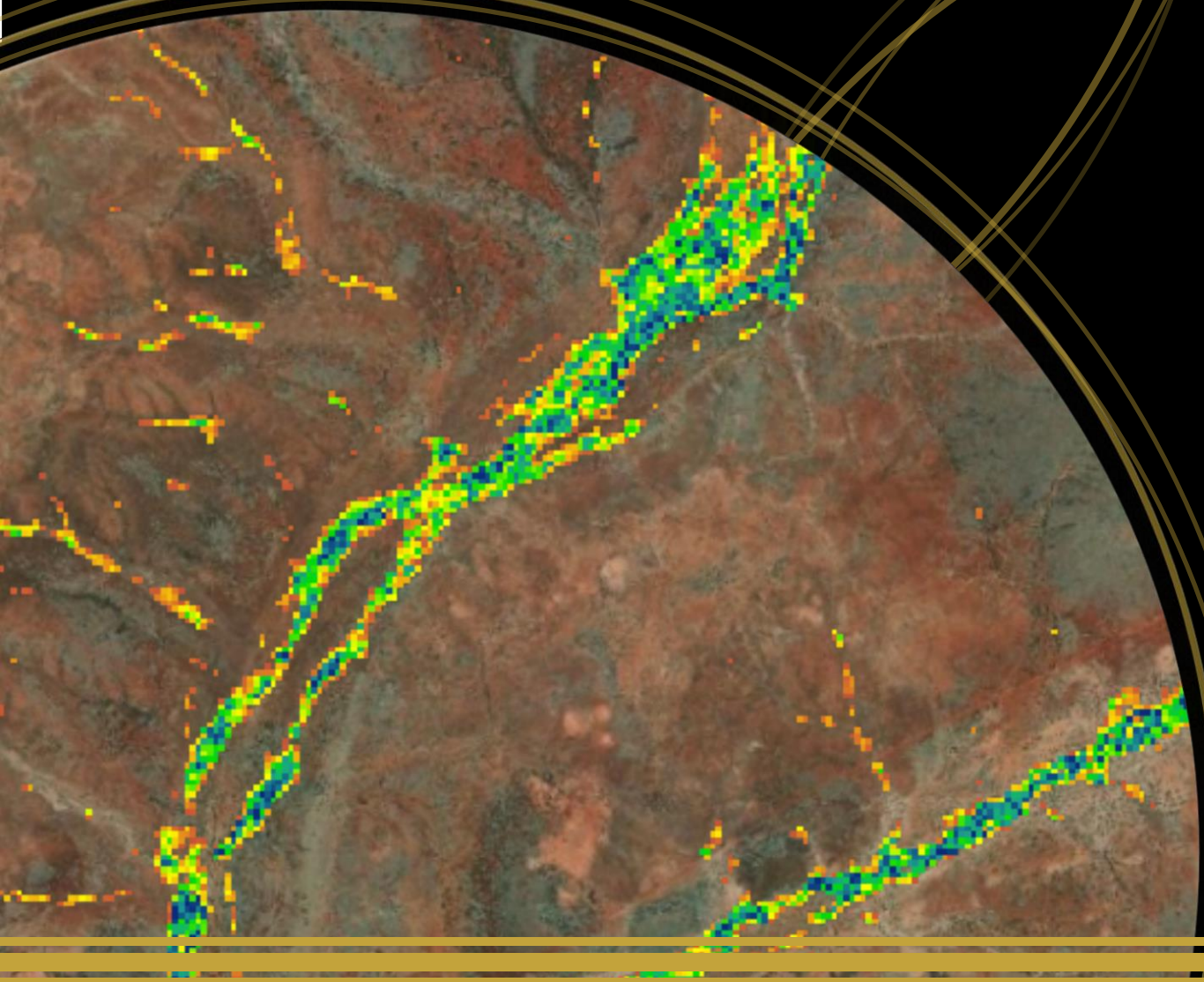
Modelling and Monitoring Groundwater Dependent Vegetation with Open Data Cube Imagery



TECHNICAL REPORT

Technical Report

Modelling and Monitoring Groundwater Dependent Vegetation with Open Data Cube Imagery

L Trotter, T.P. Robinson, A Cross

May 2020

Contents

Acknowledgements iii
Executive Summary iv
1 Introduction 1
1.1 Aims and Objectives 1
2 Materials 2
2.1 Weather and Climate 2
2.2 Study Areas 2
2.2.1 Study area 1: BHP Yandi Mine Tenement 3
2.2.2 Study area 2: Roy Hill Tenement 4
2.2.3 Study area 3: Ophthalmia Dam 4
2.3 Field Survey 4
2.3.1 Survey timing and personnel 4
2.3.2 Data collection 4
2.4 Datasets 7
2.4.1 Earth observation imagery 7
2.5 Development and software 8
3 Methods 9
3.1 GDV-likelihood Modelling 9
3.1.1 Spatial multi-criteria modelling 9
3.1.2 Spectral indices 9
3.1.3 Index seasonal medians 12
3.1.4 Invariant target standardisation 12
3.1.5 Perennial vegetation and moisture stability 13
3.1.6 Analytical hierarchy process: pairwise comparison 13
3.1.7 Accuracy assessment 14
3.1.8 GDV-likelihood dichotomisation 15
3.1.9 Bushfire detection and mapping 15
3.2 Time-series Data Analysis 15
3.2.1 Trend modelling 15
3.2.2 Change detection 16
3.2.3 Vegetative health regime and break detection 17
4 Results 18
4.1 GDV-likelihood Modelling 18
4.1.1 Accuracy assessment 22
4.1.2 GDV-likelihood dichotomisation 24
4.2 Time-series Data Analysis 27

4.2.1 Trend Modelling 27
4.2.2 Change Detection 33
5 Conclusions 36
6 References 37
Appendix A: Canopy condition ranking guide 42
Appendix B: GDV Tool workshop material 43

Acknowledgements

We would like to thank Darren Mottolini, formerly of FrontierSI and now with GHD, for bringing the proponents together to discuss the potential of earth observation imagery for groundwater dependent vegetation detection, modelling, and monitoring. We would especially like to thank Paula Fievez from FrontierSI for taking over from Darren at the beginning of the project, keeping everyone in the room talking, ensuring we met milestones, still keeping everyone in the room talking and for generally forming the nexuses to build something worthwhile. Also, from FrontierSI, we would also like to acknowledge Phil Delaney, Caitlin Adams, Jess Keysers and many others who played a very necessary role over the 12-months of the project.

This project would not have been possible without the financial support of many organisations, all of whom we thank deeply. Firstly, Chris Penning and his team from Geoscience Australia provided substantial financial support as well as access to the datacube, which was the linchpin of the whole operation. David Mickle and his associates from BHP and Harriet Davie from Roy Hill and her associates were also strong financial supporters of the project and we appreciate their time devoted to the project and for providing all the logistic support for our field validation work.

Additional in-kind support and project governance was provided by Michelle Antao from the Department of Water and Environmental Regulation, Lesley Gibson, and Denise True from the Western Australian Biodiversity Science Institute (WABSI). Additionally, we thank Tim Bleby from Astron for showing us around the tenements and providing good humour in the process.

Executive Summary

Monitoring of groundwater dependent vegetation (GDV) around mines undergoing dewatering is required to ensure water requirements from the phreatic zone are being met. Ground-based surveys of vegetation health are important in this process but have limitations that can be overcome with the assistance of earth observation imagery. This includes better targeting of potential GDV to monitor, rapid observations across entire tenements, regular repeat and passive monitoring and a large back catalogue to assist benchmarking prior to disturbance. Here, we leverage the Digital Earth Australia (DEA) Open Data Cube (ODC), which provides calibrated, analysis ready Landsat and Sentinel imagery, across Australia, through time. We test our approach on three study areas in the southeast Pilbara bioregion of Western Australia.

Initially, we develop GDV-likelihood models using a knowledge-driven approach, which requires no upfront field validation. Models are based on annual stability, greenness, and moisture, with a disproportionate amount of weight given to dry season parameters. These models reduced the search space for GDV and provided the guidance for field crews to collect validation data using tablet technology. In total, 879 data points were collected over two field campaigns during the dry season of 2019.

Using the validation data, models were permuted in pursuit of the best vegetation index. We identified the moisture adjusted vegetation index (MAVI) to outperform more commonly used indices including NDVI and SAVI irrespective of platform (Landsat or Sentinel) or study area. Therefore, **MAVI is presently the recommended index for future models** and time series analysis. **We do not recommend the use of MSVI-1/2/3** due to comparatively poor performance. The TCT Greenness index was consistently in the top three indices irrespective of platform or study area. It is therefore appropriate to **use TCT Greenness and TCT Brightness for change vector analysis**.

Model accuracy was assessed using the area under the curve (AUC) statistic, where 1 represents a perfect model. We observed an increase in AUC from 0.85 to 0.95 using MAVI when switching from 30 m to 10 m resolution imagery. Similar magnitudes occurred for all spectral indices, so we suspect **accuracy improvements were a function of spatial resolution.** Sentinel would therefore appear to be the optimal analysis ready dataset for GDV-likelihood modelling. However, we note that Sentinel lacks an archive for finding GDV that may already be in poor health, testing the significance of trend analyses and is less useful for determining benchmarks that pre-date mine disturbance. Landsat has continuity assurance and a *c.* 35-year library. Sentinel has no such assurances and has an expected mission life of seven years. Consequently, we have designed the system to accept either as input.

Trend analysis was able to detect declines in areas showing *M. argentea* and *E. camaldulensis*/*victrix* mortality. We also identified health and high recruitment in an area receiving water discharge. These results were suitable for historical reconstruction and are useful to explore the long-term trend. More testing on less binary cases is recommended, but as the trend needs to be monotonic for successive periods, **it's use for early detection of health decline is not currently recommended**. Rather, we advocate change vector analysis (CVA) for this purpose.

CVA is designed to augment the trend analysis. It requires the creation of a benchmark from any point (series of points or the entire time series) in the archive. This may be the median of years (or seasons) just prior to dewatering. **CVA allows comparison of an "ideal" to present day imagery** to identify a magnitude of change that is interpretable relative to the angle of the vector produced.

1 Introduction

Mine dewatering is the removal of unwanted groundwater from open pit mines to allow mineral extraction. In some cases, this can contribute to water table decline, which can cause water stress and mortality of plants that source most of their water requirements from the phreatic zone. Such plants are commonly labelled groundwater-dependent vegetation (GDV) or phreatophytes (Thomas, 2014). As a result of this risk, monitoring of GDV is essential on mining tenements in Australia where dewatering occurs for vegetation that has been committed to be retained and for vegetation located outside the mining tenement, and is commonly accomplished with ground-based survey assessments relative to a benchmark.

Ground-based surveys are an important part of the monitoring process but, in isolation, have limitations. Firstly, they are not necessarily targeted to visit all the GDV locations because the GDV locations are not known. Secondly, they are unlikely to sample across entire populations, especially where access is limited. Thirdly, their revisit rate may only be completed annually and can be too slow to alert and inform intervention strategies. Finally, benchmarks defining expected health are largely qualitative and more difficult to define in ecosystems that have been repeatedly disturbed over a long period of time.

Earth observation data is a proposed augmentation to better direct ground-based surveys. It has the potential to rapidly map GDV based on the premise of perennial green and moist foliage (O'Grady et al., 2011; Barron et al., 2014). Maps of GDV reduce the search space and promote more efficient surveys. Synoptic maps can easily be produced over the full extent of tenements including in hard to access locations. Sentinel and Landsat is regularly acquired (*c*. every 5 to 16 days, respectively) and thus capable of providing regular updates and early warning. Finally, imagery archives often date back to before potential disturbance activities began allowing for benchmarks to be derived retrospectively from relatively intact ecosystems.

Recently, there has been a new paradigm in the ability to access free and open satellite data known as open data cubes (ODCs). ODCs are essentially temporal stacks of analysis ready data (ARD). Digital Earth Australia is the Australian government's implementation of the ODC initiative and is leveraged in this research to process the otherwise prohibitively large spatial and temporal volumes of Landsat and Sentinel data required to inform decision makers on the location of GDV and its health trajectory. We note that, to our knowledge, no other study presently exists making use of the DEA ODC for the purposes of GDV detection or monitoring.

1.1 Aims and Objectives

This project aims to develop validated remote sensing-based methods for detecting GDV and its change over time. It has the following objectives:

1. Develop GDV likelihood models using spatial multicriteria evaluation (SMCE).
2. Optimise model input using validation statistics obtained from ground-truthing data.
3. Develop trend and change models and demonstrate their applicability over three study areas with different histories.

2 Materials

2.1 Weather and Climate

The Pilbara region has two distinct seasons: a hot wet summer (October to April) and a mild dry winter (May to September). Mean annual rainfall for the nearest official weather station, Newman Aero, is 324.3 mm (Figure 1; BoM 2020). Rainfall is highly variable, and a considerable amount of the annual total is derived from thunderstorm and cyclonic events occurring during the wet season – December through March (Van Vreeswyk 2004). Mean maximum temperatures at Newman Aero are above 30ºC for most of the year and peak at 39.1ºC during December (BoM 2020). These high temperatures result in mean and annual evaporation (> 3500 mm) that exceed mean annual rainfall totals, with deficits particularly evident in the dry season – September through November (Van Vreeswyk 2004).

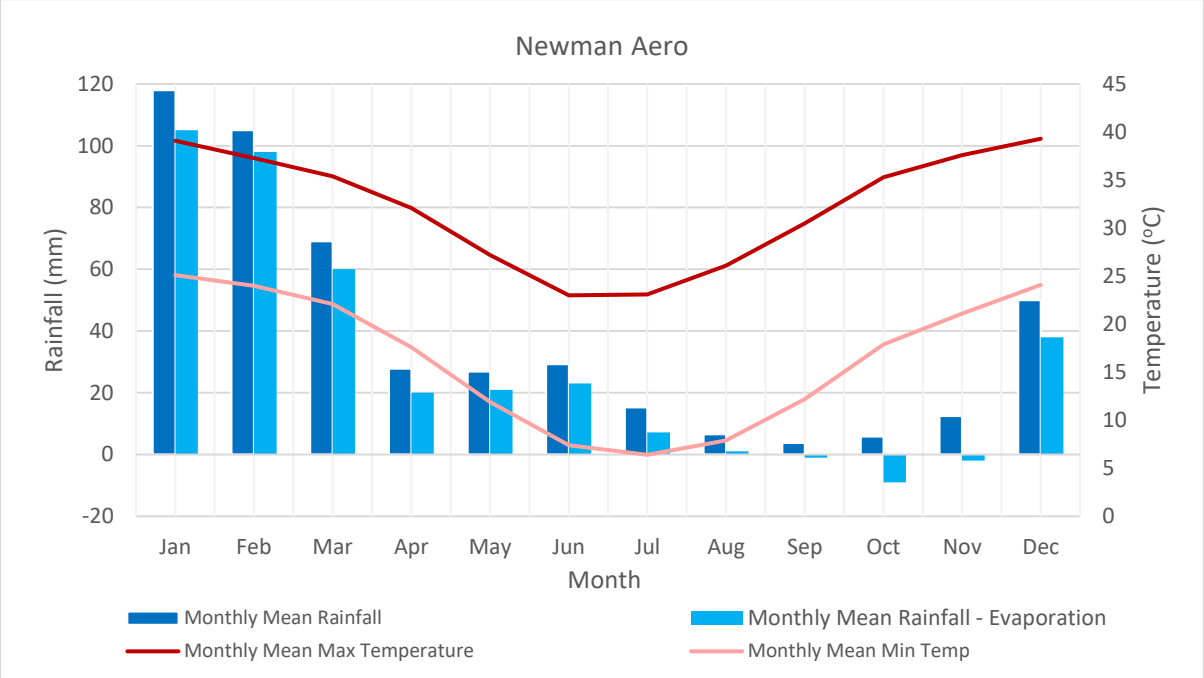


Figure 1: Climate data of Newman Aero (23.36ºS 119.73ºE).

2.2 Study Areas

Three study areas situated in the southeast Pilbara bioregion of Western Australia (WA) were selected as case studies (Figure 2). Yandi has been an operational iron ore mine for nearly three decades and presents as a suitable test case for long term dewatering activities. Roy Hill only began dewatering in August 2015 and the majority of GDV vegetation is located at some distance from current mine pits. Ophthalmia Dam is of anthropogenic origin and is considered to have caused reduced water flow north of its dam wall (Payne and Mitchell 1999) and a restriction of flood water onto adjacent plains from episodic weather events.

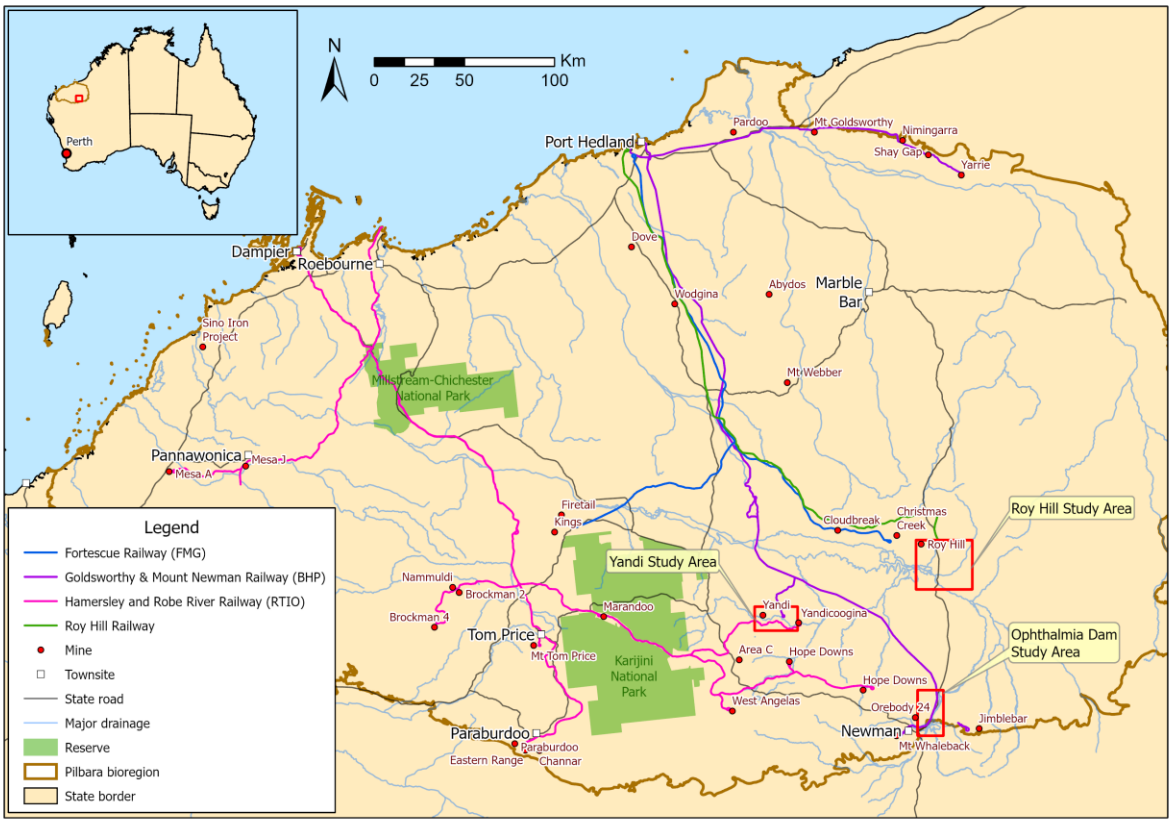


Figure 2: Regional location of the three study areas: Yandi, Roy Hill, and Ophthalmia Dam. All study areas are situated in the south-east Pilbara bioregion.

2.2.1 Study area 1: BHP Yandi Mine Tenement

BHP operates the Yandi iron ore mine located on the BHP Yandi Mine Tenement, approximately 90 km northwest of Newman in the Pilbara bioregion (Thackway and Cresswell, 1995) of Western Australia (Figure 2). Mine operations began in 1991 and mine pit dewatering and water displacement is currently in operation. Much of the mining footprint is situated adjacent to Marillana Creek (Astron 2011).

Vegetation of the area and surrounds is broadly described as spinifex steppes with a sparse Eucalyptus overstorey on undulating hills with mesas. Major creeks and riverbeds are dominated by riparian vegetation including *Eucalyptus camaldulensis* Dehnh (river red gum), *E. victrix* L.A.S. Johnson and K.D. Hill (western coollibah) and *Melaleuca argentea* W. Fitzg. (silver-leaved paperbark). All three species rely on groundwater sources for water uptake to varying degrees and as such can be adversely affected by groundwater drawdown (Astron 2011).

Decline and mortality of GDV species has been observed at isolated areas along Marillana Creek since 2000 (Astron 2011). Prior to 2000, historical reports indicate tree health along the creek was in good condition (Alan Tingay and Associates 1997). Historical survey efforts have noted gradual *M. argentea* decline or death at sites along the creek over the period 2000 to 2007. While it appears that a combination of dewatering and discharge cessation may have contributed to decline (BHP 2007), localised groundwater monitoring bores had not yet been installed, making statistical validation difficult.

Riparian tree death and decline causation has not yet been definitively explained due to a lack of widespread and historical groundwater monitoring bore data, making investigation of correlation between vegetation health and groundwater drawdown difficult. Additionally, severe bushfires and drought periods in the area have contributed to riparian tree mortality along the creek.

2.2.2 Study area 2: Roy Hill Tenement

The Roy Hill mine is located on the Roy Hill tenement approximately 120 km north of Newman in the Pilbara Region of Western Australia (Figure 2). The mine is relatively new, with construction commencing in 2011, and operations in 2015. The tenement is situated in the foothills and foot slopes of the Chichester Range and on the plains of the Fortescue Valley (Ecologia 2009).

Vegetation of the Roy Hill area is broadly described as varying widely between sparsely treed spinifex grassland steppes, mulga (*Acacia aneura*) open woodlands and tall shrublands, isolated *Acacia* sp. shrublands and individuals over rocky outcrops, and riparian vegetation including *E. camaldulensis* and *E. victrix*. These species are typically observed co-occurring with mulga, *Corymbia candida*, *Acacia coriaceae*, *Atalaya hemiglauca* and *Melcaulca glomerata* near creeks, riverbeds or floodplains (Ecologia 2009).

Both *E. camaldulensis* and *E. victrix* are dependent on groundwater within the tenement area and potentially at risk of being adversely affected by drawdown and bore-field abstraction (EPA 2019). Mine pit dewatering began in 2015. Phreatophytes are not found near mine pits and as such, GDV health decline has not yet been observed (Harriet Davie, pers. comm.). Vegetation health ground-based monitoring commenced prior to planned dewatering operations and is currently on-going (EPA 2019).

2.2.3 Study area 3: Ophthalmia Dam

The Ophthalmia Dam is located approximately 19 km east of Newman in the Pilbara bioregion of Western Australia (Figure 2). The dam was constructed in 1981 and intersects the Fortescue River just upstream of Ethel Gorge. The study area is focused on the riparian vegetation immediately surrounding the northern and southern areas of the dam wall and includes a small assessment corridor of vegetation lining the Fortescue River heading approximately 15 km north from the dam and parallel to the east of Marble bar Road.

The construction of the dam has been reported to be a major cause of severe, widespread tree and perennial grass decline and death. This has been attributed mostly to the effects the dam has had on reducing flood volume and frequency along the Fortescue River and floodplains. Payne and Mitchell (1999) report that widespread riparian vegetation decline and death occurred on parts of the Fortescue and Coolibah land systems commencing *c.* 1987, as natural flooding downstream reduced water availability to phreatophytes (mainly *E. victrix*).

2.3 Field Survey

2.3.1 Survey timing and personnel

Field surveys were carried out by Dr Todd Robinson, Lewis Trotter, Dr Adam Cross (Curtin University) and Dr Tim Bleby (Astron Environmental Services) over two consecutive visits in 2019. The Roy Hill study area survey was conducted between the 3rd and 5th of June, 2019, followed by the BHP Yandi and Ophthalmia Dam study areas between 26th and 28th of August, 2019.

2.3.2 Data collection

Data for six variables (Table 1) were recorded at the individual plant level where possible and associated with a GPS coordinate (positioning accuracy ±5 m) captured as close to the centre

of the tree canopy as possible. The “Phreatophyte” variable described whether the observed species was likely to be groundwater dependent; a value of “Yes” indicated the individual observation was either *E. camaldulensis*, *E. victrix* or *M. argentea,* greater than eight metres tall and not considered to be a sapling. Alternatively, a value of “No” indicated the observed species was unlikely groundwater dependent. This data was further elaborated using the “Species Name”, “Height” and “Photo” variables.

Table 1: Variables recorded for each observation made during field survey.

Variable	Measure	Description
Phreatophyte	Presence (Yes/No)	Is the tree a phreatophyte?
Species Name	Qualitative (description)	Name of species
Height	Quantitative (1 – 50 m)	Estimation of tree height
Overstorey Condition	Rank (1 – 5)	Overstorey condition based on estimated percentage of canopy cover remaining
Site Description	Qualitative (description)	General description of tree and surrounding species/landscape
Photo	Qualitative (photograph)	Photograph(s) of individual tree

Tree condition was recorded using a ranking system based on the estimated percentage of live foliage cover observed within the overstorey. This approach was qualitative, and field surveyors adhered to a visual guide (see Appendix A) to determine rank in the field constantly. Further guidance was provided by an experienced observer of tree condition with long-term familiarity with the vegetation of the study area when required. The ranking system utilised during surveys is provided in Table 2.

Table 2: Condition ranking system used during field surveys to represent percentage of live foliage cover in the overstorey.

Rank Value	Percentage of Live Canopy Foliage
1	<= 5%
2	6% - 35%
3	36% - 65%
4	66% - 95%
5	>= 96%

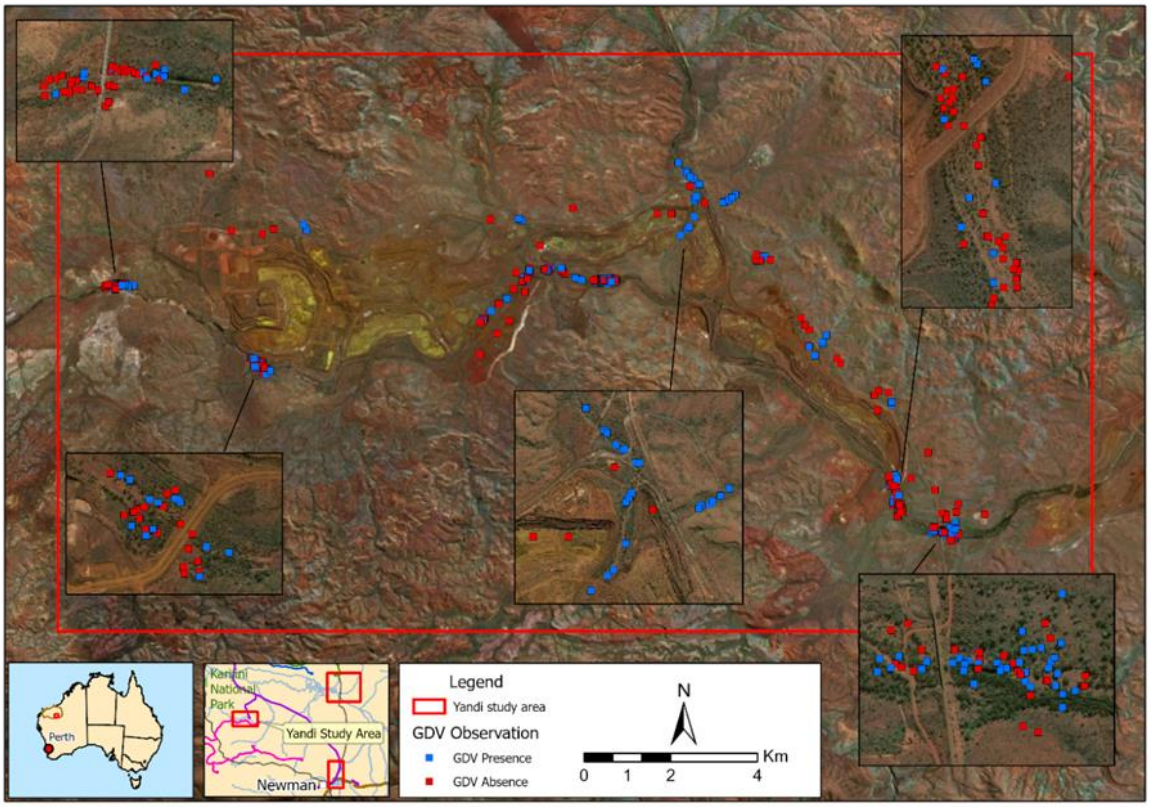


Figure 3: Survey effort at the BHP Yandi study area.

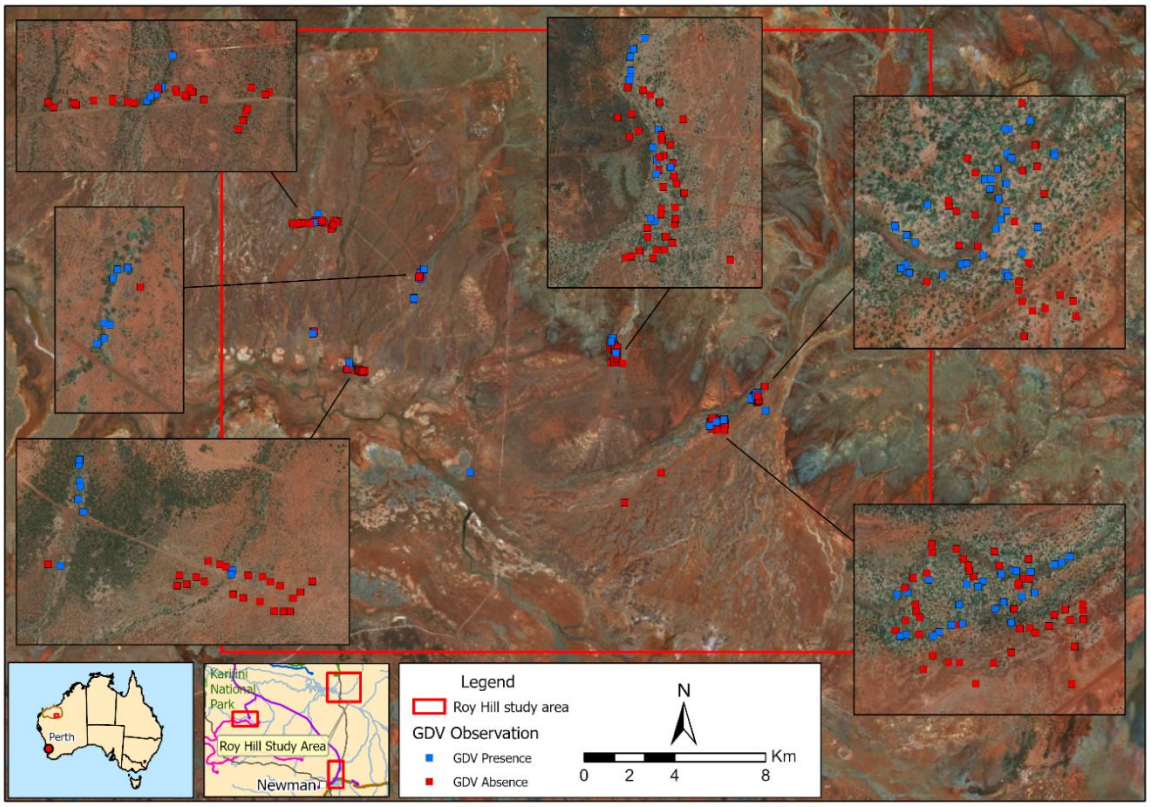


Figure 4: Survey effort at the Roy Hill study area.

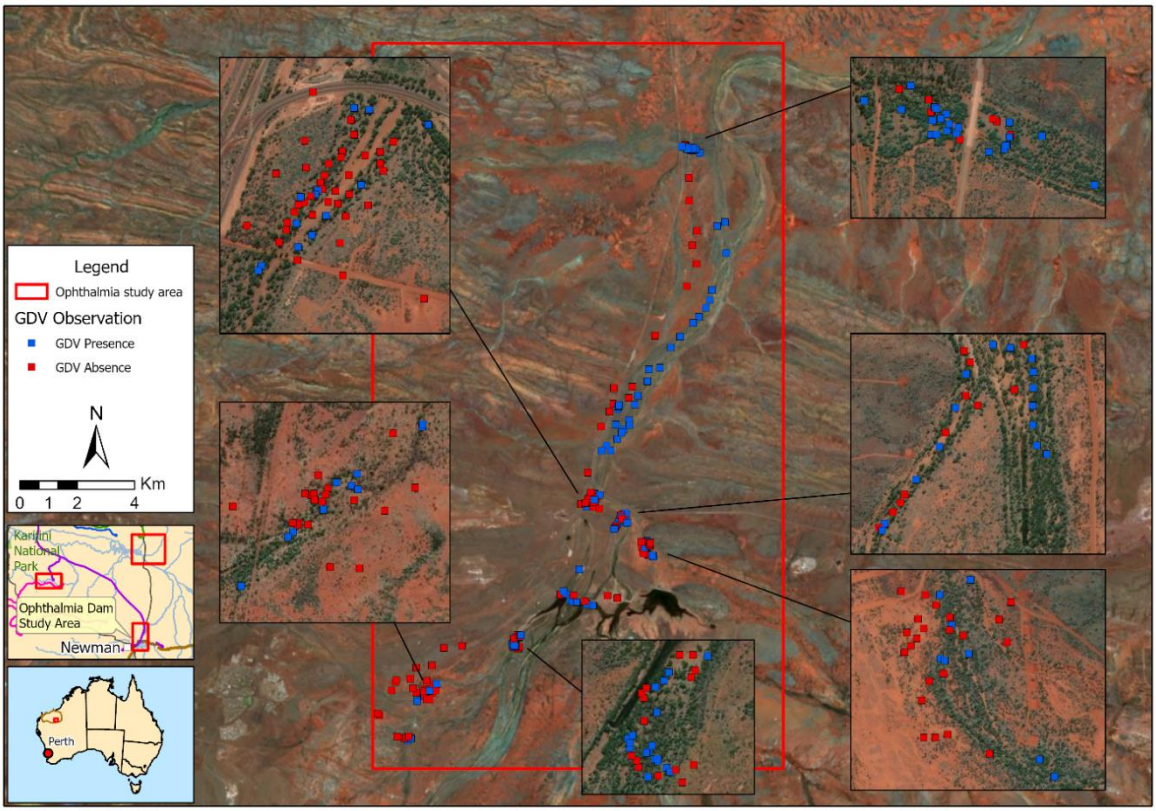


Figure 5: Survey effort at the Ophthalmia Dam study area.

2.4 Datasets

2.4.1 Earth observation imagery

A large database of a temporal series of earth observation imagery, as opposed to a snapshot in time, is required to adequately detect and monitor the trajectory of vegetation health over time (Wallace et al. 2004; Barron et al. 2014). This is due to the need for enough replicates to generate a baseline and confidence interval to enable interpretation of a significant transition or change that is not impeded by noise or short-lived aberrations.

Landsat and Sentinel-2 satellite platforms have the most desirable characteristics for this study, for varying reasons. Landsat provides specific benefits including: 1) spatial resolution of 30 m across all satellites; 2) an extensive historical image archive (Landsat 5 was launched in 1984); 3) a guaranteed commitment to continuity (Landsat 9 is scheduled for 2020); 4) a panchromatic band of 15 m allowing pan-sharpening (Landsat 7/8 and forthcoming Landsat 9); and 4) a 16-day revisit rate to the same location on earth. Sentinel-2A was launched in June 2015 and its twin, Sentinel-2B, in March 2017 and together offer several notable improvements to Landsat's characteristics, including higher resolution imagery (10 m) and 5-day temporal revisit frequency, but lack an extensive historical imagery archive. As each Sentinel mission is estimated at seven years, and there is no assurance of continuity, we have chosen to implement both data streams but, as a contingency, we have them running in parallel rather than in combination.

Digital Earth Australia (DEA) provided earth observation data for the study. The DEA platform catalogues voluminous amounts of satellite imagery covering continental Australia and is underpinned by the Open Data Cube (ODC), an open source software package that provides a Python based API for high performance querying, data access and analysis. Landsat 5 TM,

Landsat 7 ETM+, Landsat 8 OLI and Sentinel 2A and 2B satellite imagery was obtained pre-corrected to the Analysis Ready Data (ARD) standard (Giuliana et al., 2017), which includes geometric correction and bottom-of-atmosphere surface reflectance corrections using the NBAR-T (Nadir-corrected BRDF Adjusted Reflectance with Terrain Illumination correction, where BRDF stands for Bidirectional reflectance distribution function) method (Roy et al., 2016). Archived Landsat and Sentinel data up to mid-2019, each with six spectral bands (omission of the thermal band) covering the entirety of the three study areas, was made available for this study.

2.5 Development and software

Development of the GDV Tool and its algorithms, analyses and interface made extensive use of the Python programming language (www.python.org). The GDV Tool functionality was implemented and visualised using Jupyter Notebook (www.jupyter.org), a free and open-source web application designed to integrate Python development, user-interactivity, visualisation and code sharing. Access to the DEA ODC was achieved using the DEA Sandbox, a pre-built virtual environment that provides easy access to the DEA ODC, required Python packages, and the DEA Landsat and Sentinel ARD products (see Section 2.4.1). Additional Python packages utilised in this study includes pyMannKendall (Hussain and Mahmud 2019) and Ruptures (Truong et al., 2019) for Mann-Kendall trend analysis and change point detection functions, respectively.

3 Methods

3.1 GDV-likelihood Modelling

Previous research focused on the remote detection of GDV suggests that the characteristics of perennially high greenness and leaf moisture may assist its detection from earth observation imagery (O'Grady 2011; Barron et al., 2014). The scenario shown in Figure 6a represents the quintessential pattern of GDV throughout the year in an unaltered ecosystem – high greenness and leaf moisture year round. Where access to groundwater is tenuous there is likely to be a minor drop in greenness and foliage moisture in dry seasons (Figure 6b). These areas may or may not host GDV but a transition from A to B may signify GDV health decline. More intense fluctuations in greenness and foliage moisture between wet and dry seasons may simply indicate a non-GDV location when viewed in isolation (Figure 6c) but alarming loss in GDV if it represents change (e.g. A to C). Locations with invariant, low levels of greenness and foliage moisture (Figure 6d) are unlikely to be GDV locations, and if they were, would now indicate a total loss of phreatophytes. An ability to stay green and moist in the dry season thus indicates healthy GDV.

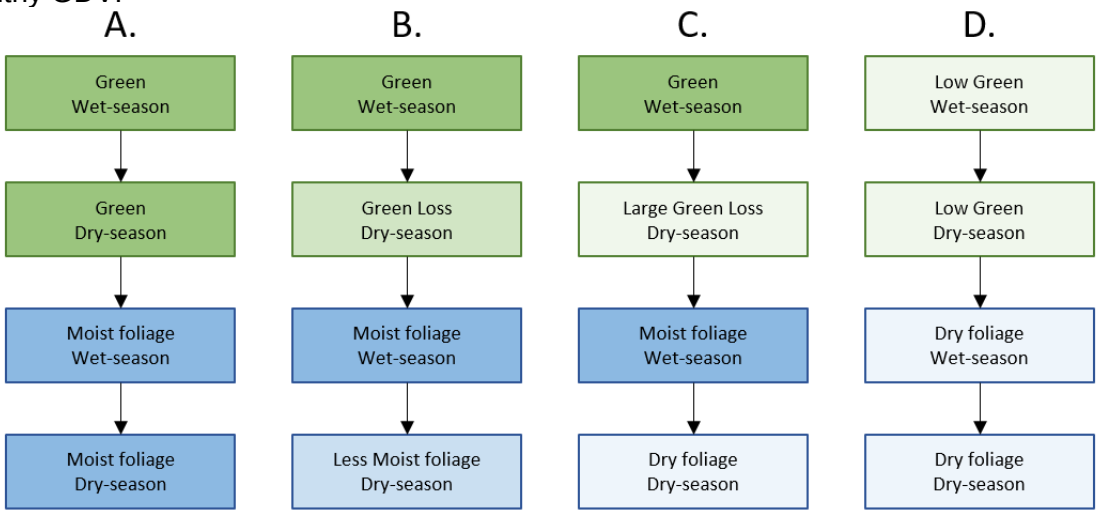


Figure 6: Conceptual model of a) Potential GDV; b) Potential GDV with diminishing access to groundwater; c) Fast-drying vegetation not accessing groundwater and not GDV; and d) invariant low greenness and low moisture and not a GDV. Transitions from pattern "a" into a subsequent pattern may indicate a decline in GDV health and triggers investigation.

3.1.1 Spatial multi-criteria modelling

Spatial multi-criteria evaluation (SMCE) refers to the use of a spatial data and Geographic Information System (GIS) to investigate the allocation of areas within a landscape to suit a purpose based on a variety of attributes or layers that the selected landscape produces. Several approaches to SMCE exist, including basic Boolean overlay to more sophisticated weighted, standardised, combinations (Eastman 1999).

As this study models GDV likelihood using multiple layers of varying importance, a weighted combination approach was utilised with fuzzy standardised layers. The flowchart of the entire GDV likelihood model developed and implemented within the GDV Tool is shown in Figure 7. Each main step of the model is described in detail in the subsections below.

3.1.2 Spectral indices

The GDV likelihood model (Figure 7) uses vegetation indices (VIs) and moisture indices (MIs) as model inputs. These indices are calculated from earth observation imagery on-demand from the DEA platform for a range of years and platform (i.e. Landsat or Sentinel) defined by the user. For every year requested, a GDV likelihood layer is generated.

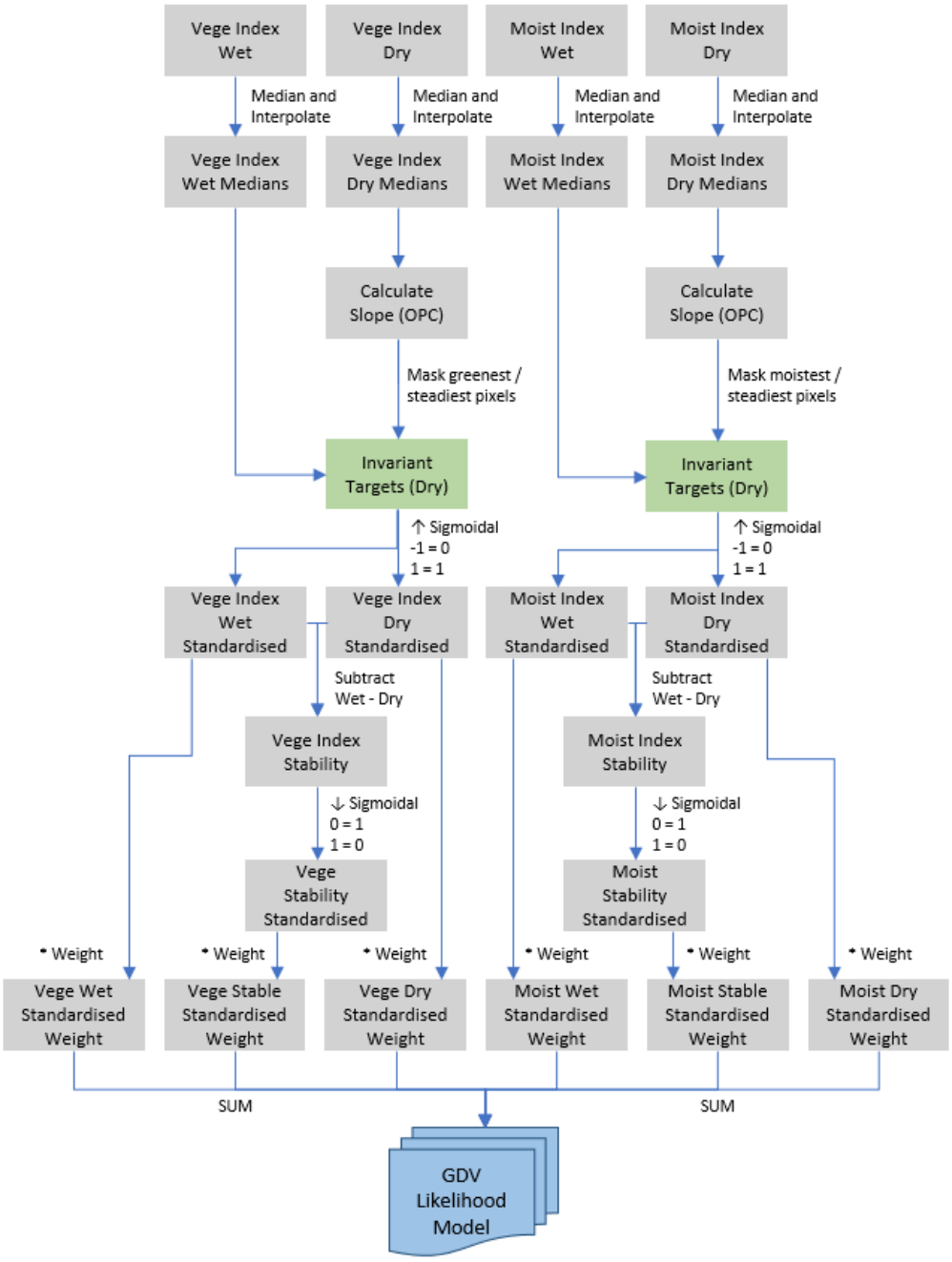


Figure 7: Flowchart for GDV Likelihood generation implemented within the GDV Tool.

An optimal vegetation index (VI) must be chosen that can dichotomise between “GDV” and “non-GDV” areas accurately and consistently at the tenement scale (Robinson et al. 2012). In practise, approaches for identifying GDV and its change over a broad landscape have applied relatively standard VIs that are assumed to provide adequate discrimination between perennial vegetation and background soil reflectance. However, these VIs may not be suitable for semi-arid regions such as the Pilbara bioregion; as most studies have focused on band configurations that only include the visible red and near infrared bands (e.g. Barron et al., 2015; Alaibakhsh et al., 2017; Emelyanova et al., 2018), little is currently known regarding the reliability of different VIs for the detection of GDV in the Pilbara bioregion, or even across neighbouring tenements with varying vegetation, soil type, topography and mining history. Here, we briefly review VIs that are compatible with the sensors carried on both Landsat and Sentinel-2. Table 3 provides references to a cross-section of relevant studies and provides

equations. Note that the VIs in Table 3 are grouped into five types based on the spectral bands they use, and on whether and how they attempt to correct for soil background reflectance.

Slope-based VIs (Table 3) are simple arithmetic combinations of reflectance measurements contrasting the high infrared and low red reflectance of photosynthesising vegetation (Jafari et al., 2007). One of the most widely used slope-based VIs is the Normalised Difference Vegetation Index (NDVI), often ascribed to Rouse et al. (1974). The soil-background corrected group of indices (Table 3) are designed to minimise the effects of soil background on the vegetation signal by incorporating a constant soil adjustment factor, *L*, which varies with the reflectance characteristics of the soil. The Soil-Adjusted Vegetation Index (SAVI) was proposed by Huete (1988), who suggested using an *L* factor of 1.0 for very low vegetation cover, 0.5 for intermediate, and 0.25 for high cover. For *L* = 0, SAVI is equivalent to NDVI (Jafari et al. 2007). For all study areas, we used a value of 0.5.

Table 3: Summary of vegetation indices (VIs) and moisture indices (MIs) used in monitoring studies classified into five groups (see text).

Group	VI/MI	Index	Example Studies	Location	Equation[1, 2, 3, 4]
Slope-based	VI	NDVI	Barron et al. (2014); Andrew and Warrener (2017)	Western Australia	$\frac{\mathrm{NIR}-\mathrm{RED}}{\mathrm{NIR}+\mathrm{RED}}$
Soil-background Corrected	VI	SAVI	Washington-Allen et al. (2006)	Utah, USA	$\frac{(1+L)(NIR-RED)}{\mathrm{NI}R+RED+L}$
Plant-water Sensitive A	MI	NDMI	O'Neill (1996); Barron et al. (2014); Andrew and Warrener (2017)	New South Wales, Western Australia	$\frac{NIR-SWIR1}{NIR+SWRI1}$
	MI	GVMI	Guerschman et al. (2009)	Australia	$\frac{(NIR+0.1)-(SWIR2+0.02)}{(NIR+0.1)+(SWIR2+0.02)}$
	VI	STVI-3	O'Neill (1996); Jafari et al. (2007)	South Australia; New South Wales	$\frac{NIR}{SWIR1+NIR}$
	VI	SLAVI	Lymburner et al. (2000)	New South Wales	$\frac{NIR}{SWIR2+NIR}$
	VI	MSVI-1	O'Neill (1996)	New South Wales	$\frac{NIR}{SWIR1}$
	VI	MSVI-2	O'Neill (1996)	New South Wales	$\frac{NIR}{SWIR2}$
	VI	MSVI-3	O'Neill (1996)	New South Wales	$\frac{NIR}{SWIR1+SWIR2}$
Plant-water Sensitive B	VI	MAVI	Zhu et al. (2014) Hazaymeh (2016)	Northwest Jordan	$\frac{NIR-RED}{NIR+RED+SWIR1}$
Orthogonal Transforms	VI	T. Cap	Jafari 2007	South Australia	Orthogonal Transformation

1 RED: red band (Landsat B4, Sentinel B4)
2 NIR: near-infrared (Landsat B5; Sentinel 2 B8)
3 SWIR1: shortwave infrared 1 (Landsat B6, Sentinel B11)
4 SWIR2: shortwave infrared 2 (Landsat B7, Sentinel B12)

The Plant-water Sensitive A group consists of VIs that do not utilise the visible portion of the spectrum but instead capitalise on the differences in reflectance of the near-infrared and shortwave-infrared areas of the electromagnetic spectrum. The rationale behind these indices is to capitalise on information relating to leaf moisture (mid-infrared, shortwave-infrared), but

also leaf structure (near-infrared; O'Neill 1996). The Normalised Difference Moisture Index (NDMI) has been included in this group as it is reliant on near- and shortwave-infrared bands for mapping leaf and canopy moisture (Lehmann et al., 2013) and has been found to be useful in monitoring vegetation in Australia's rangelands (e.g. O'Neill 1996; Andrews 2017). The Plant-water Sensitive B group adds the red region of the electromagnetic spectrum to the infrared to capitalise on information relating to chlorophyll.

The Orthogonal Transformations group contains one vegetation index ('Greenness') derived from a method known as Tasselled Cap Transformation (TCT; Kauth and Thomas 1976). TCT is considered a special case of principal components analysis which transforms all spectral bands (e.g. visible and infrared) within satellite image data to a new set of orthogonal axes often labelled as 'brightness', 'greenness' and 'wetness'. 'Greenness' is associated with green vegetation, while brightness and wetness are associated with bare-soil and water, respectively. Effectively, this means that background soil reflectance has been handled in the brightness component and therefore greenness is strictly green vegetation. For this reason, greenness has previously been used for vegetation monitoring in Australia's arid regions (e.g. Jafari 2007).

3.1.3 Index seasonal medians

Calculation of perennially high greenness and perennially high leaf moisture useful in GDV detection requires that peak wet and dry season VIs and MIs are calculated from earth observation data (O'Grady 2011; Barron et al., 2014). Months representing peak wet and dry seasons were identified to include December to February and September to November, respectively. This was achieved by charting maxima and minima monthly precipitation minus pan evaporation (Barron et al., 2014) using rainfall and evaporation data from the Australia Bureau of Meteorology (BoM 2020; see Figure 1). Earth observation data outside of peak wet and dry season months are then discarded and annual wet and dry seasonal median images are calculated from DEA platform data to develop images representing a "typical" pixel for those two seasons (Furby and Campbell 2001; Flood 2013).

Data gaps resulting from significant cloud cover or satellite malfunction and VI and MI outliers resulting from extraordinary climatic events (e.g. bushfires and post-fire greening) can be optionally corrected. Missing VI and MI imagery can be interpolated using linear interpolation for wet and dry seasons, separately. Alternatively, years with data gaps can be removed from the dataset altogether. Additionally, VI and MI outliers may be removed from the dataset by performing a z-test (Jia et al. 2014). Calculated z-scores above or below a user-defined z-score threshold (e.g. 1.96 or p-value 0.05) will be eliminated. Alternatively, outliers can be retained in the dataset if the user so choses.

3.1.4 Invariant target standardisation

Broad scale time-series monitoring of vegetation and moisture cover depends on relationships between field observations of cover and the spectral information measured by satellite imagery sensors. Variation in measured spectral information caused by factors other than variation in vegetation and moisture cover on the ground can reduce the accuracy at which vegetation cover estimates can be made from the imagery. Reduction of these variations is carried out via standardisation (Qiu 2019).

Standardisation is undertaken using a 'pseudo-invariant target site' approach (Furby and Campbell, 2001). This approach standardises all VI and MI images to the same consistently green and moist 'sites' (i.e. pixels) detected automatically across the entire time-series dataset. This process is undertaken in several steps. First, median VI and MI images are calculated using all available imagery for the dry season. Next, these images are thresholded

by a user-defined percentile (e.g. 99%); any VI pixel values lower than the user-defined percentile are masked out, leaving only the 'greenest' and 'moistest' sites.

Resulting sites are further restricted only to significantly 'steady' pixels across the entire time-series. This is achieved by computing the linear trend and slope using orthogonal polynomial coefficients (Snedecor 1956; Draper and Smith 1998). The resulting VI and MI slope layers show increasing (slope > 0), decreasing (slope < 0), and no trend at all (slope = 0). Obtaining the steadiest (i.e. no trend) VI and MI pixels is achieved by restricting slope layers to values equal or close to 0, which is also achieved by a user-defined percentile (e.g. 5%). Invariant green and moist sites are simply obtained by intersecting the 'greennest', 'moistest' and 'steadiest' masks.

VI and MI images from wet and dry seasons are ultimately standardised to the median value of all pixel values captured within the invariant target site mask as this value is the value of a typical pixel with highly stable greenness and moisture. Image values are standardised using fuzzy sigmoidal functions (Robertson et al. 2004) resulting in commensurate layers scaled between 0 and 1.

3.1.5 Perennial vegetation and moisture stability

Perennial high greenness and perennial high leaf moisture have been found to permit the detection of GDV from earth observation data (O'Grady 2011; Barron et al., 2014). The rationale is that unchanging vegetation and leaf moisture are potentially indicative of vegetation with access to plentiful groundwater, allowing it to stay green and moist even under harsh hot and dry conditions.

Stable VI and MI areas are calculated by subtracting standardised wet season VI and MI images from standardised dry season VI and MI images. Stability, which improves the likelihood of a GDV-site when coupled with high vegetation and moisture images, can be inferred from small differences (ideally 0) between these two extremes (i.e. -1 to 1). Vegetation and moisture stability layers are then standardised using an increasing-decreasing fuzzy sigmoidal function (Robertson et al., 2004), ensuring images are also on a scale of 0-1 and compatible with standardised vegetation and moisture layers.

3.1.6 Analytical hierarchy process: pairwise comparison

Weights for each layer of evidence (VIs, MIs and stability for wet and dry seasons) were developed using pairwise comparison in the context of the Analytical Hierarchy Process (Saaty, 1980). The pairwise comparison approach typically involves comparing a range of phenomena (e.g. spatial layers) in pairs to judge which of each phenomenon is more important or has a greater amount of some quantitative property, or whether or not the two entities are identical (David 1963). For each comparison made, a pairwise comparison value obtained from a nine point scale (presented in Figure 8) is entered into a pairwise comparison matrix.

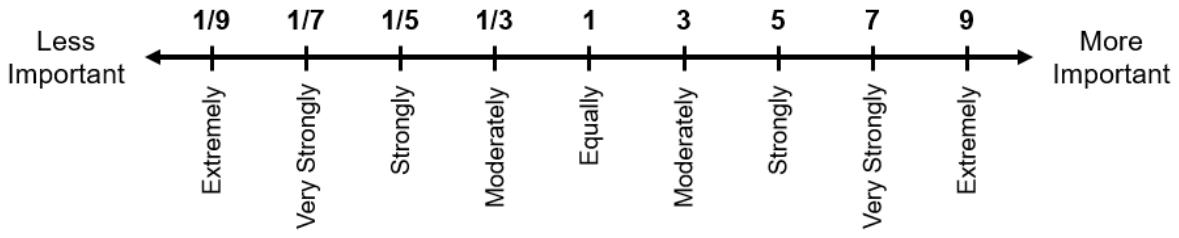


Figure 8: Nine point criteria weighting scale used during the Pairwise Comparison process (adapted from Pauer et al. 2016).

Each wet and dry season VI, MI and stability criterion were compared using the nine point comparison scale (Figure 8) and entered into comparison matrix (see Section 4.1.1). The rationale underpinning the assignment of importance in each criteria comparison included: 1) dry season layers were more important than wet season layers; 2) VI layers (representing greenness) were more important than MI layers (moisture); and 3) VI stability was more important than MI stability.

Once criteria were compared and assigned a comparison value, missing values were filled by inverting and transposing existing values. The entire comparison matrix was normalised by dividing each comparison value by the sum of its column in the matrix (David 1963). Finally, criteria weights used in the GDV likelihood model were derived by calculating the mean of each row in the normalised comparison matrix; the resulting criteria weight for each dry and wet season VI, MI and stability criteria summed to 1. Ultimately, the VI, MI and stability evidence layers for dry and wet seasons of each year are multiplied by the corresponding weights and summed to produce a GDV likelihood layer. The pairwise comparison method ensures GDV likelihood model outputs are continuous on a scale of 0-1, with 0 representing locations of very low GDV likelihood and 1 representing very high GDV likelihood.

3.1.7 Accuracy assessment

All of the VIs presented in Table 3 were tested for their ability to discriminate GDV from non-GDV at each study area. This was achieved by generating a GDV likelihood model for 2018 based on each VI (along with NDMI) (for each study area) and comparing GDV likelihood values to ground-truthed locations of GDV and non-GDV species (i.e. *E. camaldulensis*, *E. victrix* and/or *M. argentea*) via ROC analysis. Prior to ROC analysis, ground-truthed GDV tree observations were transformed into a binary ("GDV" = 1, "non-GDV" = 0) format based on observations made during field surveys. GDV species that were observed to have a condition rank less than 3 and/or recorded as a sapling were modified to represent non-GDV, which also included non-GDV species, tussock grasses and bare ground. Note, while field surveys were undertaken in 2019, the DEA Sandbox had not yet provisioned Landsat imagery for 2019 and thus there was a year difference between likelihood models and ground truthing data.

The predicted GDV likelihood values were extracted at each ground-truthing location and sorted in ascending order. The true positive rate (TPR) and false positive rate (FPR) was then calculated and a ROC plot was generated by plotting the FPR (x-axis) against the TPR (y-axis). The formula for the TPR and FPR are given below (Robinson et al., 2012):

$$TPR = \frac{nGP_i}{nGP} \quad \text{Eq. 1}$$

$$FPR = \frac{nPP_i}{nPP} \quad \text{Eq. 2}$$

where nGP_i is the cumulative number of "GDV" points at threshold *i*, *nGP* is the total number of "GDV" points, nPP_i is the cumulative number of "non-GDV" points at threshold *i* and *nPP* is the total number of "non-GDV" points.

The area under the curve (AUC) was generated from the ROC plots of each model. A GDV likelihood model that perfectly discriminates between "GDV" and "non-GDV" will have an AUC of 1, while a model no better than random chance has an AUC of 0.5 (Fielding and Bell 1997; Fawcett 2006). Based on the ranges proposed by Hosmer and Lemeshow (2000), an AUC ≥

0.7 for any model tested suggests adequate discrimination and provided the answer with statistical confidence for where remote sensing could be utilised. In contrast, a model with an AUC less than 0.7 requires additional information to provide adequate discrimination.

Histograms of AUC values were prepared for each of the GDV likelihood models for each of the different VIs listed in Table 3 to provide an overview of GDV likelihood model performance. The VI with the highest AUC is generally the best discriminator of GDV classes (Zweig and Campbell 1993) although this can depend on the application of the model (e.g. Robinson et al., 2010).

3.1.8 GDV-likelihood dichotomisation

GDV evidence layers are thresholded to dichotomise GDV locations from non-GDV locations to form contiguous GDV areas. This process is implemented in one of two ways, depending on user requirements. First, dichotomisation can be accomplished via Receiver Operating Characteristic (ROC)-slicing (e.g. Robinson et al. 2010) using observed field-based GDV species locations versus non-GDV 'background'. This approach requires a shapefile of ground-truthed GDV species locations (presence and absence) and finds the model value that maximises known GDV points and minimises false positives (also see below). As an alternative, when ground-truthing data is not available, standard deviation can be applied to threshold 'background' based on a user-defined number of standard deviations (e.g. 2.0).

3.1.9 Bushfire detection and mapping

It is expected that bushfire events have considerable influence on GDV likelihood model output layers due to the sensitivity of VIs to fire (Malak and Pausas 2006). Potential bushfire areas across years can be optionally detected from the DEA data to assist users in explaining GDV likelihood discrepancies. Based on Boer et al. (2008), this is achieved by calculating the Normalised Burn Ratio (NBR) index and, from this, the delta NBR (dNBR):

$$NBR = \frac{NIR - SWIR}{NIR + SWIR} \quad \text{Eq. 3}$$

$$dNBR = PrefireNBR - Postfire\ NBR \quad \text{Eq. 4}$$

where NIR is near-infrared reflectance (c. midpoint = 0.85 µm) and SWIR is reflectance in the short-wave infrared (c. midpoint = 1.60 µm), PrefireNBR is a preceeding NBR image for a specific year and PostfireNBR is the following NBR image.

Burn severity of a level greater than or equal to low severity (dNBR >= 0.1; Keeley 2009) are extracted from dNBR images and masked into potential burn areas. A majority filter is applied to remove isolated burn pixels.

3.2 Time-series Data Analysis

3.2.1 Trend modelling

Trend modelling is undertaken using the Mann-Kendall (MK) test (Mann, 1945; Hipel and McLeod, 2005) on a time-series of vegetation index images restricted to thresholded GDV likelihood areas derived prior. A MK test determines if there is a monotonic increase, monotonic decrease or no significant change over time over five or more sample values (i.e. five or more years). The MK non-parametric trend test is a function of the ranks of the

observations rather than their actual values and does not require the assumption of normally distributed data (Hamed, 2008) like, for example, linear regression.

MK tests were implemented using the 'original' or 'seasonal' MK test variants. The original MK test is suited for comparing anniversary dates across five or more years in a time-series (e.g. GDV health at a site every January for five years). Alternatively, the seasonal MK test (Hirsch et al., 1982) considers seasonal fluctuations and calculates the seasonal trend. This approach allows more data samples to be used each year in the time-series (e.g. vegetation health every season at a site over five years). In addition to trends, the MK test also allows the mapping of p-values and Sen's slope values (Hussain et al., 2019). The p-value provides a confidence level for interpretation, whilst the Sen's slope value provides a more robust trend that ignores any trend that does not occur for at least 30% of the time-series.

While users have complete control over trend parameters (e.g. MK type, trend date range) in the GDV Tool, trends were calculated for each study area in this report using the seasonal MK test for the time period 2010-2018 (i.e. the most recent eight years) for demonstration purposes (see section 4.2.1).

3.2.2 Change detection

An alternative to trend analysis is time-series change detection. This approach detects different types of land cover change (e.g. vegetation health increase or decline) without the need for five or more years of VI imagery. Instead, users can define a vegetation and soil baseline image and compare any date (at either seasonal or monthly temporal scales) to that baseline to detect different change types and intensity of change.

This approach is implemented using the change vector analysis (CVA) method. CVA is used to describe the characteristics of a change between two images captured at two different dates for two variables (i.e. image bands). The outputs from this process are described as a vector expressing the strength (magnitude) of change as well as the direction (angle) of change with respect to the two variables (Figure 9; Malila 1980). The technique has been successfully utilised to detect both large and small variations in vegetation change in arid and semi-arid environments (e.g. Johnson and Kasischke 1998; Palmer and van Rooyen 1998; Baker et al. 2007; Flores and Yool 2007). The two bands used in CVA for vegetation change detection includes the red and near-infrared bands used by NDVI (Table 3) or the 'greenness' and 'brightness' (or soil) bands derived from the Tasselled Cap Transformation (TCT) technique (Table 3). The use of TCT variables 'brightness' and 'greenness' have been found to be an effective method for measuring riparian and forest change in previous studies (Landmann et al., 2013; Nackaerts et al. 2005).

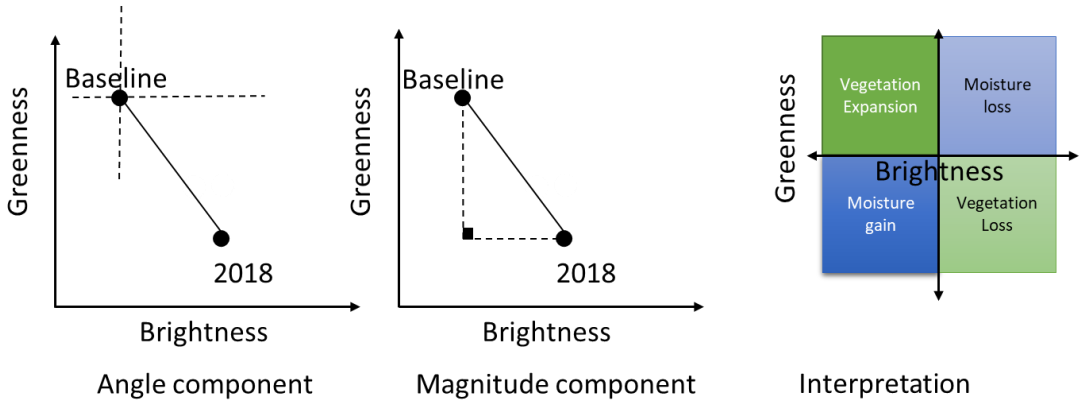


Figure 9: Change Vector Analysis produces an angle and magnitude that can be interpreted based on four quadrants.

By interpreting CVA magnitude and change angle, specific types of change can be highlighted (Figure 9). Angles between 90º to 180º can be used to separate vegetation decline (or bare soil expansion) from other change angles (Vorovencii 2014). Likewise, magnitude values at or near zero identify areas with little to no change, while higher magnitude values represent higher intensities of change. However, unchanged pixels fall within a range about the origin due to factors such as noise and imperfect normalisation. This noise can be removed by applying a threshold to magnitude values; a threshold determined by standard deviation (Baker et al., 2007; Seto et al., 2002).

Users are provided with the ability to select any range of years for their baseline 'greenness' and 'brightness' images, as well as a range of years to compare to the baseline. This provides flexibility in defining study area-specific baselines including from historical imagery acquired before any disturbance was likely to begin (where possible). Likewise, any range of angles can be defined to highlight specific change types. The TCT method is then calculated from earth observation data by converting to top-of-atmosphere reflectance using pre-defined coefficients for Landsat 5, 7 and 8 and Sentinel 2 based on Cristi (1985), Huang et al. (2002), Baig et al. (2014) and Nedkov (2017), respectively.

The main advantage of change detection over trend analysis in the context of determining vegetation decline (or growth) is that vegetation decline can be interpreted relative to a baseline image without the need for more than five consecutive years like the MK test. However, baselines do need to be chosen judiciously because pixel values above the baseline in a state of decline will not be recognised as declining until it crosses the baseline. Similarly, pixels that are below the baseline, but improving, will not be recognised as improving until they cross the baseline. This limits its ability as an early warning system and so the two tools (trend and change detection) should be used together to assist interpretation.

3.2.3 Vegetative health regime and break detection

Change point break detection is used in time-series analysis to detect abrupt shifts in time series trends (i.e. shifts in a time series' instantaneous velocity). This can be applied to a time-series of vegetation health observations to detect anomalous sequences/states in vegetation health history, and to detect dates where sudden changes occurred (e.g. Andersen 2009). Two exact segmentation change point algorithms are implemented into the tool as to assist users in visualising vegetation health break points. These algorithms are the Pruned Exact Linear Time (PELT) and Dynamic Programming search methods (Truong et al. 2019). The PELT search method detects break points from time-series data samples via the minimisation of costs, generally producing quick and consistent break results (Wambui et al. 2015). The Dynamic Programming method has more computational cost than PELT, and is generally considered less strict than PELT, returning more break points (Wambui et al. 2015).

4 Results

4.1 GDV-likelihood Modelling

The initial comparison matrix was developed using the Pairwise Comparison technique (David 1963), and comparison values were assigned using the nine point scale presented in Section 3.1.6 (Figure 8). The pairwise comparison matrix developed from this process (including inverted and transposed values) is presented in Table 4.

Table 4: Pairwise comparison matrix. Values assigned by comparing each evidence layer to each other and assigning a comparison value based on a nine-point scale (David 1963).

	VI Wet	VI Dry	VI Stability	MI Wet	MI Dry	MI Stability
VI Wet	1.00	0.11	3.00	3.00	0.20	5.00
VI Dry	9.00	1.00	5.00	7.00	3.00	7.00
VI Stability	0.33	0.20	1.00	1.00	0.14	3.00
MI Wet	0.33	0.14	1.00	1.00	0.11	3.00
MI Dry	5.00	0.33	7.00	9.00	1.00	9.00
MI Stability	0.20	0.14	0.33	0.33	0.11	1.00
Sum	15.87	1.93	17.33	21.33	4.57	28.00

VI = Vegetation Index
MI = Moisture Index

To calculate criteria weights for VIs, MIs and stability, a normalised criteria weight matrix was derived from Table 4. Each comparison value in the matrix was divided by the sum of its column, resulting in the normalised pairwise comparison matrix presented in Table 5.

Table 5: Normalised pairwise comparison matrix. Normalised comparison values derived by dividing each comparison value by the sum of its column presented in Table 4.

	VI Wet	VI Dry	VI Stability	MI Wet	MI Dry	MI Stability
VI Wet	0.0630	0.0576	0.1731	0.1406	0.0438	0.1786
VI Dry	0.5672	0.5181	0.2885	0.3281	0.6572	0.2500
VI Stability	0.0210	0.1036	0.0577	0.0469	0.0313	0.1071
MI Wet	0.0210	0.0740	0.0577	0.0469	0.0243	0.1071
MI Dry	0.3151	0.1727	0.4038	0.4219	0.2191	0.3214
MI Stability	0.0126	0.0740	0.0192	0.0156	0.0243	0.0357
Sum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

VI = Vegetation Index
MI = Moisture Index

Finally, to calculate the VI, MI and stability criteria weights, the mean of the normalised pairwise comparison values of each row presented in Table 5 were calculated. The final criteria weights used in the GDV likelihood model are already normalised and are presented in Table 6.

Table 6: Finalised criteria weights used in the GDV likelihood model. These weights are derived from the Pairwise Comparison in the context of the Analytical Hierarchy process.

Evidence Layer	Season	Weights
VI	Dry	0.434845
VI	Wet	0.109446
VI Stability	Wet and Dry	0.061272
MI	Dry	0.309005
MI	Wet	0.055179
MI Stability	Wet and Dry	0.030255
Total		1.0

Figure 10, Figure 11 and Figure 12 provide an overview of GDV likelihood model outputs for the BHP Yandi, Roy Hill and Ophthalmia Dam study areas, respectively. Figures are based on the optimal VI determined for each study area during the accuracy assessment process (see Section 4.1.1 for a full description). GDV likelihood model outputs are provided for both satellite platforms for the year 2018 (i.e. the most recent, full satellite year available in the DEA

ODC) in addition to GDV likelihood median summary models of the last decade (2009 - 2018) for Landsat and last four years (2015 – 2018) for Sentinel.

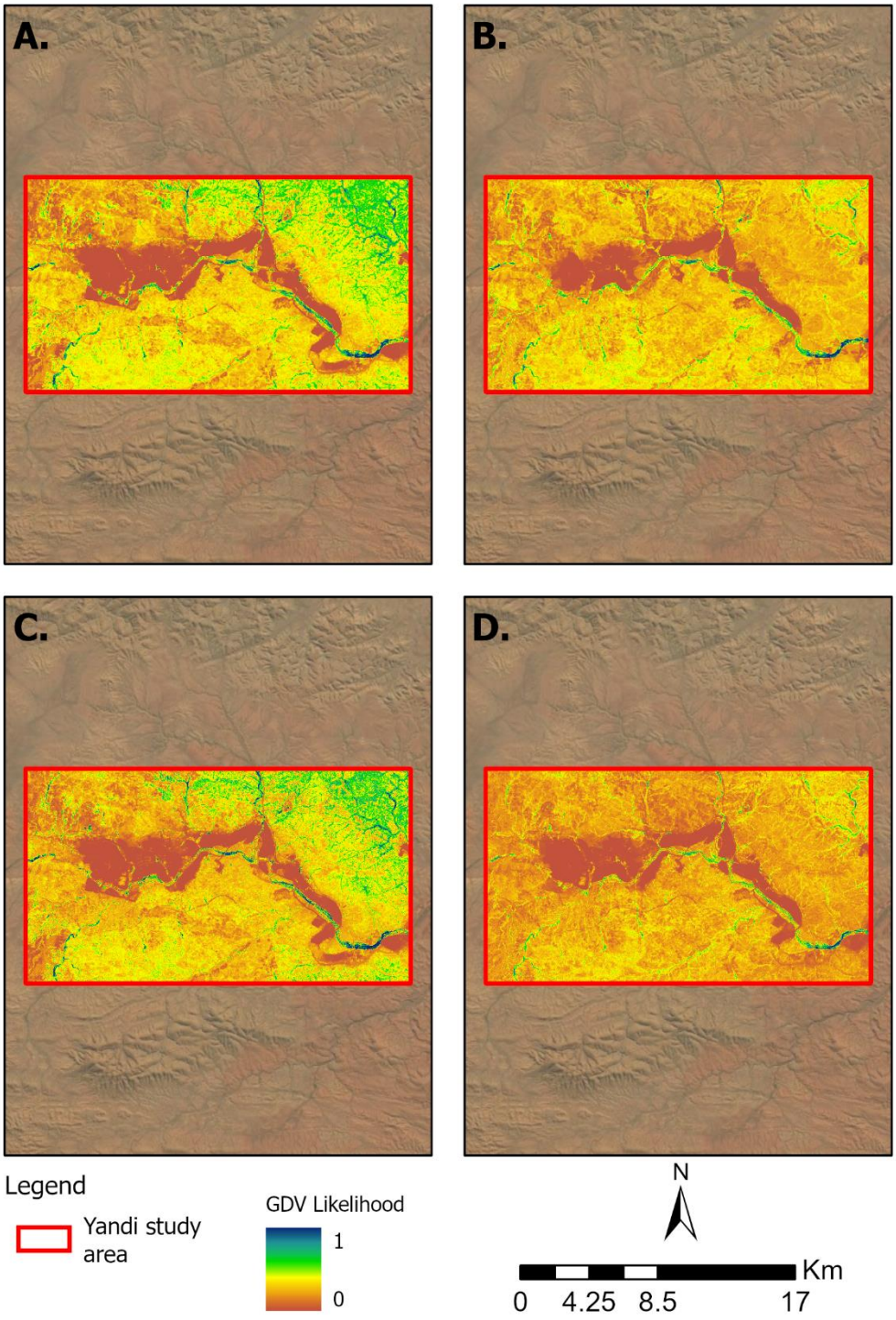


Figure 10: GDV likelihood models for the BHP Yandi study area based on the optimal parameters: a) Landsat-based MAVI for 2018; b) Landsat-based MAVI for median of years 2008-2018; c) Sentinel-based MAVI for 2018; and d) Sentinel-based MAVI for median of years 2015-2018.

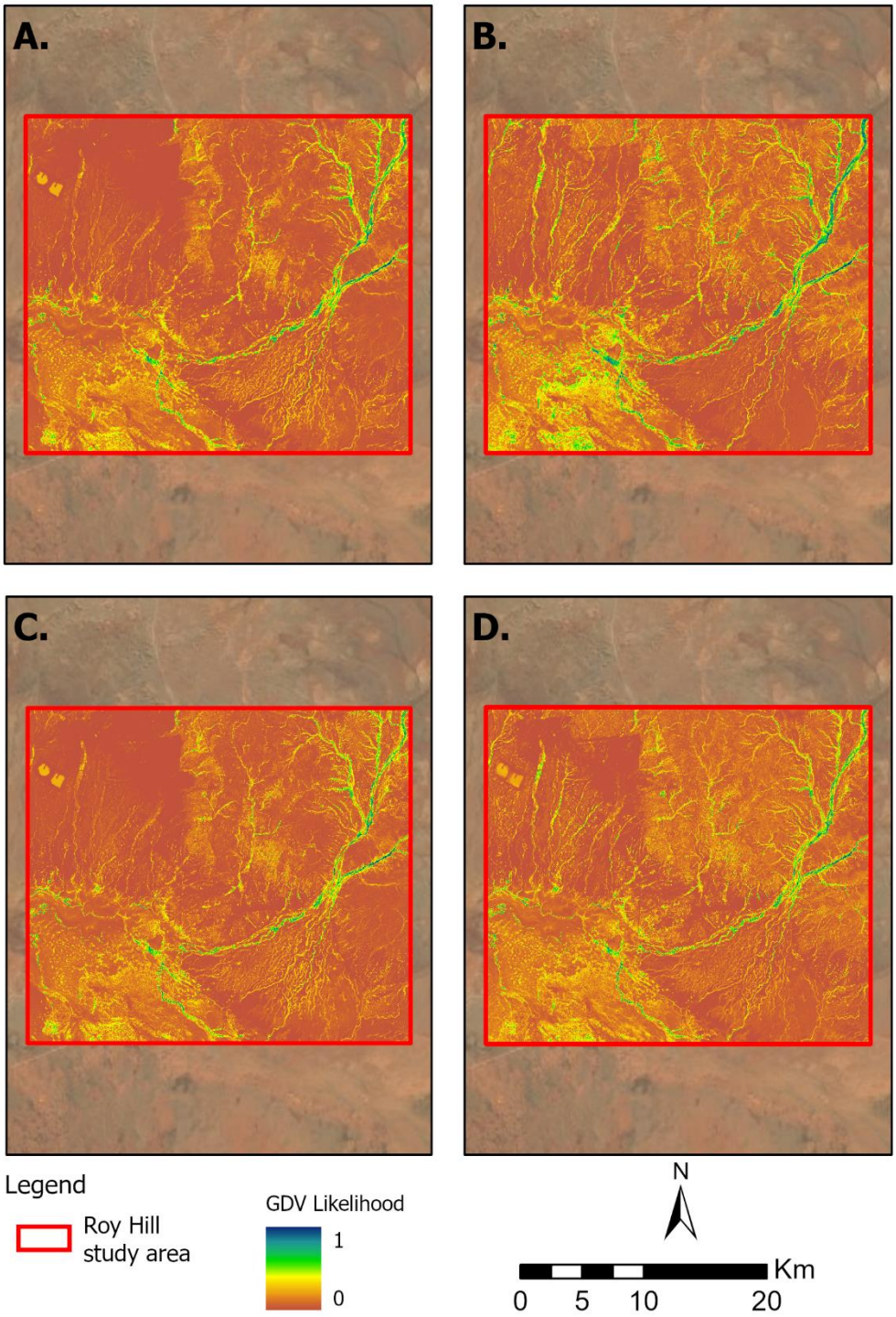


Figure 11: GDV likelihood models for the Roy Hill study area based on the optimal parameters: a) Landsat-based MAVI for 2018; b) Landsat-based MAVI for median of years 2008-2018; c) Sentinel-based MAVI for 2018; and d) Sentinel-based MAVI for median of years 2015-2018.

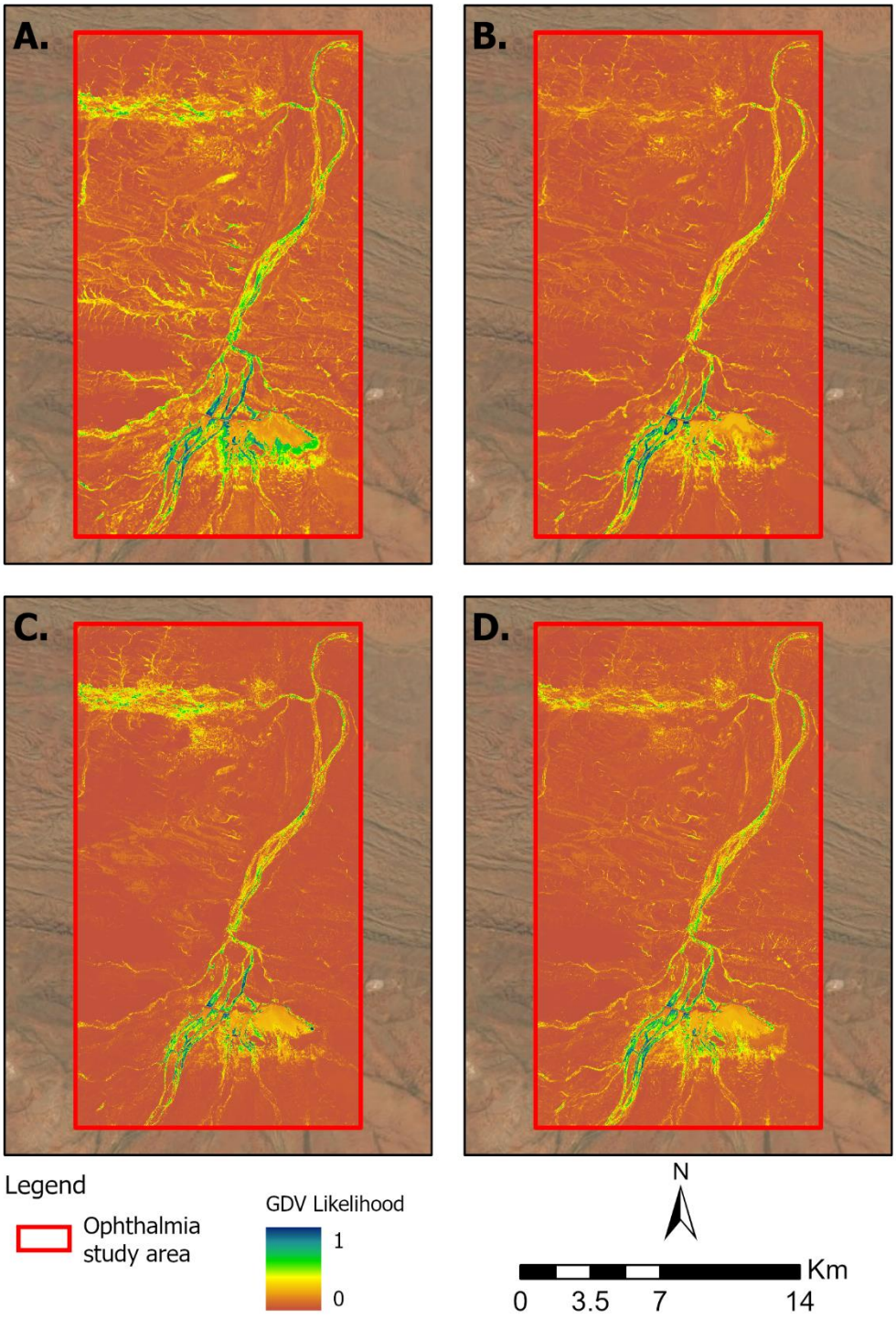


Figure 12: GDV likelihood models for the Ophthalmia Dam study area based on the optimal parameters: a) Landsat-based MAVI for 2018; b) Landsat-based MAVI for median of years 2008-2018; c) Sentinel-based MAVI for 2018; and d) Sentinel-based MAVI for median of years 2015-2018.

4.1.1 Accuracy assessment

Summary statistics of the AUC and confidence interval (CI) determined from ROC analysis for each study area, satellite platform (Landsat and Sentinel) and VI (see Table 3 for VI and MI details) is provided in Table 7. The highest possible discrimination observed for the GDV likelihood model generated for the BHP Yandi study area, based on 332 "GDV" and "non-GDV" observations, was obtained via MAVI (Plant-water Sensitive B group) for both Landsat (AUC = 0.823, CI = 0.781 – 0.868) and Sentinel (AUC = 0.948, CI = 0.904 – 0.967) satellite imagery. This was closely followed by NDVI (Slope-based) and TCT Greenness (Transformation-based) for both sensors. Conversely, the poorest performing VIs were from the Plant-water Sensitive A group, including MSVI-1 for Landsat and MSVI-3 for Sentinel-2 (Table 7).

At the Roy Hill study area, the highest possible AUC obtained was also attributed to MAVI for both Landsat and Sentinel (Table 7). This was followed by SLAVI and TCT Greenness for Landsat and TCT-Greenness and STVI-3 for Sentinel, although the differences were marginal (Table 7). comparatively. Like the BHP Yandi study area, MSVI-1 produced the weakest discriminatory capability between "GDV" and "non-GDV" for Landsat and MSVI-3 for Sentinel.

Table 7: Summary of the AUC for GDV likelihood models derived from each of the nine VIs, for each of the three study areas and satellite platforms (Landsat and Sentinel). Each AUC is derived from the most recent fully available satellite imagery year (2018). Top three AUC's are in bold type font.

Study Area	VI[1]	N[2]	Landsat 5, 7, 8		Sentinel 2	
			AUC	CI	AUC	CI
BHP Yandi	NDVI	332	**0.821**	0.789 – 0.856	**0.929**	0.899 – 0.962
	SAVI	332	0.807	0.771 – 0.839	0.914	0.879 – 0.945
	STVI-3	332	0.786	0.745 – 0.807	0.895	0.861 – 0.947
	SLAVI	332	0.801	0.777 – 0.844	0.914	0.867 – 0.951
	MSVI-1	332	0.748	0.723 – 0.783	0.884	0.836 – 0.921
	MSVI-2	332	0.761	0.733 – 0.789	0.889	0.863 – 0.933
	MSVI-3	332	0.759	0.729 – 0.778	0.878	0.858 – 0.921
	MAVI	332	**0.823**	0.781 – 0.868	**0.948**	0.904 – 0.967
	TCT Greenness	332	**0.817**	0.785 – 0.852	**0.919**	0.852 – 0.943
Roy Hill	NDVI	267	0.799	0.758 – 0.827	0.902	0.895 – 0.941
	SAVI	267	0.781	0.765 – 0.829	0.896	0.871 – 0.939
	STVI-3	267	0.804	0.778 – 0.838	**0.915**	0.884 – 0.943
	SLAVI	267	**0.821**	0.781 – 0.853	0.912	0.875 – 0.940
	MSVI-1	267	0.767	0.727 – 0.799	0.889	0.867 – 0.937
	MSVI-2	267	0.784	0.753 – 0.823	0.893	0.875 – 0.938
	MSVI-3	267	0.771	0.758 – 0.808	0.879	0.857 – 0.922
	MAVI	267	**0.836**	0.787 – 0.872	**0.938**	0.897 – 0.968
	TCT Greenness	267	**0.813**	0.789 – 0.838	**0.919**	0.889 – 0.944
Ophthalmia Dam	NDVI	280	0.853	0.815 – 0.897	0.938	0.917 – 0.961
	SAVI	280	0.849	0.812 – 0.881	0.935	0.909 – 0.953
	STVI-3	280	0.869	0.821 – 0.912	**0.951**	0.934 – 0.987
	SLAVI	280	**0.871**	0.842 – 0.908	0.932	0.896 – 0.955
	MSVI-1	280	0.821	0.783 – 0.847	0.899	0.867 – 0.943
	MSVI-2	280	0.832	0.787 – 0.876	0.901	0.859 – 0.948
	MSVI-3	280	0.819	0.777 – 0.844	0.897	0.851 – 0.937
	MAVI	280	**0.891**	0.861 – 0.952	**0.963**	0.922 – 0.978
	TCT Greenness	280	**0.884**	0.857 – 0.913	**0.957**	0.927 – 0.976

1 Number of observations the AUC is derived from during ROC analysis.
2 SAVI L value = 0.5.
CI = Confidence Interval

Ophthalmia Dam followed the same pattern with the top three VIs as Roy Hill. The highest AUC produced at Ophthalmia Dam study area was again obtained via MAVI for both Landsat (AUC = 0.891, CI = 0.861 – 0.952) and Sentinel (AUC = 0.963, CI = 0.922 – 0.978). Notably, both Landsat- and Sentinel-derived VIs were overall higher (by about 0.05 on average) than

the other study areas. The TCT Greenness VI was the next highest both Landsat and Sentinel (Table 7). This was followed by SLAVI (Landsat) and STVI-3 (Sentinel). The weakest performing VI was MSVI-3 for both Landsat and Sentinel (Table 7).

Figure 13 presents the AUC calculated for each VI at each study area for Landsat (Figure 13a) and Sentinel (Figure 13b) satellite platforms. The best performing VIs using Landsat imagery were MAVI (mean AUC = 0.852), TCT Greenness (mean AUC = 0.838) and SLAVI (mean AUC = 0.831) and MAVI (mean AUC = 0.949), TCT Greenness (mean AUC = 0.932) and NDVI (mean AUC = 0.923) when using Sentinel imagery. In contrast, the lowest performing Landsat-derived VIs, on average, were MSVI-1 (mean AUC = 0.778), MSVI-3 (mean AUC = 0.783) and MSVI-2 (mean AUC = 0.792). Similarly, the lowest performing Sentinel-derived VIs irrespective of study area, on average, were MSVI-3 (mean AUC = 0.884), MSVI-1 (mean AUC = 0.891) and MSVI-2 (mean AUC = 0.894).

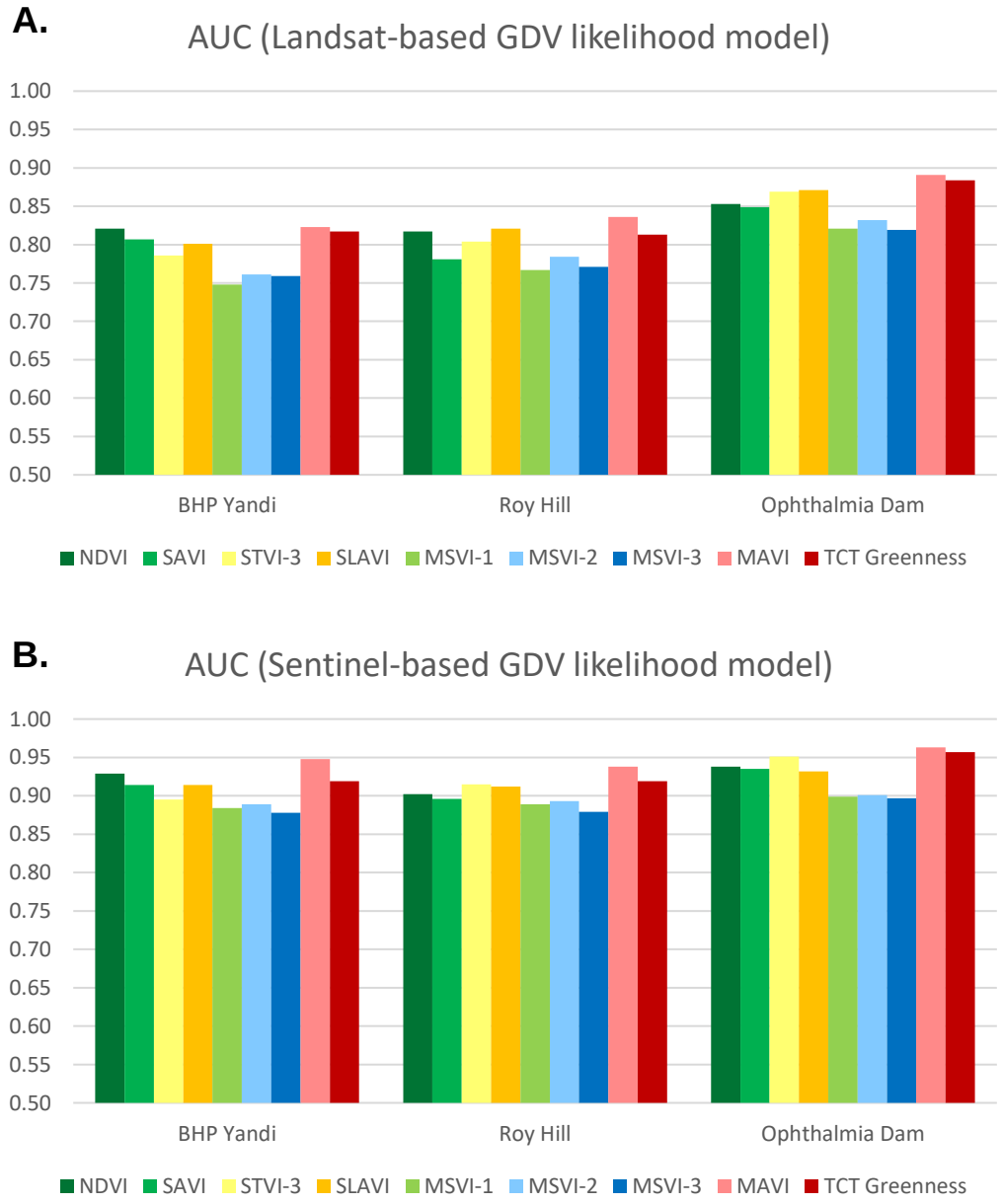


Figure 13: AUC statistic for the nine VIs tested at the three study areas in the year 2018 for the Landsat and Sentinel satellite platforms.

4.1.2 GDV-likelihood dichotomisation

Finally, Figure 14, Figure 15 and Figure 16 provide an overview of the dichotomised GDV likelihood model outputs for the BHP Yandi, Roy Hill and Ophthalmia Dam study areas, respectively.

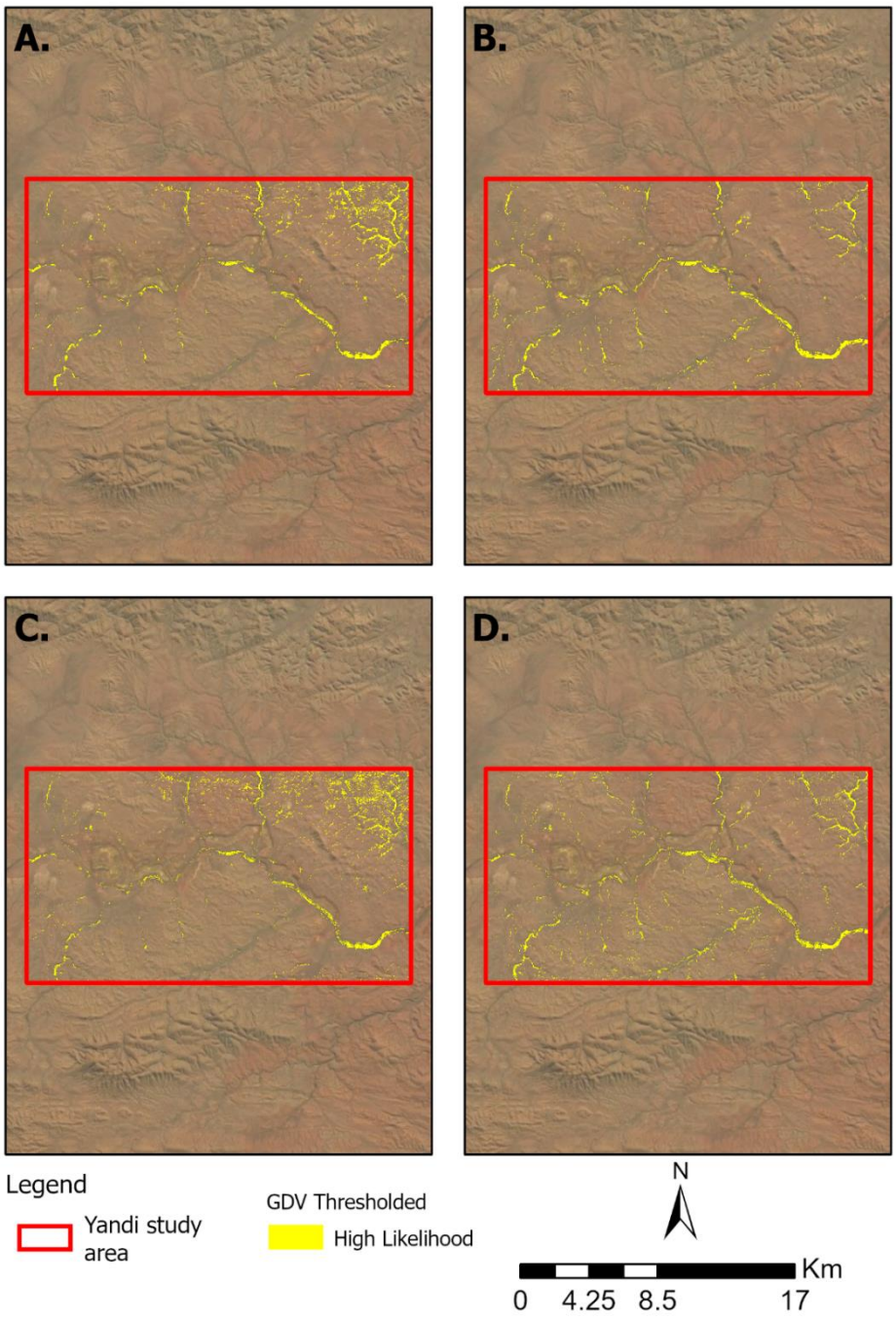


Figure 14: Thresholded GDV likelihood models for the BHP Yandi study area based on the optimal parameters: a) Landsat-based MAVI for 2018 thresholded via ROC analysis; b) Landsat-based MAVI for median of years 2008-2018 thresholded via standard deviation = 2.0); c) Sentinel-based MAVI for 2018 thresholded via ROC analysis; and d) Sentinel-based MAVI for median of years 2015-2018 thresholded via standard deviation = 2.0.

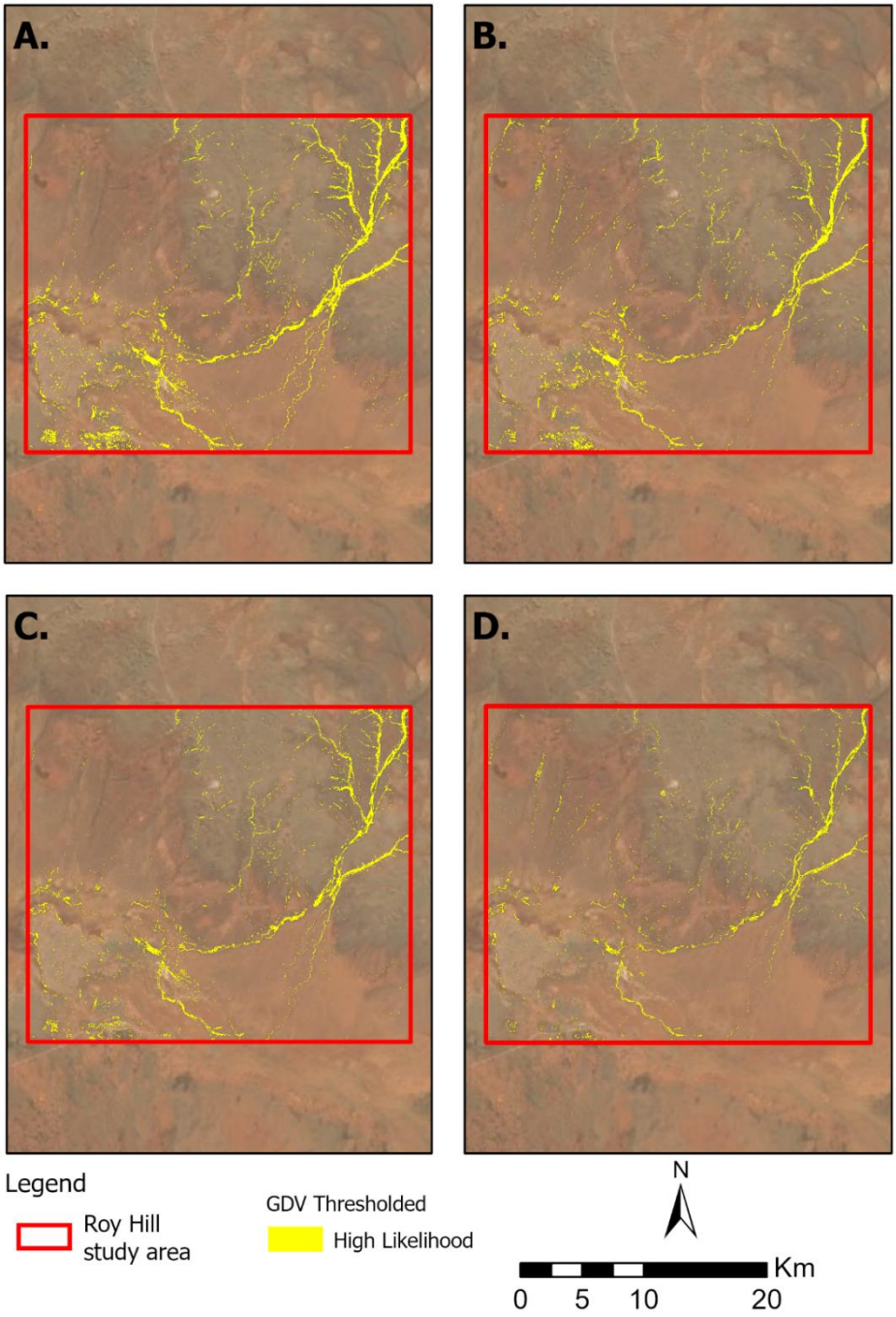


Figure 15: Thresholded GDV likelihood models for the Roy Hill study area based on the optimal parameters: a) Landsat-based MAVI for 2018 thresholded via ROC analysis; b) Landsat-based MAVI for median of years 2008-2018 thresholded via standard deviation = 2.0); c) Sentinel-based MAVI for 2018 thresholded via ROC analysis; and d) Sentinel-based MAVI for median of years 2015-2018 thresholded via standard deviation = 2.0.

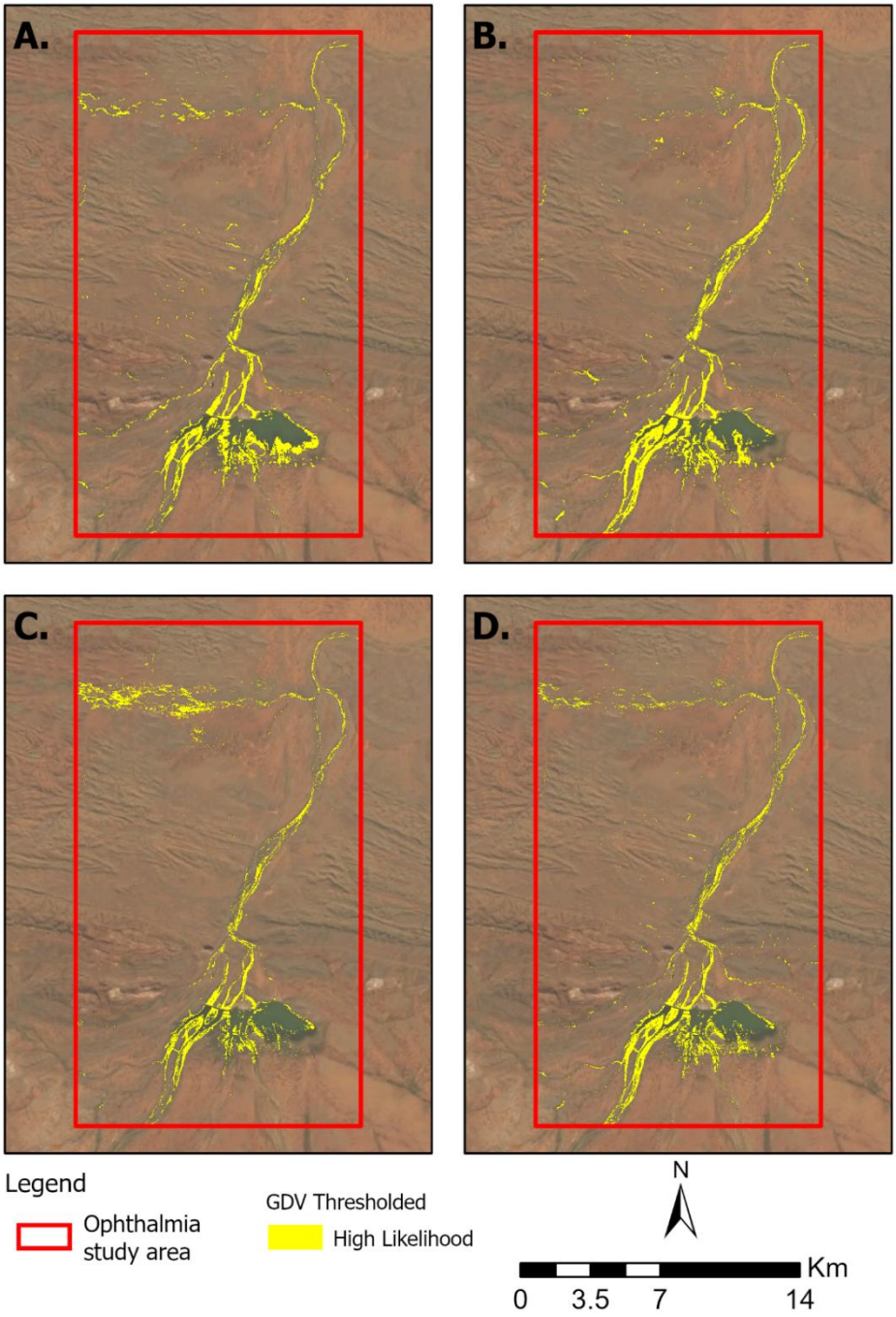


Figure 16: Thresholded GDV likelihood models for the Ophthalmia Dam study area based on the optimal parameters: a) Landsat-based MAVI for 2018 thresholded via ROC analysis; b) Landsat-based MAVI for median of years 2008-2018 thresholded via standard deviation = 2.0); c) Sentinel-based MAVI for 2018 thresholded via ROC analysis; and d) Sentinel-based MAVI for median of years 2015-2018 thresholded via standard deviation = 2.0.

4.2 Time-series Data Analysis

4.2.1 Trend Modelling

A seasonal Mann-Kendall (MK) test (Mann, 1945; Hirsch et al., 1982) was performed for each study area. This was undertaken within thresholded GDV likelihood areas generated for the period 1987-1997. This range of years was selected to capture areas of both historic and existing GDV to clearly demonstrate the influence of vegetation loss on the MK trend. As such, some potential GDV areas may occur in unexpected, recently cleared areas. The optimal VI MAVI, based on Landsat satellite imagery, was generated for each study area, and four seasonal medians were generated for every year over the period 2010-2018 (i.e. four seasonal vegetation samples at every image pixel for eight consecutive years). A seasonal MK test was then undertaken on each study area, resulting in a Tau (i.e. trend directions) and confidence (i.e. p-values) layer of GDV for each study area.

The Tau map highlights areas of vegetation that have experienced on-going growth, health increase or 'greening' (monotonic ascending trend; Tau value > 0), on-going health decline or 'green loss' (monotonic declining trend; Tau value < 0), or no clear health trend (no monotonic trend; Tau = 0) occurring over at least five years (Eastman et al. 2013). Additionally, a confidence map that provides the p-value of each Tau trend pixel is provided, offering helpful context on the confidence level for interpretation (Mann, 1945). Combining the Tau trend map with areas on the confidence map where $p < 0.05$, a significant Tau map can be derived. Figure 17, Figure 18 and Figure 19 present an overview of the raw MK Tau and significant MK Tau (where $p < 0.05$) layers for the BHP Yandi, Roy Hill and Ophthalmia Dam study areas, respectively.

A range of trends within potential GDV areas can be observed at each study area. Based on Figure 17a, the BHP Yandi study area MK test suggested most trends over the 2010-2018 test period were relatively reasonably stable (see yellow colouring; Tau value close or equal to 0). Potential GDV areas within a closer proximity to mine developments that have occurred over the period 2010-2018 (see Figure 17a, b MD1 and MD2) show areas of significant decline (Figure 17b). In contrast, significant increasing trends were also observed in the Fortescue Creek alongside older mine development areas (see Figure 17a, b MD3).

The Roy Hill and Ophthalmia Dam study areas also show a similar range of trends. The Roy Hill area, based on Figure 18a, presents a range of Tau values ranging from low to high-moderate vegetation increase within potential GDV for most of the study area over the period 2010-2018. The exception are the significant decline areas (see Figure 18a, b MD1 and MD2) that has been cleared for recent mine development. The Ophthalmia Dam area MK trend results (Figure 19) suggest vegetation lining the Fortescue River has also experienced low to moderate vegetation increase over the period 2010-2018. Areas closer to nearby mining activity (MD1) and the dam itself (MD2) show areas of vegetation decline.

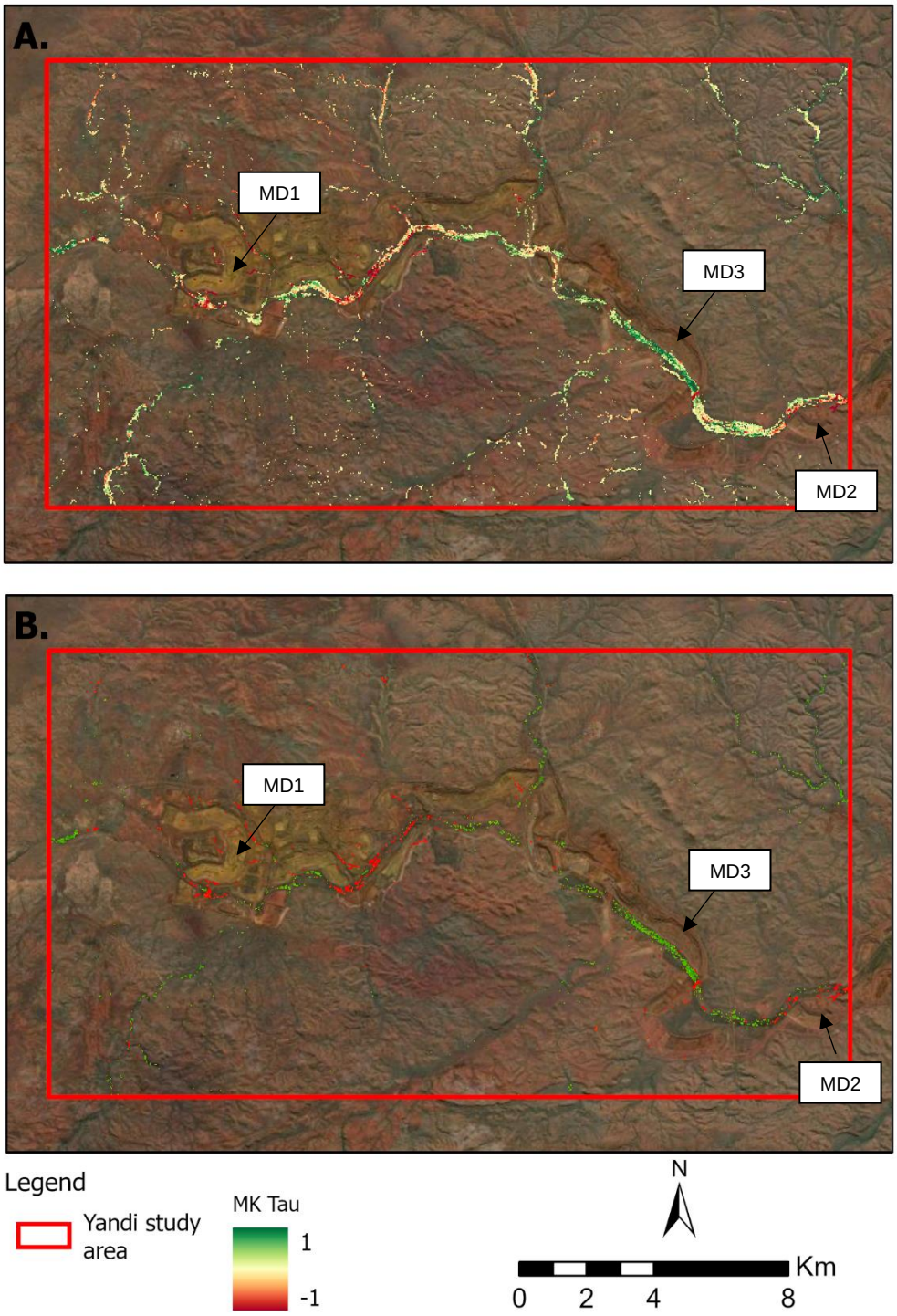


Figure 17: Overview of Tau layer for the BHP Yandi study area derived from seasonal Mann-Kendall (MK) test. MK test was undertaken using MAVI-based images for period 2010-2018 (four seasonal samples per year). A Tau of -1 is perfectly continuous decline, 1 is a perfectly continuous increase, and 0 is no trend. Not all MK trends are statistically significant: a) shows all trends with p-value <= 1; and b) shows only significant trends (p-value < 0.05).

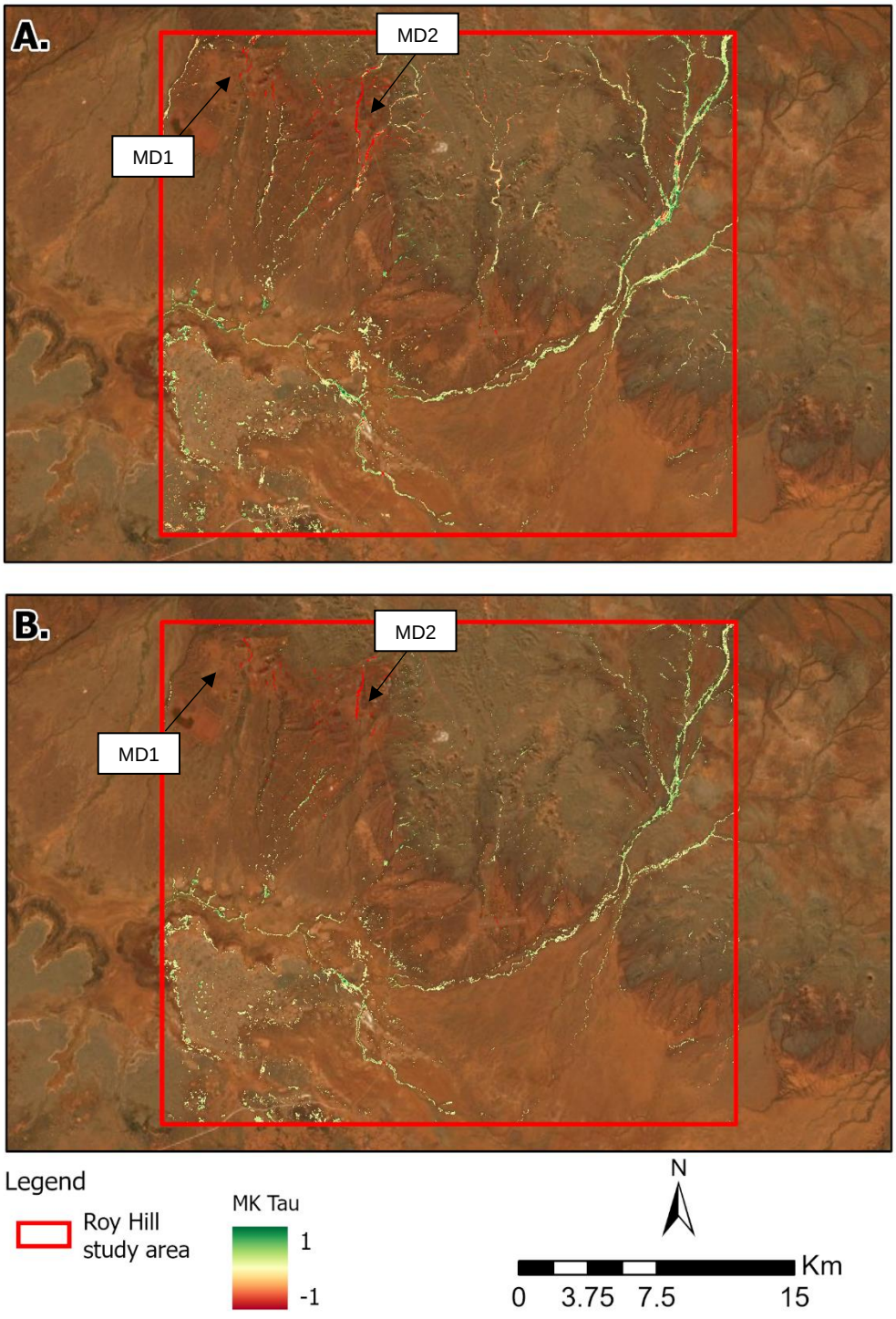


Figure 18: Overview of Tau layer for the Roy Hill study area derived from seasonal Mann-Kendall (MK) test. MK test was undertaken using MAVI-based images for period 2010-2018 (four seasonal samples per year). A Tau of -1 is perfectly continuous decline, 1 is a perfectly continuous increase, and 0 is no trend. Not all MK trends are statistically significant: a) shows all trends with p-value <= 1; and b) shows only significant trends (p-value < 0.05).

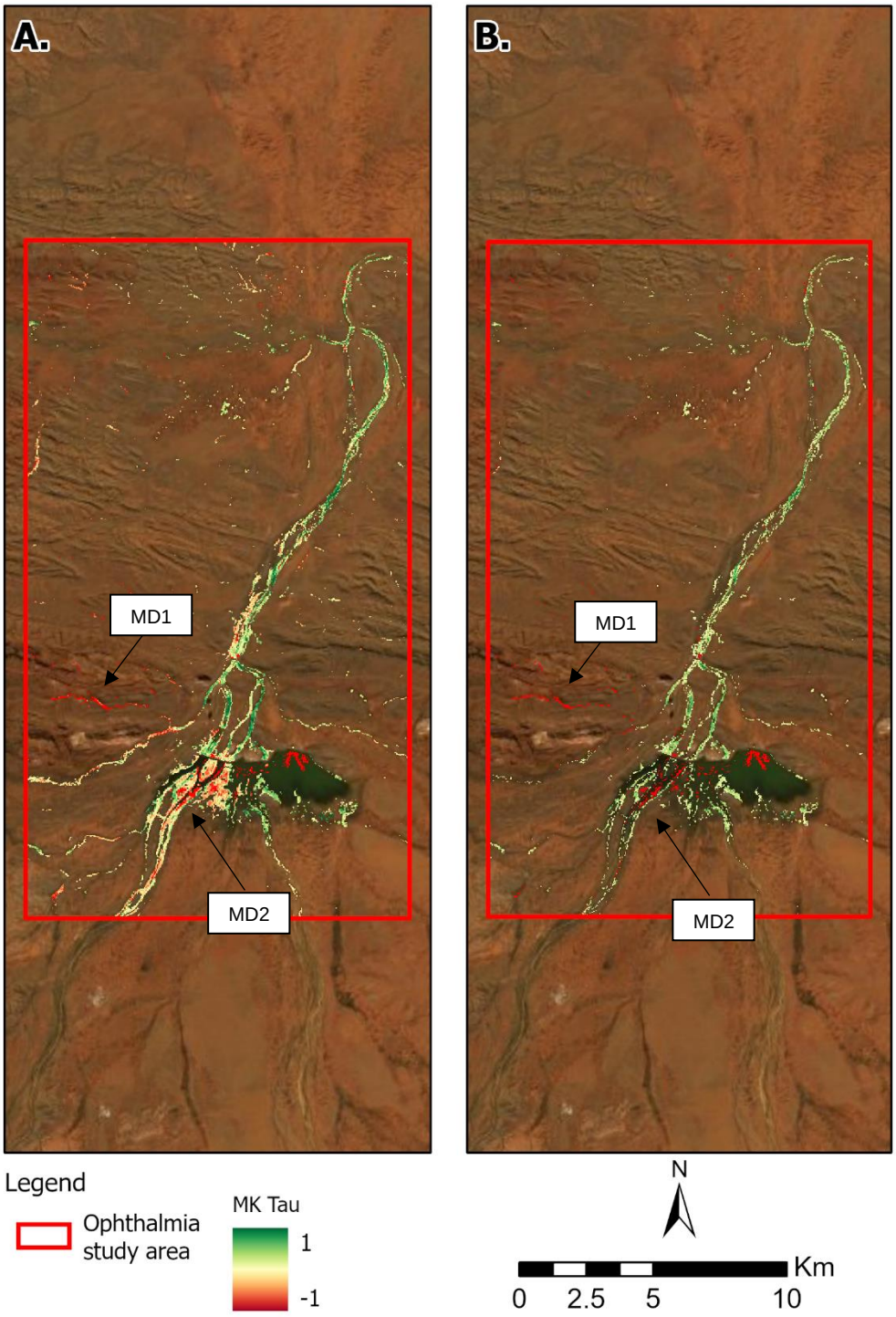


Figure 19: Overview of Tau layer for the Ophthalmia study area derived from seasonal Mann-Kendall (MK) test. MK test was undertaken using MAVI-based images for period 2010-2018 (four seasonal samples per year). A Tau of -1 is perfectly continuous decline, 1 is a perfectly continuous increase, and 0 is no trend. Not all MK trends are statistically significant: a) shows all trends with p-value <= 1; and b) shows only significant trends (p-value < 0.05).

Several vegetated areas containing GDV species that were visited during field surveys and observed to have experienced significant health impacts were assessed against the MK test results (Figure 20). These areas included: cleared GDV for mine infrastructure (Figure 20a); historically dead *M. argentea* individuals (*c.* 10 year old death) possibly resulting from nearby water abstraction (Figure 20b); significant GDV health increase from water discharge and pumping (Figure 20c); and significant *Eucalyptus camaldulensis*/*victrix* death resulting from drainage obstruction (Figure 20d).

In each example presented in Figure 20, the MK trend result detected a trend that correlated with ground-truthed observations relatively well. The trend, vegetation graphing and change point detection results presented are derived from the GDV Tool. The area at Figure 20a shows a haul road that intersected riparian vegetation containing *M. argentea,* represented as the red Tau declining trend area. Examination of Google Earth aerial imagery shows the haul road was constructed between December 2012 and 2013; a significant decline in MAVI values can be seen in early 2013 on the corresponding graph (Figure 20a).

An area of *M. argentea* mortality believed to have occurred approximately 10 years ago was also detected by a seasonal MK test undertaken over the period 2000-2018, and is presented as red Tau areas in Figure 20b. Five health regimes can be observed from the area; average greenness values peaked over period 1990 to mid-1994; a period of decline from mid-1994 to mid-2007 is observed, followed by a significantly sharp decline at the end of 2007. The decline period continues until early 2011, when a small increase in greenness is registered until late 2013, in which a gradual decline regime begins until the latest image (Figure 20b).

The influence of discharge and pumping on vegetation health was also observed. Vegetation containing *M. argentea* shown in Figure 20c was healthy and dense due to on-going water discharge into the drainage system. Increasing greenness, as presented as dark green areas surrounding the drainage line, appears to continuously increase from 1990 to mid-2012, when greenness peaks and then plateaus until now; late-2012 is detected as a new health regime.

The area of significant historical *E. camaldulensis*/*victrix* tree death observed at the final monitoring area is presented in Figure 20d. The MK trend results detected this area as a significant decline area (Figure 20d), which occurred around mid-2007 based on the associated vegetation health signature (Figure 20d).

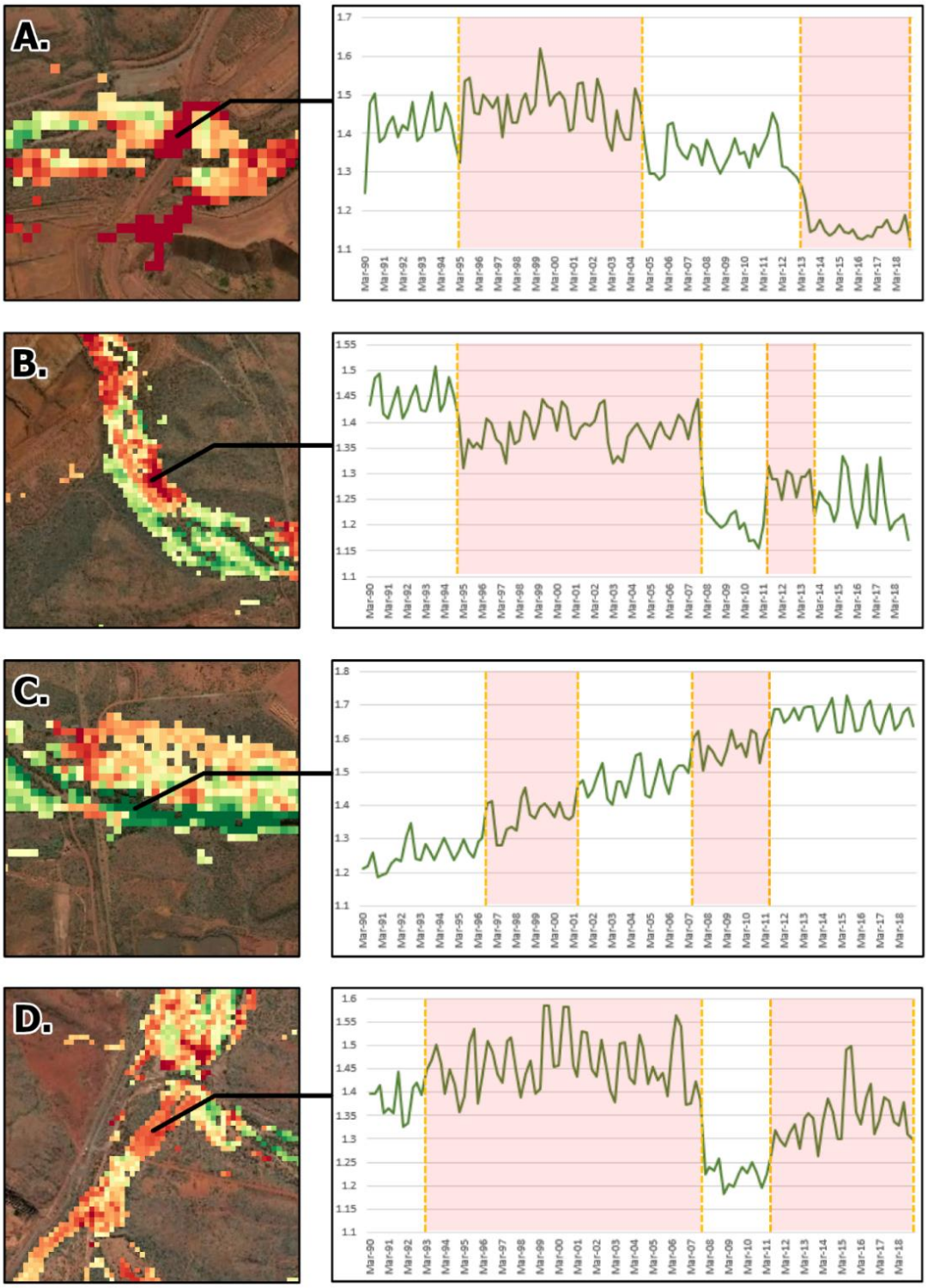


Figure 20: Overview of Mann-Kendall seasonal trend areas showing: a) haul road placed across vegetation and creek between 2012-2013; b) ground-truthed *Melaleuca argentea* open woodland that has experienced heavy mortality over last 10 years; c) ground-truthed discharge area; and d) ground-truthed *Eucalyptus camaldulensis*/*victrix* open woodland with numerous dead individuals.

4.2.2 Change Detection

The four impacted areas outlined above in Section 4.2.1 (Figure 20) were revisited using the change vector analysis (CVA) approach provided by the GDV Tool. A baseline image was generated using the median of all available dry-season (September to November) Landsat-derived TCT greenness and brightness (i.e. soil) images in the DEA ODC: this ranged from 1986-2018. A range of comparison dry-season greenness and soil images for years 1990 – 2018 was generated, and a CVA was performed for every year within this range and compared to the same baseline. The CVA angle ranges of 90º-180º and 270º-360º were applied via the GDV Tool, returning change maps showing vegetation decline and vegetation increase, respectively (Vorovencii 2014). Finally, very low magnitude pixels resulting from factors such as noise or imperfect normalisation was removed using a threshold value of 2 standard deviations (Baker et al., 2007; Seto et al., 2002).

The CVA results of the four impacted GDV areas visited during the field surveys are presented in Figure 21 and Figure 22. Figure 21a suggests an area of moderate to high magnitude vegetation decline is first observed in the dry season of 2012, which continues to increase in both magnitude and spatial extent from 2013 to 2014 (and on-going). This change is reflected in the associated graph (Figure 21a); greenness appears to decline sharply while bare soil (brightness) increases sharply late 2012, reflecting the removal of vegetation and exposure of bare soil observed at the haul road site. The change point detection break also highlights that late 2012 was a date of significant greenness and soil change. The CVA result reflects the MK trend observed prior (see Figure 20).

The CVA approach also detected the *M. argentea* impact site (Figure 21b). Areas of low to low-moderate vegetation decline magnitude are first observed in the dry season 2007, followed by significant increase in magnitude and spatial extent in 2008 and 2009 (and on-going). This change is reflected in the associated graph also; greenness sharply declines, and soil sharply increases end of 2007, suggesting significant vegetation loss at the area.

CVA was less consistent in detecting on-going vegetation increase at the observed discharge area (Figure 22a) in comparison to the MK trend (Figure 20c). Nevertheless, areas of vegetation increase were detected at the area, with a break detected occurring at the end of 2012 where vegetation increase began to plateau (see Figure 20c and Figure 22a graphs).

The *E. camaldulensis*/*victrix* decline area shown in Figure 22b was clear using the CVA approach. Early signs of low magnitude vegetation decline were first detected in the dry season of 2006; by dry season 2007, significantly high vegetation decline magnitude was apparent and continuous through 2008 and beyond. This change was detected occurring late 2007 by change point detection (see Figure 22b graph), where a significantly sharp decline in greenness occurred simultaneously with soil increase over the period 2007-2008.

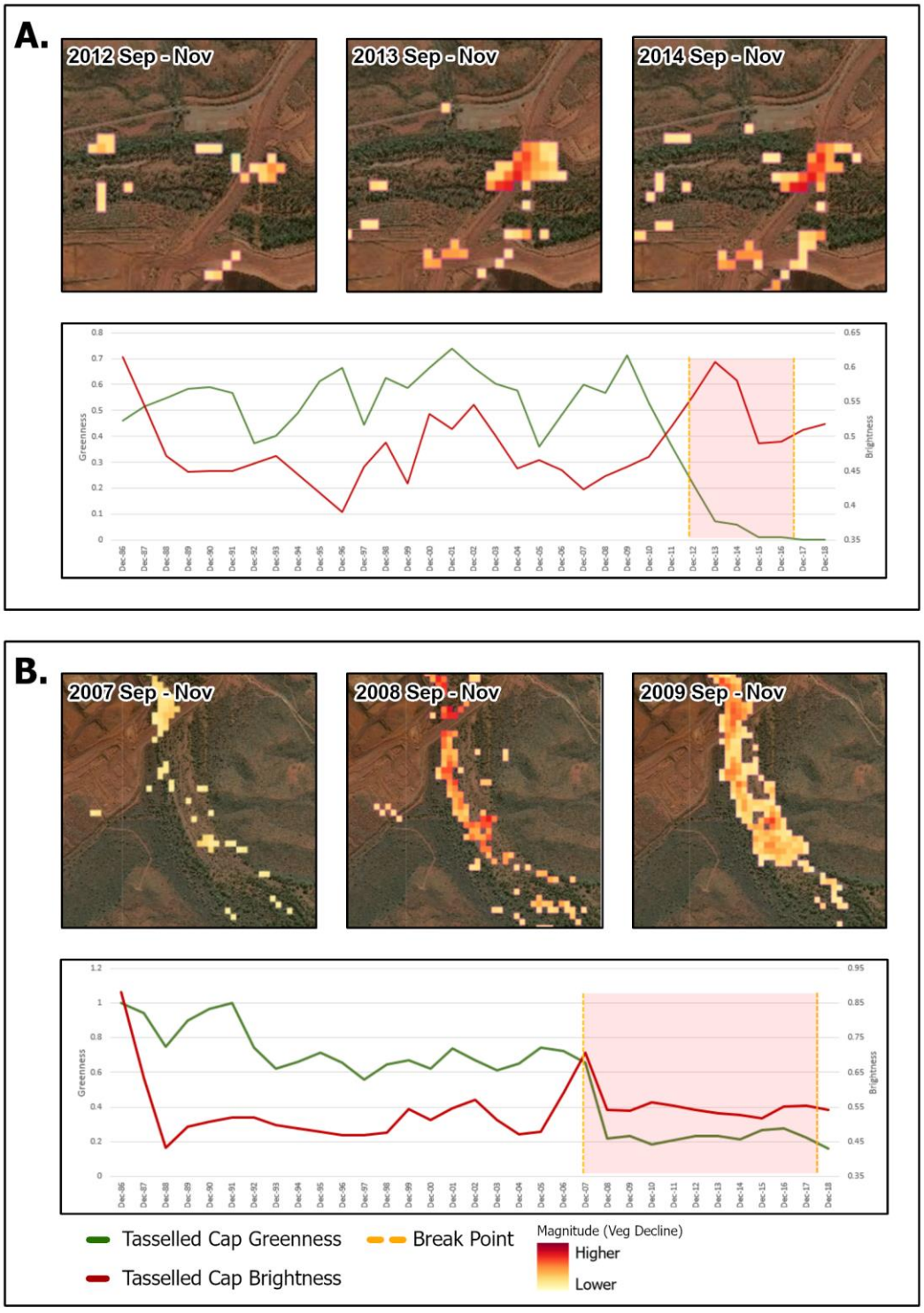


Figure 21: Overview of Change Vector Analysis vegetation decline (change angle = 90^{o}-180^{o}) areas showing: a) haul road placed across vegetation and creek between 2012-2013; and b) ground-truthed *Melaleuca argentea* open woodland that has experienced heavy mortality over last 10 years.

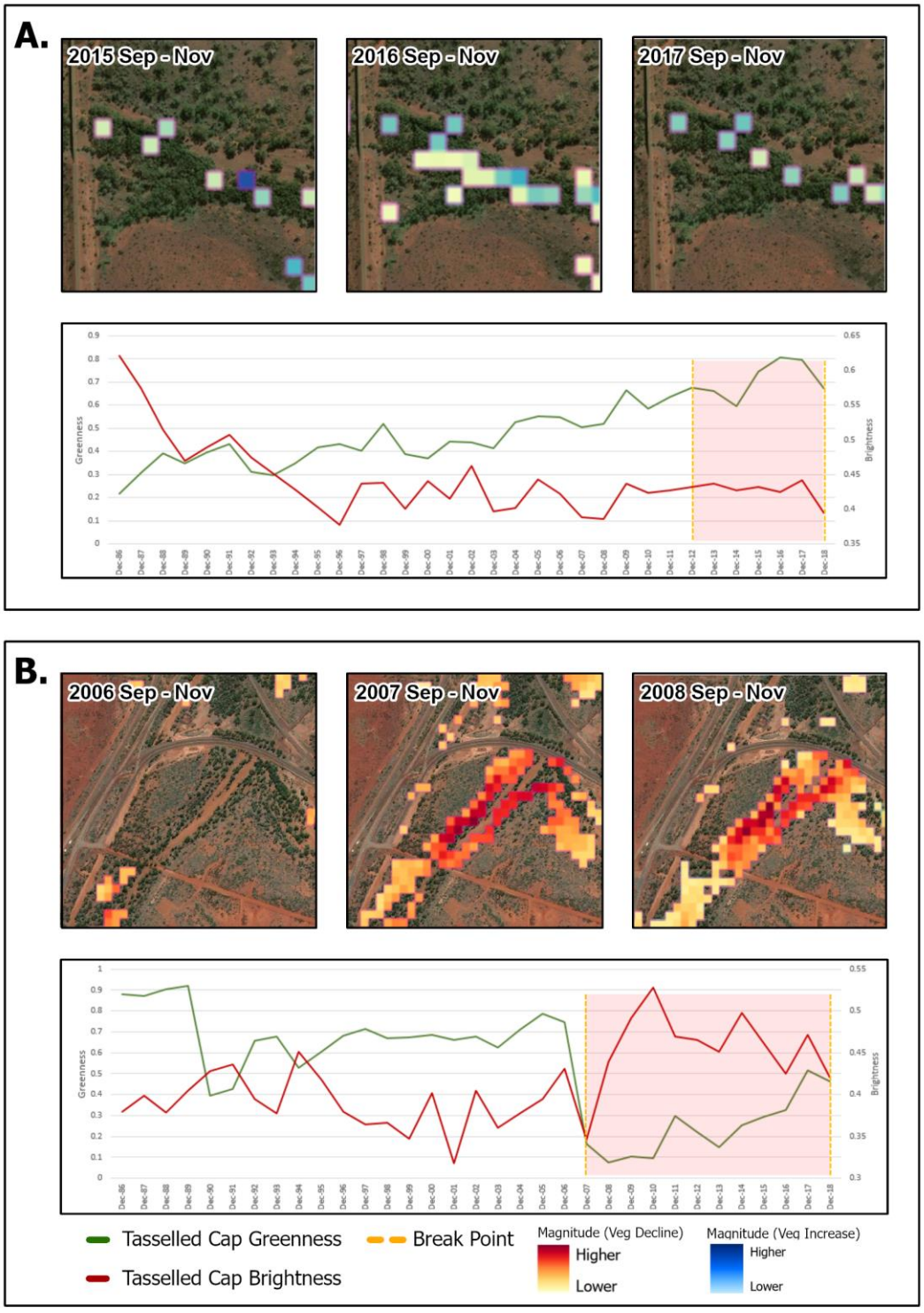


Figure 22: Overview of Change Vector Analysis vegetation increase (change angle = 270º-360º) and decline (change angle = 90º-180º) areas showing: a) ground-truthed discharge area; and d) ground-truthed *Eucalyptus camaldulensis*/*victrix* open woodland with numerous dead individuals.

5 Conclusions

GDV models based on spatial multicriteria evaluation with spectral inputs have proven to have excellent potential for detecting groundwater dependent vegetation in three study areas within the Pilbara bioregion. These models require no upfront training data and so can be created quickly and easily to reduce the search space for GDV even in unexplored territory. Our findings based on three study areas in the southeast of the Pilbara bioregion identified the moisture adjusted vegetation index (MAVI) to outperform more commonly used indices including NDVI and SAVI irrespective of platform (Landsat or Sentinel). For optimal model performance, it is the recommended index for future models and time series analysis. We do not recommend the use of MSVI-1/2/3 due to poor performance in these areas. The TCT Greenness index was consistently in the top three indices irrespective of platform or study area. As change vector analysis requires brightness and greenness to operate, we suggest the continued use of TCT images for this purpose. Furthermore, the addition of brightness to greenness in the temporal graphs adds further details for interpretation.

Area under the curve statistics improved with higher resolution imagery. For example, MAVI averaged an AUC of 0.852 using Landsat but this rose to about 0.949 using Sentinel imagery. This would therefore appear to be the optimal analysis ready dataset to use, but we note that it still lacks a sufficient historical archive for trend analysis as well as for determining benchmarks that pre-date mine disturbance or at least can be estimated from large tracts of vegetation that are relatively undisturbed. Furthermore, we are not aware of a Sentinel program that will continue should Sentinel A/B reach the end of their seven-year expected missions. In contrast, Landsat imagery will continue to be collected, and at higher temporal resolution, with the launch of Landsat 9 later this year (2020). In summary, we see value in both platforms that can be tapped. Therefore, we chose to allow the user to pick their platform based on their intended analyses; both platforms run in parallel.

We tested the Mann-Kendall Tau statistic for trend analysis using observations of decline and increase observed in the field. Our results were able to pick up an area of dead *M. argentea* that were still standing and possibly the result of water abstraction as well as dead E. *camaldulensis*/*victrix,* in locations that were historically more mesic. We also identified health and density increases in an area receiving water discharge. These results were suitable for historical reconstruction and are useful to explore the long-term trend. However, the statistic requires at least five years of data to create meaningful trends and, by itself, is not a great tool for early warning intervention purposes, particularly in areas with highly dynamic water tables. We recommend further research in the detection of near real time change.

To augment the trend analysis, change vector analysis was introduced. This tool allows for the creation of a long-term benchmark from historical imagery to be compared to present day imagery to identify a magnitude of change that is interpretable relative to four quadrants: vegetation loss, moisture loss, moisture gain, vegetation gain. This worked well for all areas of known loss but was less effective at observable gains. Originally, this was surprising, but it may be that the use of an earlier benchmark when the landscape was undisturbed, and vegetation was historically greener, magnifies the losses and supresses the gains to some degree. This also needs further research for clarity.

6 References

Alaibakhsh, M., Emelyanova, I., Barron, O., Khiadani, M. and Warren, G., 2017. Large-scale regional delineation of riparian vegetation in the arid and semi-arid Pilbara region, WA. Hydrological Processes, 31(24), pp.4269-4281.

Alan Tingay and Associates, 1997. Yandi Iron Ore Mine Marillana Creek Environmental Study 1991-97. Review of Study Design. Report No: 97/28.

Andersen, T., Carstensen, J., Hernandez-Garcia, E. and Duarte, C.M., 2009. Ecological thresholds and regime shifts: approaches to identification. Trends in Ecology & Evolution, 24(1), pp.49-57.

Andrews, M.E. and Warrener, H., 2017. Detecting micro-refugia in semi-arid landscapes from remotely sensed vegetation dynamics. Remote Sensing of Environment, 200, pp.114-124.

Assal, T.J., Anderson, P.J. and Sibold, J., 2016. Spatial and temporal trends of drought effects in a heterogeneous semi-arid forest ecosystem. Forest Ecology and Management, 365, pp.137-151.

Astron, 2011. Marillana Creek: Survey of Riparian Tree Health and Reconnaissance of Additional RVMP Sites. Environmental Survey Report.

Baig, M.H.A., Zhang, L., Shuai, T. and Tong, Q., 2014. Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance. Remote Sensing Letters, 5(5), pp.423-431.

Baker, C., Lawrence, R.L., Montagne, C. and Patten, D., 2007. Change detection of wetland ecosystems using Landsat imagery and change vector analysis. Wetlands, 27(3), p.610.

Barron, O.V., Emelyanova, I., Van Niel, T.G., Pollock, D. and Hodgson, G., 2014. Mapping groundwater-dependent ecosystems using remote sensing measures of vegetation and moisture dynamics. Hydrological Processes, 28(2), pp.372-385.

BHP, 2007. Marillana Creek Tree Health Physical Monitoring December 2007. BHPBIO Environment Section Report.

Boer, M.M., Macfarlane, C., Norris, J., Sadler, R.J., Wallace, J. and Grierson, P.F., 2008. Mapping burned areas and burn severity patterns in SW Australian eucalypt forest using remotely-sensed changes in leaf area index. Remote Sensing of Environment, 112(12), pp.4358-4369.

Crist, E.P., 1985. A TM tasseled cap equivalent transformation for reflectance factor data. Remote Sensing of Environment, 17(3), pp.301-306.

Curry, P., Zdunic, K., Wallace, J. and Law, J., 2008. Landsat monitoring of woodland regeneration in degraded mulga rangelands: implications for arid landscapes managed for carbon sequestration. 14th ARSPC, Sept 29-Oct2, Darwin.

David, H.A., 1963. The method of paired comparisons (Vol. 12). London.

Draper, N.R. and Smith, H., 1998. *Applied regression analysis* (Vol. 326). John Wiley & Sons.

Eastman, J.R., 1999. Multi-criteria evaluation and GIS.*Geographical information systems*, *1*(1), pp.493-502.

Eastman, J.R., Sangermano, F., Machado, E.A., Rogan, J. and Anyamba, A., 2013. Global trends in seasonality of normalized difference vegetation index (NDVI), 1982–2011. Remote Sensing, 5(10), pp.4799-4818.

Ecologia, 2009. Roy Hill 1 Project: Flora and Vegetation Assessment. Report for Hancock Propsescting Pty. Ltd.

Emelyanova, I., Barron, O. and Alaibakhsh, M., 2018. A comparative evaluation of arid inflow-dependent vegetation maps derived from LANDSAT top-of-atmosphere and surface reflectances. *International journal of remote sensing*, *39*(20), pp.6607-6630.

EPA, 2019. Roy Hill Iron Ore Pty Ltd – Revised Proposal for the Roy Hill Iron Ore Mine. Environmental Review Document.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern recognition letters, 27(8), pp.861-874.

Flood, N., 2013. Seasonal composite Landsat TM/ETM+ images using the medoid (a multi-dimensional median). *Remote Sensing*, *5*(12), pp.6481-6500.

Flores, S.E. and Yool, S.R., 2007. Sensitivity of change vector analysis to land cover change in an arid ecosystem. International Journal of Remote Sensing, 28(5), pp.1069-1088.

Fielding, A.H. and Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation, 24(1), pp.38-49.

Furby, S.L. and Campbell, N.A., 2001. Calibrating images from different dates to 'like-value'digital counts. Remote Sensing of Environment, 77(2), pp.186-196.

Giuliani, G., Chatenoux, B., De Bono, A., Rodila, D., Richard, J.P., Allenbach, K., Dao, H. and Peduzzi, P., 2017. Building an earth observations data cube: Lessons learned from the Swiss data cube (SDC) on generating analysis ready data (ARD). Big Earth Data, 1(1-2), pp.100-117.

Graetz, R.D., Pech, R.P. and Davis, A.W., 1988. The assessment and monitoring of sparsely vegetated rangelands using calibrated Landsat data. International Journal of Remote Sensing, 9(7), pp.1201-1222.

Guerschman, J.P., Van Dijk, A.I.J.M., Mattersdorf, G., Beringer, J., Hutley, L.B., Leuning, R., Pipunic, R.C., Sherman, B.S. (2009) Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia, Journal of Hydrology, 369 (1-2), 107-119.

Hamed, K.H., 2008. Trend detection in hydrologic data: the Mann–Kendall trend test under the scaling hypothesis. Journal of hydrology, 349(3-4), pp.350-363.

Hazaymeh, K., 2016. Development of a Remote Sensing-Based Agriculture Monitoring Drought Index and Its Application Over a Semi-Arid Region. University of Calgary Press.

Hipel, K.W. and McLeod, A.I., 1994. Time series modelling of water resources and environmental systems. Elsevier.

Hirsch, R.M., Slack, J.R. and Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. Water resources research, 18(1), pp.107-121.

Homer, C.G., Aldridge, C.L., Meyer, D.K. and Schell, S.J., 2012. Multi-scale remote sensing sagebrush characterization with regression trees over Wyoming, USA: laying a foundation for monitoring. International Journal of Applied Earth Observation and Geoinformation, 14(1), pp.233-244.

Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013. Applied logistic regression (Vol. 398). John Wiley & Sons.

Huang, C., Wylie, B., Yang, L., Homer, C. and Zylstra, G., 2002. Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. International journal of remote sensing, 23(8), pp.1741-1748.

Huete, A., 1988. Huete, AR A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment. Remote sensing of environment, 25, pp.295-309.

Hussain, M. and Mahmud, I., 2019. pyMannKendall: a python package for non-parametric Mann Kendall family of trend tests. Journal of Open Source Software, 4(39), p.1556.

Jafari, R., Lewis, M.M. and Ostendorf, B., 2007. Evaluation of vegetation indices for assessing vegetation cover in southern arid lands in South Australia. The Rangeland Journal, 29(1), pp.39-49.

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B. and Wang, X., 2014. Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote sensing, 6(11), pp.11518-11532.

Johnson, R.D. and Kasischke, E.S., 1998. Change vector analysis: A technique for the multispectral monitoring of land cover and condition. International Journal of Remote Sensing, 19(3), pp.411-426.

Kauth, R.J. and Thomas, G.S., 1976, January. The tasselled cap--a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In LARS symposia (p. 159).

Keeley, J.E., 2009. Fire intensity, fire severity and burn severity: a brief review and suggested usage. *International Journal of Wildland Fire*, *18*(1), pp.116-126.

Landmann, T., Schramm, M., Huettich, C. and Dech, S., 2013. MODIS-based change vector analysis for assessing wetland dynamics in Southern Africa. Remote sensing letters, 4(2), pp.104-113.

Lehmann, E.A., Wallace, J.F., Caccetta, P.A., Furby, S.L. and Zdunic, K., 2013. Forest cover trends from time series Landsat data for the Australian continent. International Journal of Applied Earth Observation and Geoinformation, 21, pp.453-462.

Lymburner, L., Beggs, P.J. and Jacobson, C.R. (2000) Estimation of Canopy-Average Surface-Specific Leaf Area Using Landsat TM Data, Photogrammetric Engineering and Remote Sensing, Vol. 66 (2), pp. 183-191.

Malak, D.A. and Pausas, J.G., 2006. Fire regime and post-fire Normalized Difference Vegetation Index changes in the eastern Iberian Peninsula (Mediterranean basin). International Journal of Wildland Fire, 15(3), pp.407-413.

Malila, W.A., 1980, January. Change vector analysis: an approach for detecting forest changes with Landsat. In LARS symposia (p. 385).

Mann, H.B., 1945. Nonparametric tests against trend. Econometrica: Journal of the Econometric Society, pp.245-259.

Nackaerts, K., Vaesen, K., Muys, B. and Coppin, P., 2005. Comparative performance of a modified change vector analysis in forest change detection. International Journal of Remote Sensing, 26(5), pp.839-852.

Nedkov, R., 2017. Orthogonal transformation of segmented images from the satellite Sentinel-2. Comptes rendus de l'Academie bulgare des Sciences, 70(5), pp.687-692.

O'Grady, A.P., Carter, J.L. and Bruce, J., 2011. Can we predict groundwater discharge from terrestrial ecosystems using existing eco-hydrological concepts? Hydrology & Earth System Sciences, 15(12).

O'Neill, A.L., 1996. Satellite-derived vegetation indices applied to semi-arid shrublands in Australia. The Australian Geographer, 27(2), pp.185-199.

Palmer, A.R. and van Rooyen, A.F., 1998. Detecting vegetation change in the southern Kalahari using Landsat TM data. Journal of Arid Environments, 39(2), pp.143-153.

Pauer, F., Schmidt, K., Babac, A., Damm, K., Frank, M. and von der Schulenburg, J.M.G., 2016. Comparison of different approaches applied in analytic hierarchy process–An example of information needs of patients with rare diseases. BMC medical informatics and decision making, 16(1), p.117.

Payne, A L, and Mitchell, A A., 1999. An assessment of the impact of Ophthalmia Dam on the floodplains of the Fortescue River on Ethel Creek and Roy Hill Stations. Department of Agriculture and Food, Western Australia, 124, pp.66.

Qiu, S., Lin, Y., Shang, R., Zhang, J., Ma, L. and Zhu, Z., 2019. Making Landsat time series consistent: evaluating and improving Landsat analysis ready data. Remote Sensing, 11(1), p.51.

Robertson, M.P., Villet, M.H. and Palmer, A.R., 2004. A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects. Diversity and Distributions, 10(5-6), pp.461-474.

Robinson, T., Novelly, P., Watson, I., Corner, R., Thomas, P., Schut, T., Jansen, S. and Shepherd, D., 2009. Towards an approach for remote sensing-based rangeland condition assessment in north Western Australia. In Proceedings of the Surveying and Spatial Sciences Institute Biennial International Conference (pp. 899-913). Surveying and Spatial Sciences Institute.

Robinson, T.P., van Klinken, R.D. and Metternicht, G., 2010. Comparison of alternative strategies for invasive species distribution modeling. Ecological Modelling, 221(19), pp.2261-2269.

Robinson T.P., Novelly P.E., Corner R., Thomas P.A., Russell-Brown A. 2012. Pastoral Lease Assessment Using Geospatial Analysis, Department of Agriculture and Food.

Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication, 351, p.309.

Roy, D.P., Zhang, H.K., Ju, J., Gomez-Dans, J.L., Lewis, P.E., Schaaf, C.B., Sun, Q., Li, J., Huang, H. and Kovalskyy, V., 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. Remote Sensing of Environment, 176, pp.255-271.

Saaty, R.W., 1987. The analytic hierarchy process—what it is and how it is used. Mathematical modelling, 9(3-5), pp.161-176. McGraw-Hill, New York.

Seto, K.C., Woodcock, C.E., Song, C., Huang, X., Lu, J. and Kaufmann, R.K., 2002. Monitoring land-use change in the Pearl River Delta using Landsat TM. International Journal of Remote Sensing, 23(10), pp.1985-2004.

Thackway, R. & Cresswell, I.D. (1995). An Interim Biogeographic Regionalisation for Australia: A framework for setting priorities in the National Reserves System Cooperative Program, Version 4.0. Australian Nature Conservation Agency, Canberra.

Thomas F.M. (2014) Ecology of Phreatophytes. In: Lüttge U., Beyschlag W., Cushman J. (eds) Progress in Botany. Progress in Botany (Genetics - Physiology - Systematics - Ecology), 75, Springer, Berlin, Heidelberg.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8, pp.127-150.

Truong, C., Oudre, L. and Vayatis, N., 2019. Selective review of offline change point detection methods. *Signal Processing*, p.107299.

Van Vreeswyk, A.M.E., Leighton, K.A., Payne, A.L. and Hennig, P., 2004. An inventory and condition survey of the Pilbara region, Western Australia.

Vorovencii, I., 2014. A change vector analysis technique for monitoring land cover changes in Copsa Mica, Romania, in the period 1985–2011. Environmental monitoring and assessment, 186(9), pp.5951-5968.

Wallace, J.F. and Thomas, P.W.E., 1998. Rangeland monitoring in northern Western Australia using sequences of Landsat imagery. Report to National Landcare Program Project No. 953024. Agriculture Western Australia, unpublished.

Wallace, J.F., Caccetta, P.A. and Kiiveri, H.T., 2004. Recent developments in analysis of spatial and temporal data for landscape qualities and monitoring. Austral Ecology, 29(1), pp.100-107.

Washington-Allen, R.A., West, N.E., Ramsey, R.D. and Efroymson, R.A., 2006. A protocol for retrospective remote sensing–based ecological monitoring of rangelands. Rangeland Ecology & Management, 59(1), pp.19-29.

Wambui, G.D., Waititu, G.A. and Wanjoya, A., 2015. The power of the pruned exact linear time (PELT) test in multiple changepoint detection. American Journal of Theoretical and Applied Statistics, 4(6), pp.581-586.

Zhu G, Ju W, Chen JM, Liu Y (2014) A Novel Moisture Adjusted Vegetation Index (MAVI) to Reduce Background Reflectance and Topographical Effects on LAI Retrieval. PLoS ONE 9(7): e102560

Zweig, M.H. and Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry, 39(4), pp.561-577.

Appendix A: Canopy condition ranking guide

>= 96% of live canopy (rank 5)

66% - 95% live canopy (rank 4)

36% - 65% live canopy (rank 3)

6% - 35% live canopy (rank 2)

<= 5% live canopy (rank 1)

Figure A1: Canopy condition ranking guide used during field surveys.

Appendix B: GDV Tool workshop material

PART 1

DIGITAL EARTH AUSTRALIA TRAINING WORKSHOP

DEA SANDBOX SIGN-UP, ACCESS AND JUPYTER INTRODUCTION

Authors: Caitlin Adams, Jess Keysers, Alex Leith, Robbi Bishop-Taylor and Claire Phillips

Modified by: Lewis Trotter

Introduction

This workshop will introduce working with Digital Earth Australia (DEA) data in the DEA Sandbox environment for the Open Data Cube (ODC). The workshop is broken into the following sections:

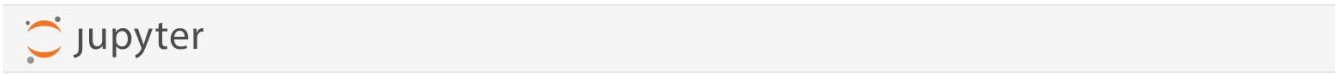
1. Getting started – access the sandbox
2. Learning Jupyter – explore what a Jupyter Notebook is

At the end of the workshop you will know how to use a Jupyter Notebook in conjunction with the ODC to access and analyse Earth observation data. The workshop should take around an hour to complete.

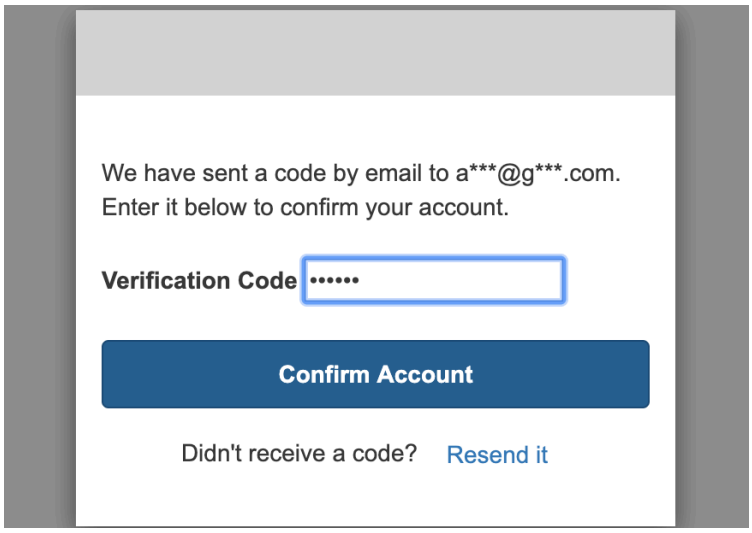
Getting started

Sign up for a DEA Sandbox Account

The DEA Sandbox uses requires you to create an account to log in. Please visit https://app.sandbox.dea.ga.gov.au/ to sign up for a new account, or log in if you already have one.

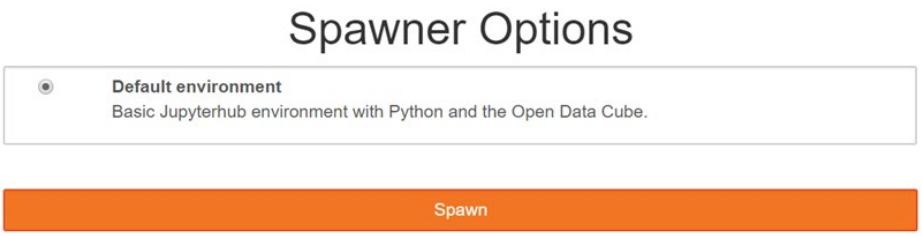


You'll need to be able to access the email you register with in order to proceed.



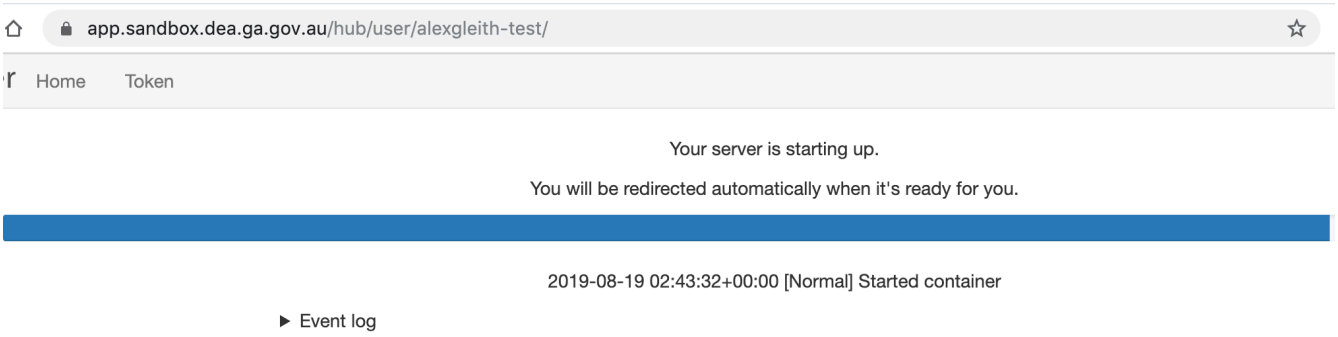
Spawn the DEA Sandbox

After confirming the account, spawn the DEA Sandbox by clicking 'Spawn'.

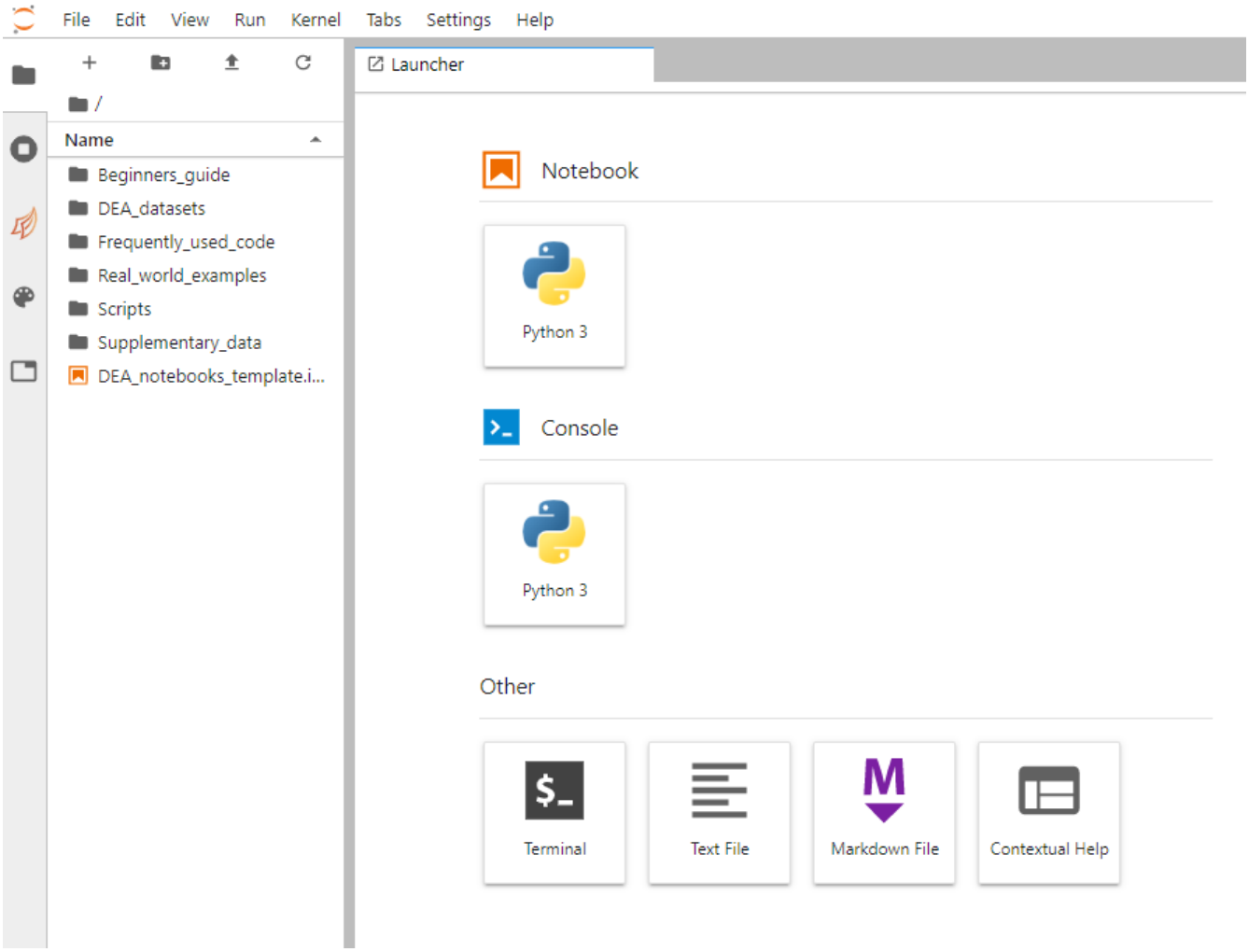


Accessing the DEA Sandbox

After signing into the DEA Sandbox, your Jupyter environment will be created and you should see a loading screen (also shown below) while the system is working to prepare the environment.



Once signed in, the JupyterLab homepage should appear (see below).


File
Edit
View
Run
Kernel
Tabs
Settings
Help
Launcher
/
Name
Beginners_guide
DEA_datasets
Frequently_used_code
Real_world_examples
Scripts
Supplementary_data
DEA_notebooks_template.i...
Notebook
Python 3
Console
Python 3
Other
Terminal
Text File
Markdown File
Contextual Help

Learning Jupyter

Overview

Jupyter is an interactive coding environment. The name 'Jupyter' comes from Julia, Python and R, which are all programming languages that are used in scientific computing.

Jupyter started as a purely Python-based environment, called iPython, but there has been rapid progress over the last few years, and now many large organisations like Netflix[1] are using the system to analyse data.

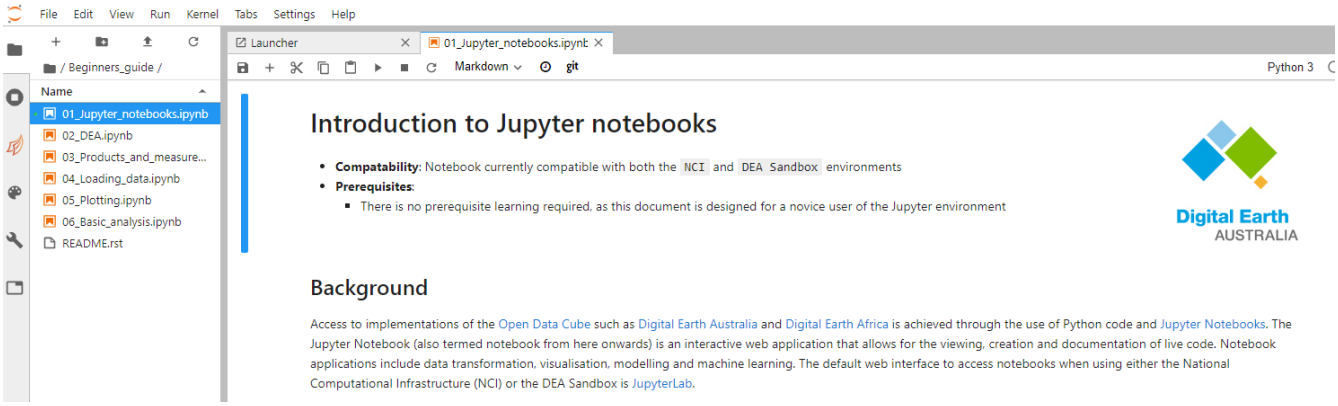
Since the ODC is a Python library, the workshop will cover working with Earth observation data in Python-based notebooks.

Explore a basic notebook

The first exercise is to explore a very basic notebook. The goal is to understand the key features of notebooks.

> *If you've used Jupyter before, you may want to skip this step.*

Open the 'Beginners_guide' folder, and then open the file named '01_Jupyter_notebooks.ipynb'.



To learn about the notebook, read through the text in the notebook and run each code cell step (using 'shift + enter' or the play button ▶).

Feel free to change the code sections to explore how it works.

> *Note that when the section to the side of a cell is showing an asterisk, that means it is running. This is most important when running a data load that may take more than a few seconds*

The next image displays an executed cell.

[1] https://medium.com/netflix-techblog/notebook-innovation-591ee3221233

Run the cell below:

```
[2]: print("I ran a cell!")
```

```
I ran a cell!
```

Cell status

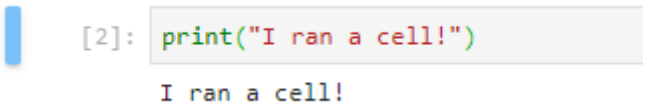
The `[ ]:` symbol to the left of each Code cell describes the state of the cell:

- `[ ]:` means that the cell has not been run yet.
- `[*]:` means that the cell is currently running.
- `[1]:` means that the cell has finished running and was the first cell run.

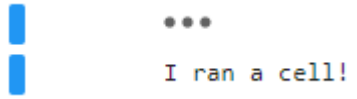
Code cells can also be hidden and revealed.

This is useful for keeping notebooks presentable to the user.

To hide a code cell such as the one displayed above, hover your mouse to the left of the code cell and a blue bar should appear – click that blue bar once.



The raw Python code will be hidden and replaced with a '…' button show below.



Simply click the '…' button again to reveal the code.

Continue on with the remainder of the notebook to learn about different types of code cells.

References

- The Open Data Cube website: https://www.opendatacube.org/
- Geoscience Australia (including DEA) metadata: http://cmi.ga.gov.au/
- DEA data on Amazon Web Services' S3: http://dea-public-data.s3-ap-southeast-2.amazonaws.com/index.html
- National Map to view data: https://nationalmap.gov.au/

OPEN DATA CUBE

Digital Earth AUSTRALIA

Australian Government
Geoscience Australia

FRONTIER SI

PART 2

GROUNDWATER DEPENDENT VEGETATION TOOL WORKSHOP

DETECT POTENTIAL GDV AND EXPLORE GDV HEALTH TRENDS/CHANGE OVER TIME

Authors: Lewis Trotter, Todd Robinson, Paula Fiévez

Curtin University

Introduction

This workshop will walk you through the Groundwater Dependent vegetation (GDV) detection and monitoring tool (hereafter referred to as the GDV Tool).

Here you will learn how to detect potential GDV in a particular study area, calculate vegetation health trends and explore vegetation changes across time, and determine dates in which significant breaks in vegetation health occurred

This tool has been developed using the Jupyter Notebook open-source web application and makes considerable use of Digital Earth Australia (DEA) data in the DEA Sandbox environment for the Open Data Cube (ODC).

This workshop is broken into the following sections:

1. Getting started: prepare your DEA Sandbox and download GDV Tool code;
2. Section 00: initialise tool code and modules;
3. Section 01: generate GDV likelihood areas
4. Section 02: monitor GDV health trends via Mann-Kendall trend analysis
5. Section 03: monitor GDV change via spectral change vector analysis

At the end of the workshop you will know how to use every main feature offered by the GDV Tool 1-2 hours to complete.

Getting started

Prerequisites

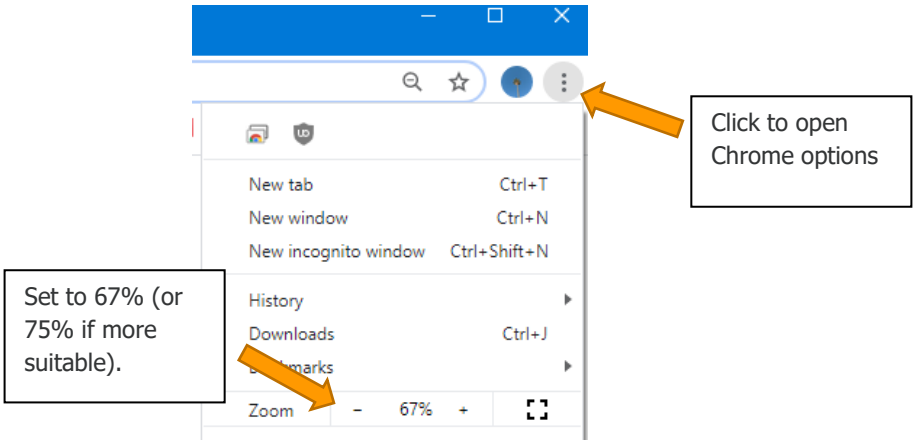
This workshop requires that you have signed up for access to the DEA Sandbox account. If you have not, please read the associated DEA training material (part 1 document) to set up a new DEA Sandbox account to obtain access.

It is also recommended that a stable internet connection is available and you are using the Google Chrome browser (https://www.google.com/chrome/).

Chrome settings

The GDV Tool was developed with medium to large sized screens in mind. Smaller screens can make some of the Jupyter control panels difficult to use. It is recommended, if on a smaller laptop screen, to use the Chrome 'zoom out' ability to provide Jupyter with more screen space.

In the top right corner of Chrome browser, click the three dots and set the zoom to 67%. If this is too zoomed out, try 75%.



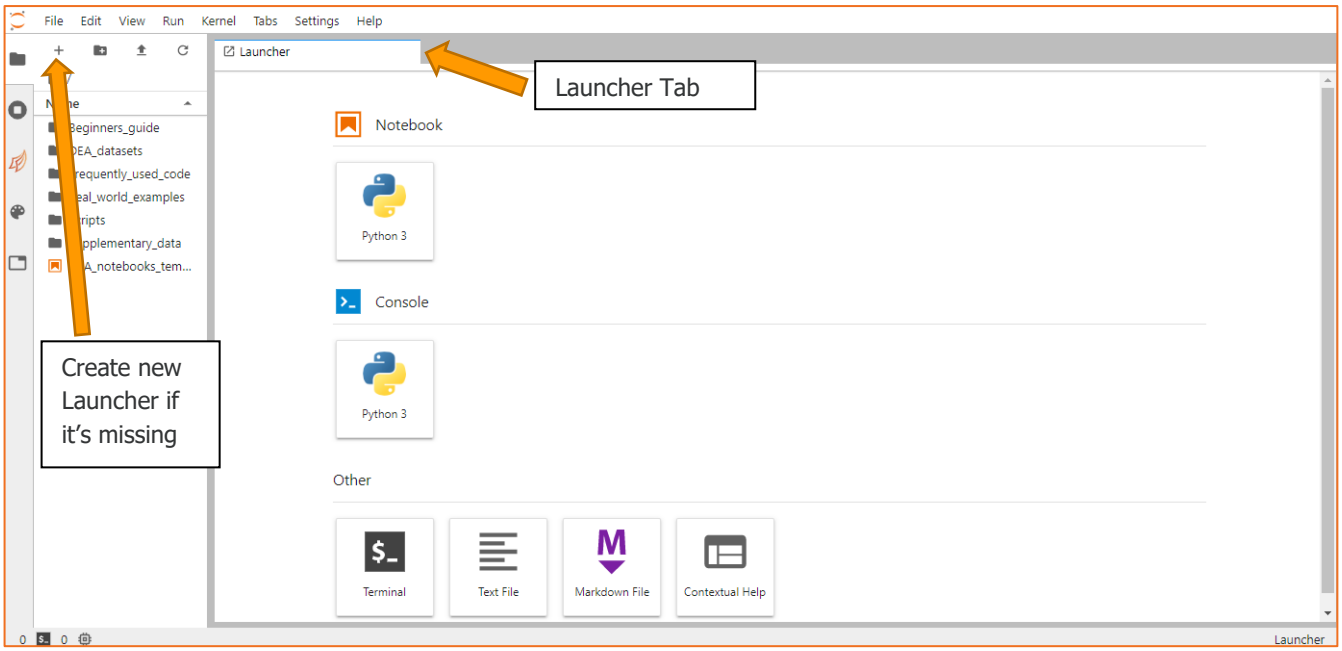
Downloading the GDV Tool

Note: You will only ever have to do this once, or whenever there is a version change.

At this stage, the GDV Tool has not been added to your DEA Sandbox environment.

The GDV Tool source code is fully available at 'GitHub'. Although you can view that here: https://github.com/lewistrotter/GDV_TOOL, we will use the command line to do this for us next.

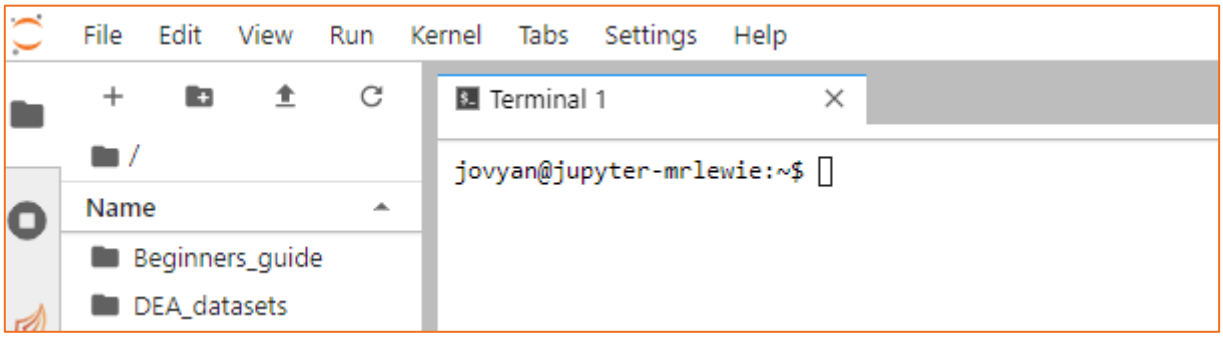
Go to your Launcher tab, shown below. If you don't have one, create a new one first (also shown below).



You will need to start a new terminal session to communicate with GitHub and download the code.

Click the 'Terminal' icon in the launcher page (at the bottom - it has a dollar sign on it).

A prompt similar to below will appear – you should see your username positioned after 'joyvan@jupyter'.



Click once within the terminal page area and type in:

```
cd
```

Press enter to issue the command. If nothing happens, that's fine – proceed.

On the next line, type in:

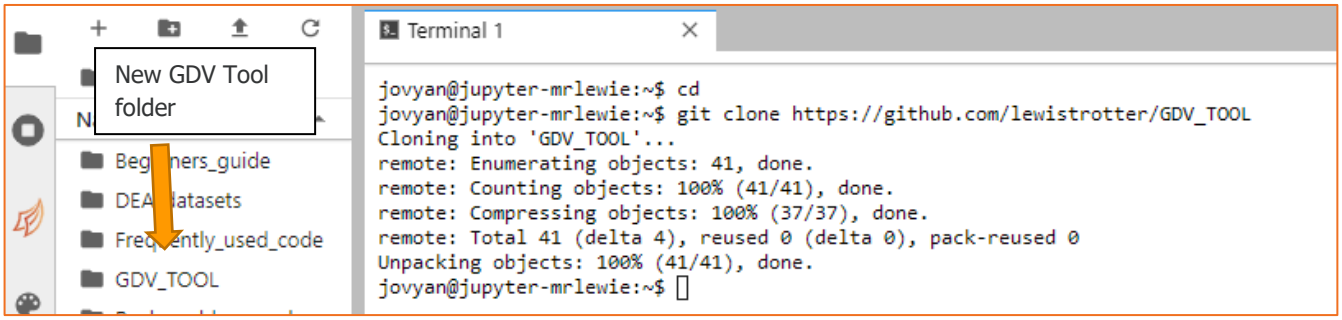
```
git clone https://github.com/lewistrotter/GDV_TOOL
```

Press enter again.

You should see some information about cloning and unpacking.

This is good; the GDV Tool is being downloaded and cloned to your Sandbox.

Eventually a new folder called 'GDV_TOOL' will be created in the 'Left Sidebar' area of your screen, as presented below:



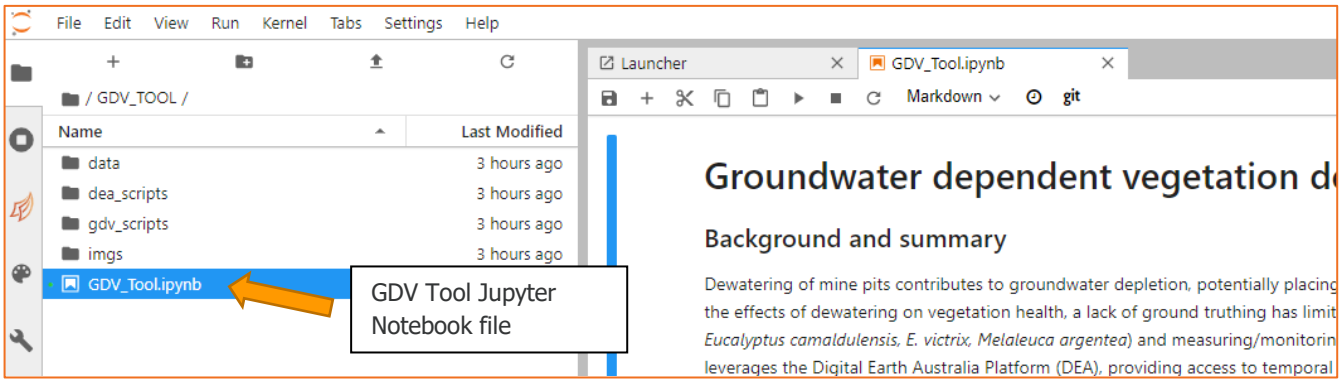
Double click on the GDV_TOOL folder to open its contents.

Starting the GDV Tool

Within the GDV Tool folder are numerous subfolders containing scripts and files needed by the tool. We can ignore these.

Most important is the Jupyter Notebook file named 'GDV_Tool.ipynb'. This file essentially brings everything together into a user-friendly interface.

Double click on the 'GDV_Tool.ipynb' file shown below to initialise the GDV Tool.



A new page should load in a new tab to the right. This is where we will spend our time during this workshop.

The tool itself contains a lot of helpful notes and information regarding every aspect of the tool.

We will summarise most of this in this workshop document, but feel free to read the tool notes as we go.

GDV Tool sections

The GDV Tool is separated into four main sections to provide a clear sense of workflow:

- Section 00: Initialise code – loads required modules, scripts, files, storage, etc. used by the tool;
- Section 01: Generate GDV likelihood layers - uses satellite images and spatial modelling techniques to detect areas in your study area that are likely (and unlikely) to be GDV;
- Section 02: Perform Mann-Kendall trend analyses – detect and explore trends (e.g. increasing/decreasing) in GDV health across five or more years for any season using several Mann-Kendall trend analyses;

Section 03: Perform spectral change vector analysis – detect and explore vegetation change for any season of any date relative to a long-term seasonal median baseline.

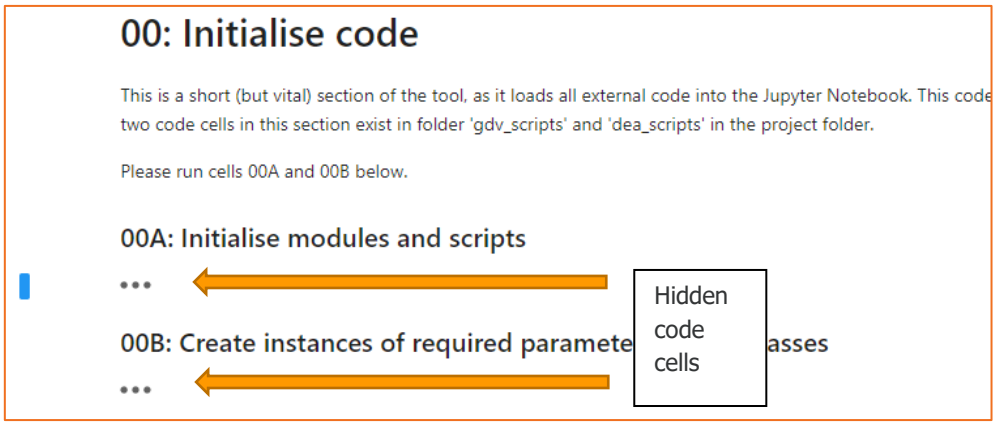
Note: Section 00 and Section 01 must be run first. Following that, users can then use either Section 02 or Section 03 in any order.

Section 00: Initialise code

This section of the tool is a short but vital one; it is used to initialise all external code into Jupyter.

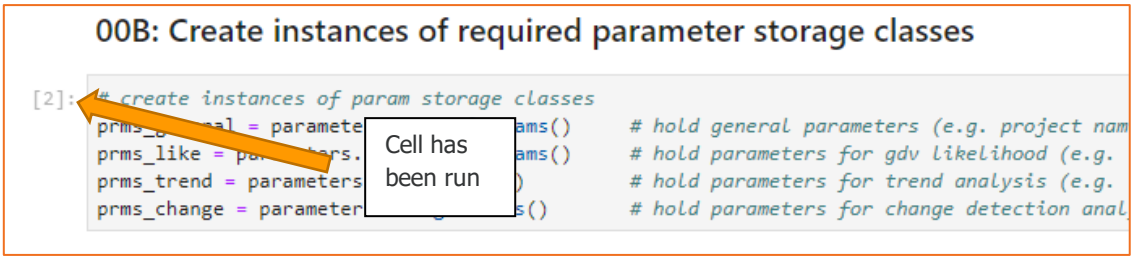
If the code cells here are not run, the GDV Tool will not work.

You should see two steps (00A and 00B) with hidden code underneath each one, like so:



Run both of these code cells, starting with 00A, by either pressing 'shift + enter' or the ▶ button at the top of the page.

When a code cell has been successfully run, it will show a number to the left of the code cell, like so:



It is worth checking your code cells have a number next to them before proceeding to ensure the code has actually be run.

After you've checked the code cells have been run, you can hide the cell again by clicking the blue line to the left of the cell.

Section 01: Generate GDV likelihood layers

The first section of this tool involves the use of spatial multi-criteria evaluation modelling techniques to differentiate GDV from non-GDV to an acceptable level of accuracy using freely available remote sensing satellite imagery.

It is hypothesised that GDV can be differentiated from non-GDV on satellite imagery by combining metrics of greenness, leaf moisture and stability with the construction of a weighting scheme disproportionally in favour of the dry season (September-November) vs the wet season (January-March).

The key processing steps applied in this section are as follows:

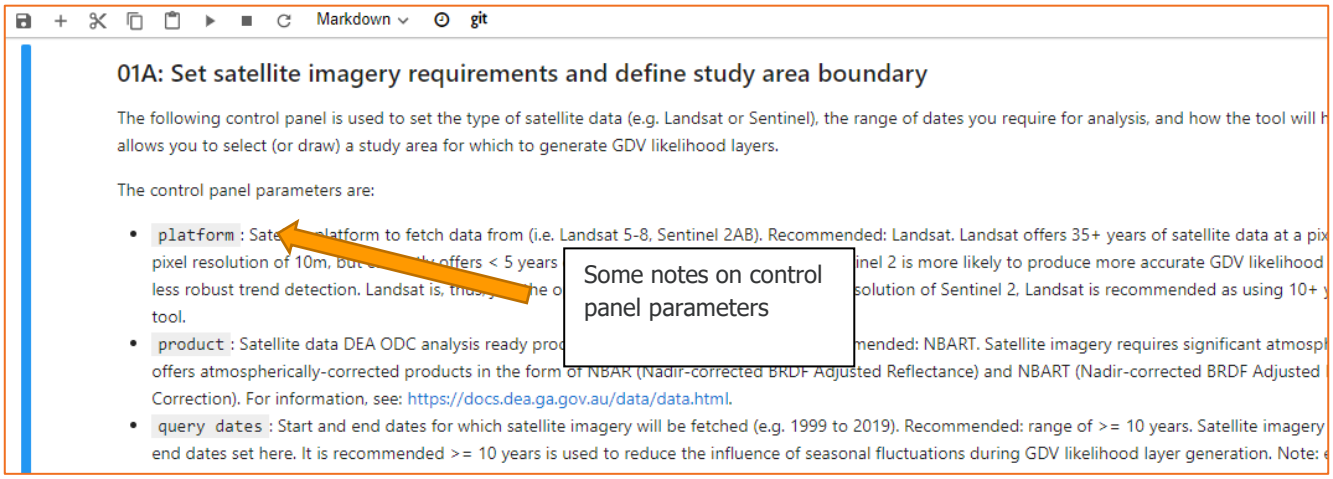
1. Fetch a 'cube' of Landsat or Sentinel satellite data for a user-defined study area and timespan using the DEA ODC;
2. Generate vegetation and moisture indices based on user-selection;
3. Reduce data to wet (Jan-Mar) and dry (Sep-Nov) annual medians and interpolate missing data (optional);
4. Standardise vegetation and moisture using an invariant target site and orthogonal polynomial coefficients approach;
5. Re-scale values using fuzzy sigmoidal functions;
6. Generate vegetation and moisture stability for each year using wet and dry seasons;
7. Produce a GDV likelihood layers via Analytical Hierarchy Process (AHP) Pairwise Comparison weighting approach;
8. Slice GDV likelihood layers into GDV/non-GDV areas using ground-truth point data (optional) or basic standard deviation;
9. Export results as GDV likelihood layers (median GDV likelihood of all time or per year in data 'cube').

When ready, move to step 01A to begin.

Step 01A: Set satellite imagery requirements and define study area boundary

This step involves setting satellite platform (e.g. Landsat or Sentinel), analysis date range, and satellite image cloud handling. It also involves selecting (or drawing) a study area for which to perform analysis within.

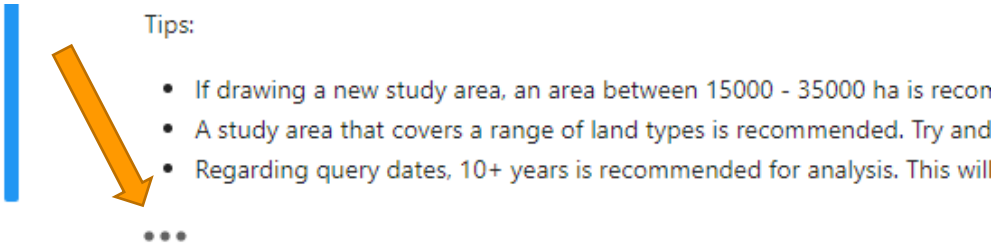
Go to Step 01A in your Jupyter Notebook and you should see some notes, like so:



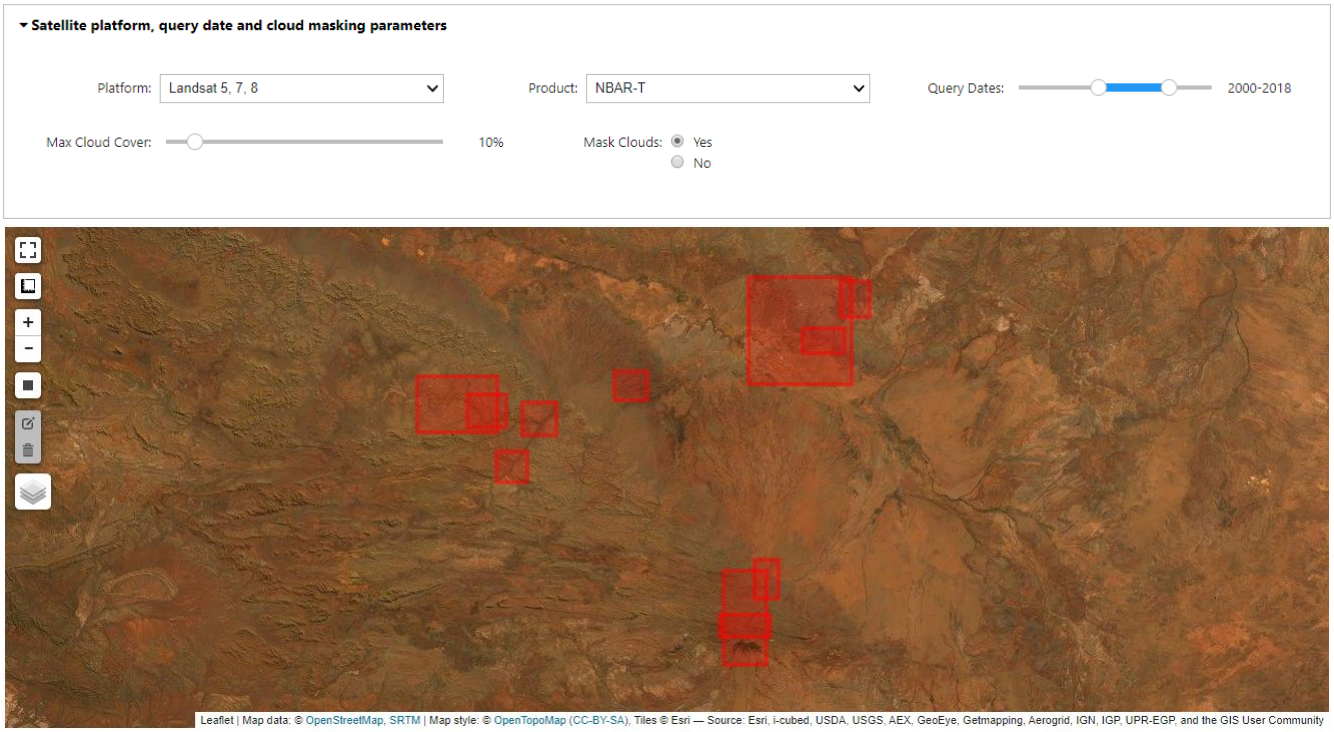
Important notes are embedded throughout the GDV Tool interface. These notes are there to explain various analysis options, provide background information, and/or provide tips on obtaining optimal analysis results.

Before you read over the notes, it might help to reveal the control panel and interactive map that the notes are referring to.

Run the code cell located just underneath the note block, as shown below:



The resulting control panel and interactive map should appear:



Now that we can see these panels, the notes mentioned earlier will make more sense.

It is recommended you read over the aforementioned notes at the start of Step 01A to get an understanding of the control panel and interactive map parameters.

When you are ready to move on, set the control panel parameters to match the following:

- Platform: Landsat 5, 7, 8
- Product: NBART
- Query Dates: 2000-2018
- Max Cloud Cover: 10%
- Mask Clouds: Yes

!!! Warning! Do not re-run control panel code cell after you've selected your options – they will be reset !!!

Parameter explanation:

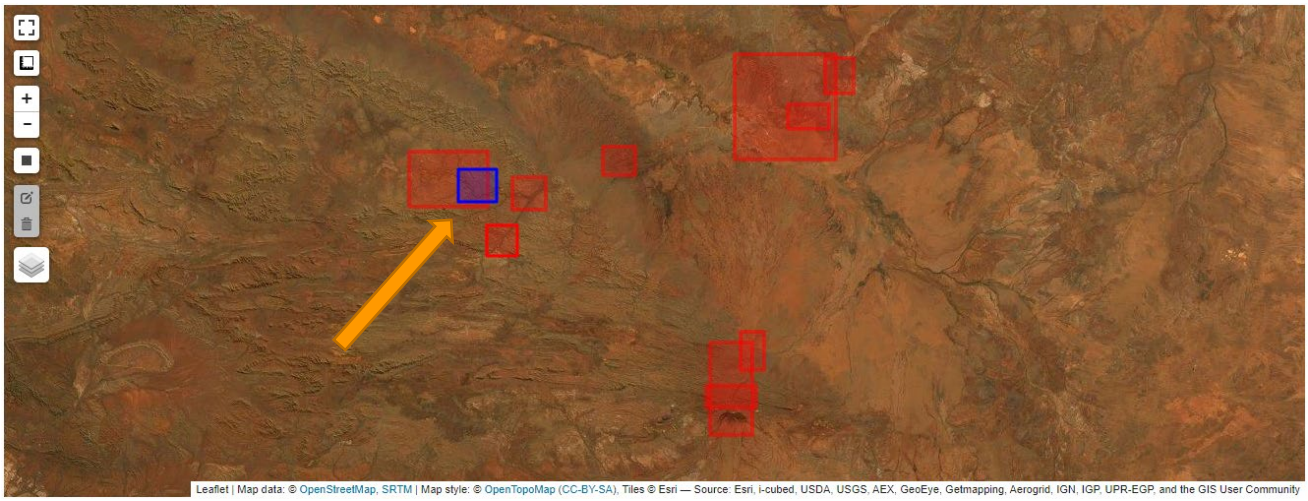
Here, you have opted to use Landsat 5, 7 and 8 satellite imagery as the data behind analysis. You have also chosen to use the NBART (Nadir-corrected BRDF Adjusted Reflectance with Terrain Illumination Reflectance Correction) analysis ready data product offered by DEA. This product is satellite imagery that has been corrected for atmospheric and terrain distortions (more here: https://docs.dea.ga.gov.au/data/data.html).

Additionally, your analysis will be undertaken from the year 2000 to 2018 (inclusive), and any satellite images with more than 10% visible cloud cover will be excluded from analysis. You've also opted to mask out any cloud pixels detected in the satellite image that do happen to exist.

Next, we need to select a study area to perform analysis within.

We will use a subset of the one of the study areas that the research team visited in late 2019. This trip was undertaken to collect field data for model validation purposes. Later, you will use some of this field data to validate our model accuracy within this tool itself.

On the interactive map, click on the study area boundary shown below so that it turns blue:



The blue highlight indicates we've successfully selected a study area.

When ready, proceed to step 01B.

Step 01B: Fetch satellite imagery from Open Data Cube

We will now fetch all available Landsat satellite imagery for the options set on the above control panel and map. This data will be fetched from the DEA ODC.

Digital Earth Australia (DEA) is a digital platform that catalogues large amounts of Earth observation data covering Australia. It is underpinned by the Open Data Cube (ODC), which is an open source software package that provides a Python-based API for high performance querying and data access (see https://www.ga.gov.au/dea/odc).

Quickly read the tool notes (particularly the dot points) to get an understanding of what this step is doing behind the scenes.

When ready, simply run the code cell beneath the tool notes to begin the fetching process. This process may take several minutes to complete.

When the process is finished, you should see a summary output similar to below:

```
Beginning satellite data fetch. This can take awhile.

=== Your Chosen Parameters ===
Platform: ls.
Product: nbart.
Date Range: 2000-01-01 to 2018-12-31.
Cloud Cover: 10.0.
Mask Clouds: True.
NW Corner: (-1311179.195452772, -2513202.72392482) | SE Corner (-1299406.1343468376, -2521332.647970429).
Input EPSG: EPSG:3577 | Output EPSG EPSG:3577.
=== Your Chosen Parameters ===

Loading ga_ls5t_ard_3 data
    Filtering to 103 out of 130 observations
    Applying pixel quality mask
Loading ga_ls7e_ard_3 data
    Ignoring SLC-off observations for ls7
    Filtering to 50 out of 404 observations
    Applying pixel quality mask
Loading ga_ls8c_ard_3 data
    Filtering to 100 out of 127 observations
    Applying pixel quality mask
Combining and sorting data
    Masking out invalid values
    Returning 253 observations as a dask array

Successfully fetched Landsat satellite data!
Renaming Landsat bands for clarity.
Doing some remaining dataset clean up (e.g. dropping mask var, normalising time, sorting)
Reducing dataset down into wet (JFM) and dry (SON) seasons only.

Disabling above dashboard and map. Please re-run the tool if you require a new study area.
```

1

2

3

Result explanation:

This output reiterates what our selected ODC query parameters were (1). It also provides a helpful summary of exactly how many satellite images were loaded for each satellite platform (ls5t, ls7e and ls8c represent Landsat platforms 5, 7 and 8, respectively) (2). Seems we have 253 satellite images in total. You can also see that satellite bands were renamed, some basic clean-up was performed, and the dataset was reduced to wet (Jan-Mar) and dry (Sep-Nov) seasons to reduce memory issues (3).

When ready, proceed to step 01C.

Step 01C: Set vegetation and moisture indices and options

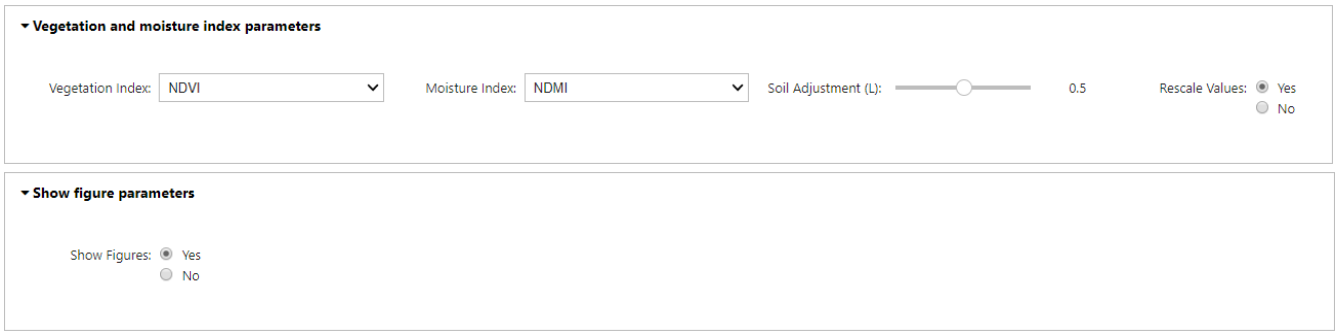
Now that we have obtained all available Landsat data for our study area, the next step is to derive vegetation and moisture indices from their spectral bands. These indices are used in the spatial multi-criteria evaluation modelling techniques implemented throughout this tool.

As remote sensing of vegetation in semi-arid regions like the Pilbara is challenged by significant soil background reflectance, choice of vegetation index can vary the result and accuracy of the GDV likelihood modelling outputs.

A suite of remote sensing indices was tested during the development of this tool, and those found to produce optimal results across multiple tests sites have been provided to users. Based on these preliminary findings, SLAVI, MAVI, Tasselled Cap Greenness and NDVI produced optimal GDV likelihood models.

Run the code cell directly beneath the step 01C notes to reveal the control panel, then spend some time reading the tool notes for an overview of the available vegetation and moisture indices and control panel parameter explanations.

The control panel should appear as below:



When you are ready to move on, set the control panel parameters to match the following:

- Vegetation index: NDVI
- Moisture index: NDMI
- Soil Adjustment (L): 0.5 *(only relevant for use with SAVI)*
- Rescale values: Yes
- Show figures: Yes

Parameter explanation:

Based on these choices, you have opted to use Normalised Difference vegetation Index (NDVI) and Normalised Difference Moisture Index (NDMI). The soil adjustment value is not considered when using any index other than SAVI, so we can leave this as default. We have also chosen to rescale index values – this is done to simplify some processing behind the scenes and is highly recommended. Finally, we have opted to show figures instead of hide them once processing is done.

When ready, proceed to step 01D.

Step 01D: Generate vegetation and moisture indices

We will now generate NDVI and NDMI using the spectral bands embedded in the ODC satellite images.

Read the tool notes (particularly the dot points) to get an understanding of what this step is doing behind the scenes.

When ready, simply run the code cell beneath the tool notes to begin generating vegetation and moisture indices. This process may take several minutes to complete.

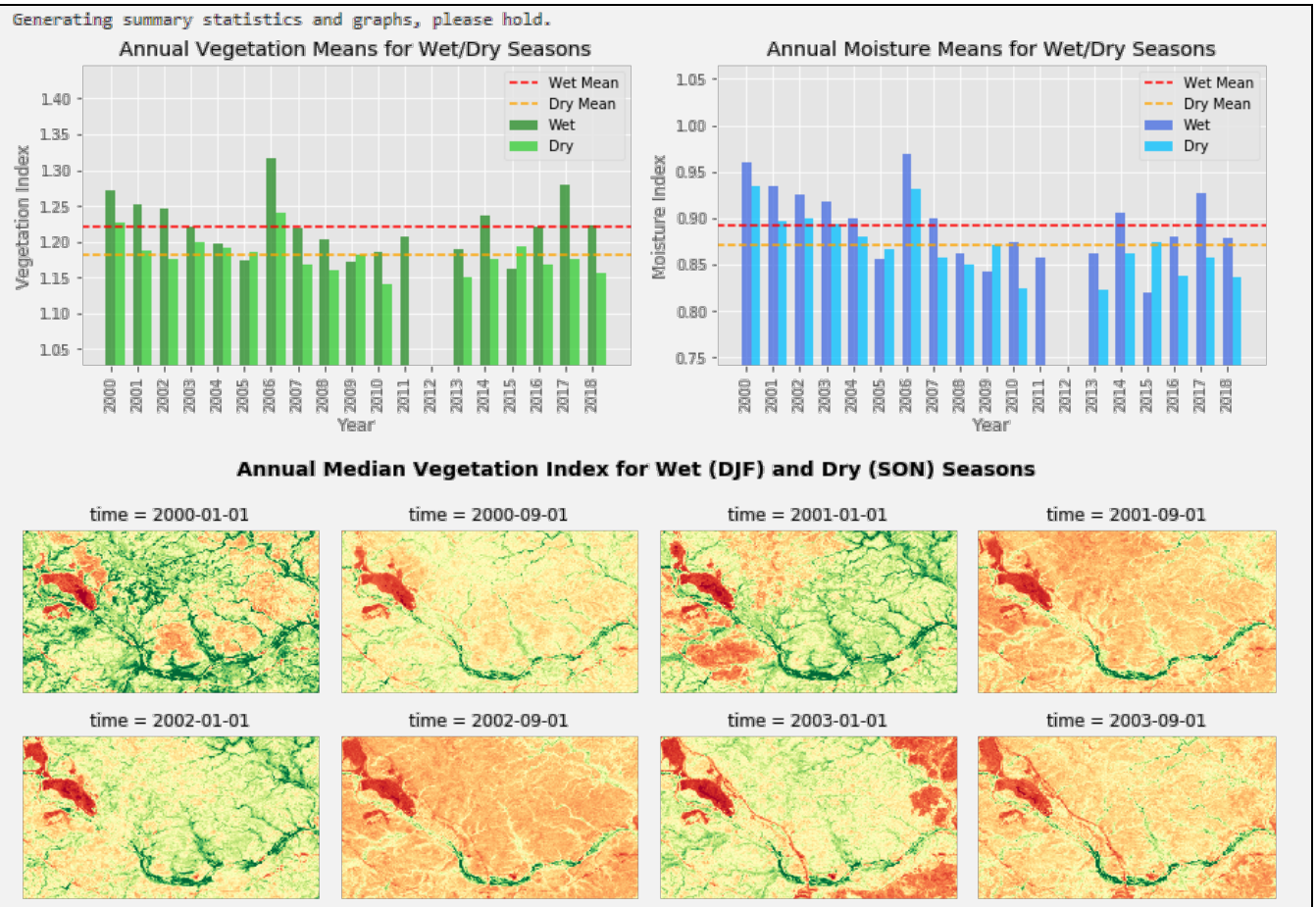
Eventually you should see an output similar to below:

```
Calculating vegetation index: ndvi. Please wait.
Successfully generated vegetation index: ndvi. Proceeding.
Calculating moisture index: ndmi. Please wait.
Successfully generated moisture index: ndmi. Proceeding.
Dropping original blue, green, red, nir, swir1, swir2 bands.

Resampling data down to annual wet and dry seasonal medians.
Computing dask during reduction. This can take quite awhile. Please wait.
Computated dataset successfully! Proceeding.

Disabling above dashboard. Please re-run the tool if you require a different index.

=== Summary dataset statistics ===
> Total number of dates in dataset: 38
> Total number of years in dataset: 19        1
> Number of wet (DJF) dates in dataset: 19
> Number of dry (SON) dates in dataset: 19
> Years with missing season(s): []
> Years with extra season(s): []
> Completely missing vegetation data on dates: ['2011-09-01', '2012-01-01', '2012-09-01']   2
> Completely missing moisture data on dates: ['2011-09-01', '2012-01-01', '2012-09-01']
> Partially missing vegetation data on dates: []   3
> Partially missing moisture data on dates: []
> Mean values for entire wet season: vege: 1.221, moist: 0.893
> Mean values for entire dry season: vege: 1.181, moist: 0.871    4
> Median values for entire wet season: vege: 1.214, moist: 0.907
> Median values for entire dry season: vege: 1.177, moist: 0.885
> Standard deviation values for entire wet season: vege: 0.084, moist: 0.118
> Standard deviation values for entire dry season: vege: 0.06, moist: 0.106
=== Summary dataset statistics ===

Performing Z-score test to determine data outliers. Please wait.
=== Z-Score Results using score of 1.96 ===
> Vege Outlier Dates: ['2006-01-01', '2006-09-01']   5
> Moisture Outlier Dates: ['2000-09-01']
=== Z-Score Results using score of 1.96 ===
```

Result explanation:

From this output we can see there are 38 (19 wet season and 19 dry season) satellite images that have been returned from ODC from 19 different years (1). Vegetation and moisture images are completely missing for 2011-09-01, 2012-01-01 and 2012-09-01 (2). No images are missing any pixels (i.e. partially missing) – great (3). Summary statistics are provided as mean, median and standard deviations (4). Finally, a preliminary z-test result is presented (5). The z-test is used to detect any potential vegetation or moisture outlier images in the dataset to a significance of 0.05 (i.e. z-score of 1.96). Two vegetation outliers (2006-01-01 and 2006-09-01) and one moisture outlier (2000-09-01) were detected. You will be able to remove these later, if desired.

A graph of annual wet/dry season vegetation/moisture means should also appear, as should a figure of vegetation and moisture satellite images, as shown below:



It is worth taking some time to explore these outputs. They can highlight exceptionally wet or dry years, satellite image data gaps, and provide a visual of vegetation and moisture change over time within your study area.

Most gaps and outliers can be corrected in the next step of the tool.

When ready, proceed to step 01E.

Step 01E: Set vegetation and moisture index corrections and interpolation options

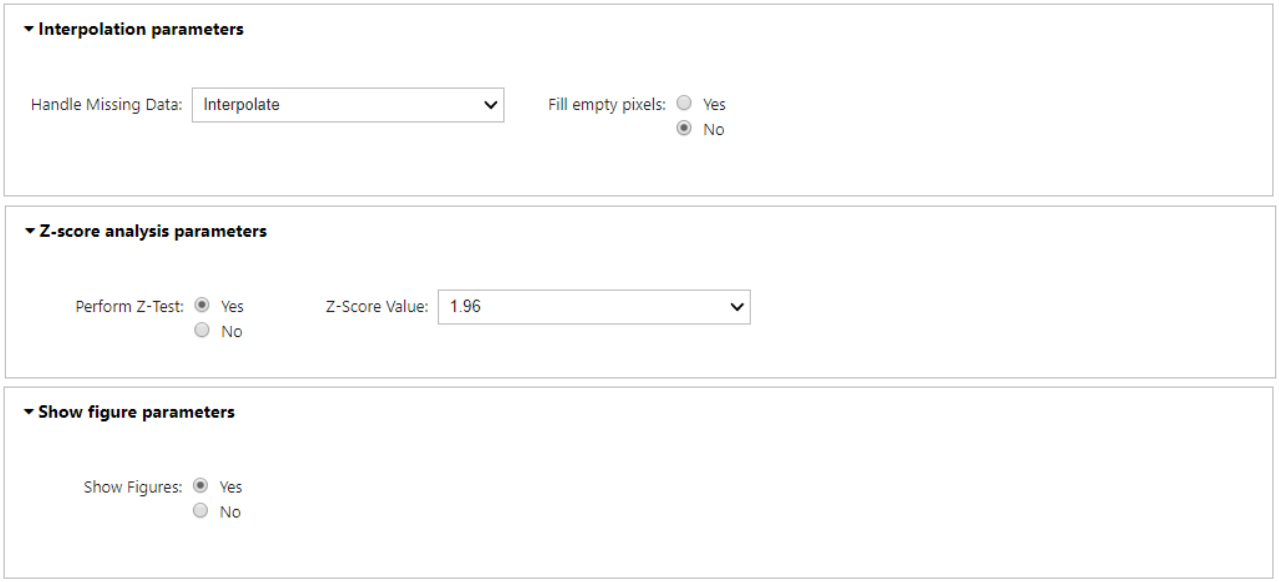
As you can see from the above figure, satellite imagery is often prone to data gaps (usually due to the presence of considerable cloud cover), missing pixels and occasional outliers (e.g. post-bushfire greening, climate events).

Some of these gap issues can be corrected through linear interpolation, and outliers can be detected and even removed using outlier testing (e.g. z-test, see: https://en.wikipedia.org/wiki/Z-test).

This step involves identifying some of these issues in our dataset and correcting them through statistical outlier removal and linear interpolation. Note: interpolation and outlier removal are completely optional in this tool (but recommended).

Run the code cell directly beneath the step 01E notes to reveal the control panel, then spend some time reading the tool notes for an overview of the available correction and interpolation methods and control panel parameters.

The control panel should appear as below:



When you are ready to move on, set the control panel parameters to match the following:

- Handle missing data: Interpolate
- Fill empty pixels: No
- Perform z-test: Yes
- Z-score value: 1.96
- Show figures: Yes

Parameter explanation:

Based on these choices, you have opted to interpolate missing vegetation and moisture images using linear interpolation. This should interpolate vegetation/moisture images for our missing dates (e.g. 2011-09-01, 2012-01-01 and 2012-09-01) using the image before and after within the associated season. We have decided not to fill in missing pixels (as we do not have any). Finally, we will perform a z-test with a z-score value of 1.96 (equivalent to a p-value < 0.05) to remove outliers that were highlighted in step 01D (i.e. 2000-09-01, 2006-01-01 and 2006-09-01).

When ready, proceed to step 01F.

Step 01F: Correct and interpolate vegetation and moisture indices

We will now correct and interpolate missing satellite imagery and outliers.

Again, have a quick read of the tool notes for step 01F and, when ready, run the code cell beneath the tool notes to begin correcting and interpolating potential issues and gaps.

This process shouldn't take long to complete. Eventually you should see an output similar to below:

```
Performing Z-score test to determine data outliers. Please wait.
=== Z-Score Results using score of 1.96 ===
> Vege Outlier Dates: ['2006-01-01', '2006-09-01']
> Moisture Outlier Dates: ['2000-09-01']
=== Z-Score Results using score of 1.96 ===

Checking if first year has missing seasons, and dropping if so. Please wait.
Sufficient data in first year: 2000. No data removed. Proceeding.
Checking if final year has missing seasons, and dropping if so. Please wait.
Sufficient data in final year: 2018. No data removed. Proceeding.

Checking for years with wet/dry season data gaps, will flag and remove if found. Please wait.
Following years contain wet/dry season gaps: [2000, 2006, 2011, 2012]
Beginning linear interpolation of missing data. Please wait.
Successfully interpolated missing data. Proceeding.

Disabling above dashboard. Please re-run the tool if you require a different index.

=== Summary dataset statistics ===
> Total number of dates in dataset: 38
> Total number of years in dataset: 19
> Number of wet (DJF) dates in dataset: 19
> Number of dry (SON) dates in dataset: 19
> Years with missing season(s): []
> Years with extra season(s): []
> Completely missing vegetation data on dates: []
> Completely missing moisture data on dates: []
> Partially missing vegetation data on dates: []
> Partially missing moisture data on dates: []
> Mean values for entire wet season: vege: 1.211, moist: 0.887
> Mean values for entire dry season: vege: 1.173, moist: 0.86
> Median values for entire wet season: vege: 1.204, moist: 0.901
> Median values for entire dry season: vege: 1.169, moist: 0.874
> Standard deviation values for entire wet season: vege: 0.08, moist: 0.115
> Standard deviation values for entire dry season: vege: 0.057, moist: 0.103
=== Summary dataset statistics ===
```

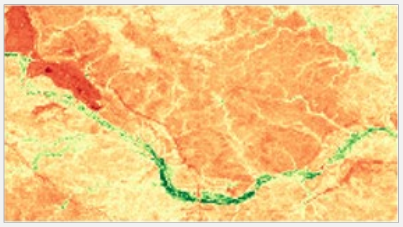
1
2
3

Result explanation:

The output above provides useful information about what happened to the satellite imagery after interpolation and outlier removal. The vegetation and moisture outliers mentioned prior (i.e. 2000-09-01, 2006-01-01 and 2006-09-01) were removed and re-interpolated (1). As an example:

Before outlier correction
time = 2006-01-01

After outlier correction
time = 2006-09-01



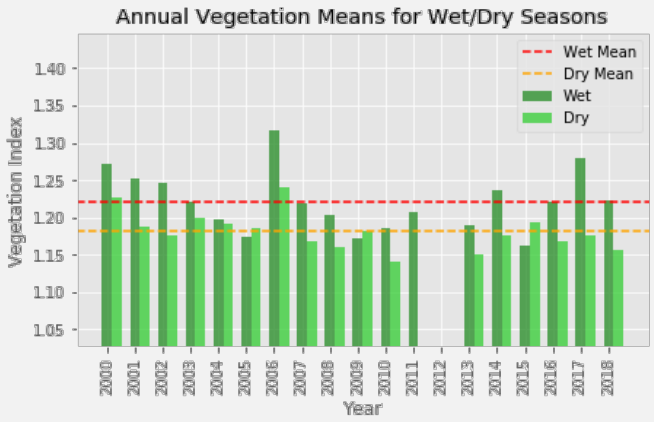
You can also see that starting and ending years of our dataset (2000, 2018) contained both wet and dry season images (2). This is important to note, as if one season is missing in either of these years, the whole year has to be dropped.

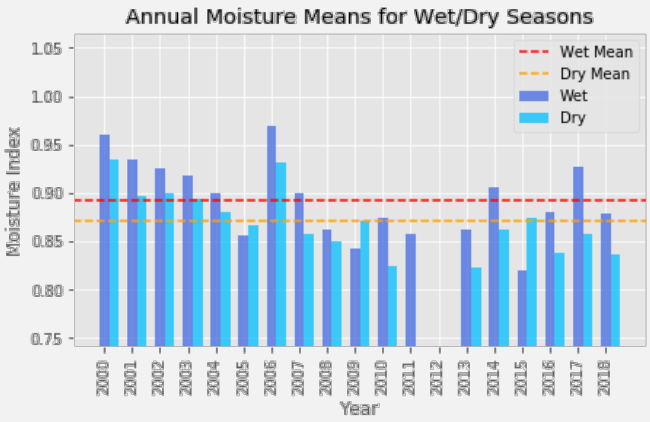
It also appears that years 2000, 2006, 2011 and 2012 are missing one or both wet and dry season images. These whole years were dropped and interpolated (3). As an example:

Finally, it can be helpful to compare your earlier vegetation and moisture graphs with the newly interpolated and corrected ones. Doing so provides a quick visual of how the corrections and interpolation processes have changed your vegetation and moisture images.

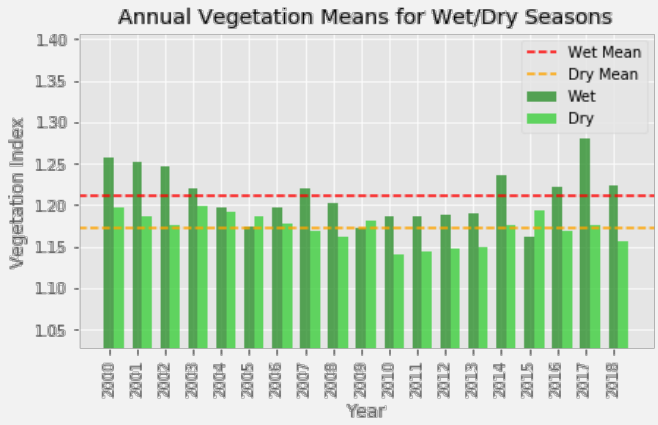
Spend some time exploring the difference between graphs and slides, that show appear like so:

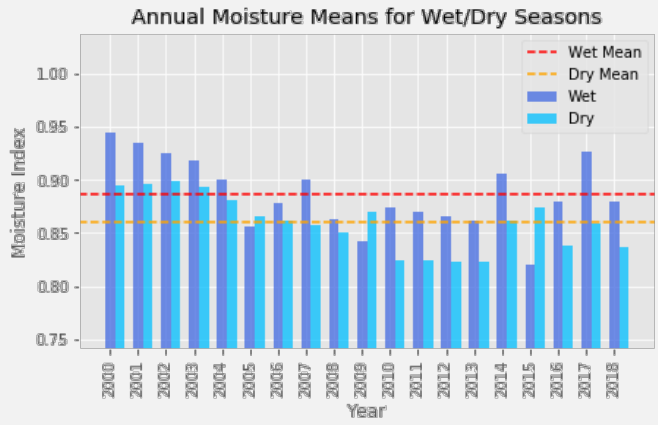
Before interpolation and outlier removal





After interpolation and outlier removal





Also take some time to explore your vegetation and moisture figure.

When ready, proceed to step 01G.

Step 01G: Prepare invariant target sites for vegetation and moisture standardisation

The next several steps of the tool (01G, 01H and 01I) are used to standardise and re-scale vegetation/moisture images to make them commensurate across seasons and years.

Variation in measured satellite sensor spectral information caused by factors other than variation in vegetation and moisture cover on the ground can reduce the accuracy at which vegetation estimates can be made from the imagery. We will attempt to correct some of these variations via standardisation.

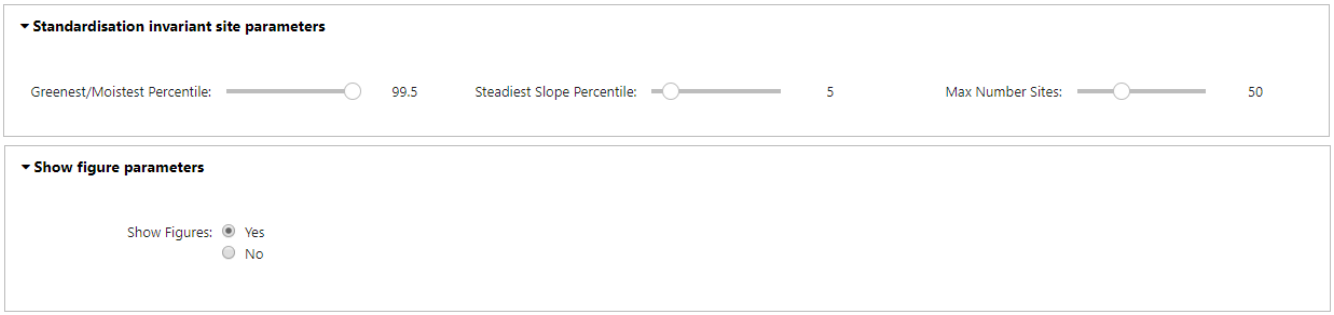
The GDV Tool uses a robust standardisation process that involves the generation of 'invariant target sites', which are essentially the most green, moist and steady (i.e. consistently un-changing) image pixels across all dry season images.

The most green and moist pixels are obtained from vegetation/moisture median baseline images with values above the percentile set by the user. Steadiness is determined from vegetation/moisture trend slope images calculated via orthogonal polynomial regression (Snedecor 1956) for dry season images. Steadiest pixels can be obtained where slope equals (or is close to) 0.

These invariant target sites, along with fuzzy sigmoidal functions (Robertson et al. 2004), are used to standardise and re-scale image values to a range from 0-1.

Run the code cell directly beneath the step 01G notes to reveal the control panel, then spend some time reading the tool notes for an overview of the control panel parameters.

The control panel should appear as below:



When you are ready to move on, set the control panel parameters to match the following:

- Greennest/Moistest Percentile: 99.5
- Steadiest Slope Percentile: 5
- Max Number Sites: 50
- Show figures: Yes

Parameter explanation:

Based on these choices, you have opted to designate only the upper 99.5% of vegetation/moisture values in our all-time dry season baseline images as the 'greennest' and 'moistest' pixels. This is quite strict. You have also selected the orthogonal polynomial regressions slope values closest to 5% either side of 0 (i.e. only the steadiest of vegetation and moisture across all-time). Finally, you have limited the number of invariant target sites to 50. If the process captures more sites than 50, that number will then be reduced to 50.

When ready, proceed to step 01H.

Step 01H: Generate and visualise invariant target sites parameters prior to standardisation

We will now generate and visualise the invariant target sites.

Please read the tool notes for step 01H and when ready, run the code cell beneath the tool notes to begin the process. This process shouldn't take long to complete.

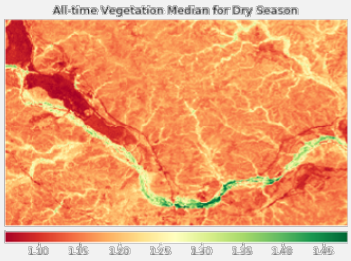
Eventually you should see an output similar to below:

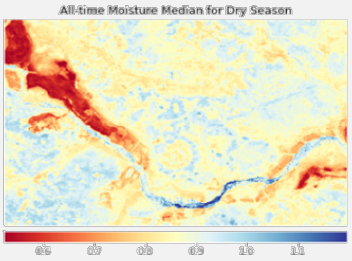
```
Generating invariant target sites for vegetation/moisture for dry season. Please wait.

Getting highest valued vegetation and moisture pixels. Please wait. 1

Generating orthogonal polynomial coefficient and slope for vegetation and moisture data. Please wait.
> Generating dry season slope for var: veg_idx. Number of coefficients: 19. Sum of squares: 570. Constant: 1. 2
> Generating dry season slope for var: mst_idx. Number of coefficients: 19. Sum of squares: 570. Constant: 1.
> Getting steadiest parts of slope for vegetation and moisture data. Please wait.

Building invariant green/moist sites. Please wait. 3

Building reduced invariant green/moist sites. Please wait.
> Number of invariant sites lower than user defined. Skipping reduction. 4
> Number of invariant sites lower than user defined. Skipping reduction.
Preparing various standardisation figures. Please wait.

Successfully generated invariant green/moisture sites for standardisation.
```

Result explanation:

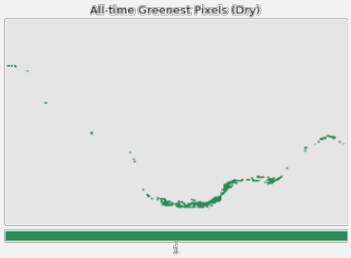
From this output we can see the 'greennest' and 'moistest' pixels across all-time (dry season) were extracted successfully; no errors were flagged (1). We are provided with some information about the orthogonal polynomial regression parameters (2). It appears invariant target sites were generated successfully also, as no errors were raised (3). Interestingly, the number of invariant target sites appears to be lower than our maximum number of 50 we set earlier (4). We may need to adjust our control parameters in step 01G to correct this (or ignore and proceed).

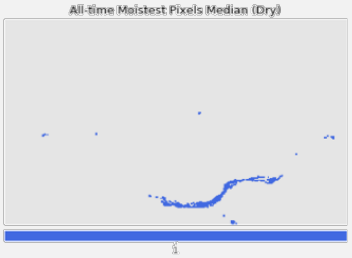
Let's see how the invariant sites look visually before we proceed. See the following page for an overview of all outputs.



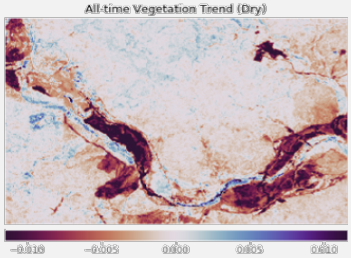


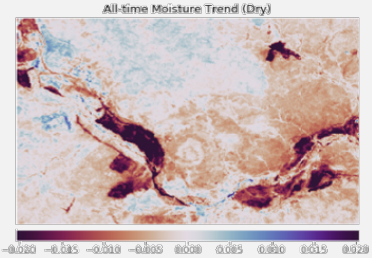
These two images show the vegetation and moisture all-time dry season medians that the 'greenest' and 'moistest' pixels are generated from.



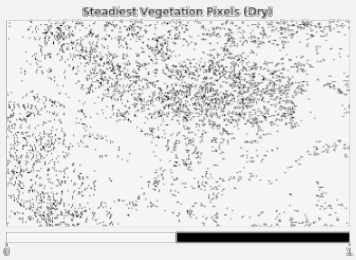


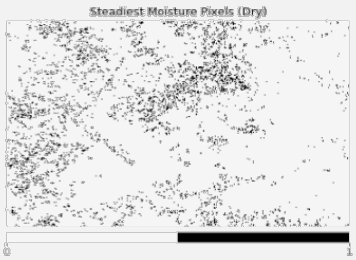
From the above all-time medians, these are the 'greenest' and 'moistest' pixels based on the 99.5% percentile value set earlier.



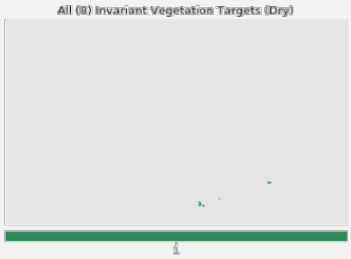


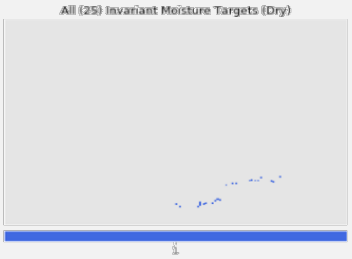
This is the slope generated from orthogonal polynomial regression. White pixels (values 0 or close to 0) are steadiest. Blue and Red are not.



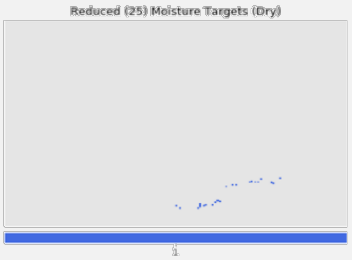


From the slope image above, black pixels shown here are the steadiest pixels of the slope using the 5% percentile set earlier.





These pixels are where 'greennest' and 'moistest' areas intersect 'steadiest' pixel i.e. our invariant target sites Notice we have < 50 for both.



The reduced invariant target sites as set by user (i.e. 50 maximum). No reduction occurred as we did not have > 50 sites.

Based on the above figure, we can see we did not reach the invariant target site maximum number (50) that we requested during step 01G.

Considering the current study area is relatively small, it is possible the current selection of invariant target sites would suffice. However, for demonstration purposes, let's try and increase the number to 50.

Return back to the control panel at step 01G and change the parameters to match the following:

- Greennest/Moistest Percentile: 98
- Steadiest Slope Percentile: 5
- Max Number Sites: 50
- Show figures: Yes

Parameter explanation:

We have simply reduced the upper 'greennest' and 'moistest' percentile so as to allow for more vegetation and moisture pixels to be returned.

Now, go back to the step 01H code cell and re-run it

This will re-generate and re-visualise the invariant target sites using our new parameter values.

You should the following at the end of your figure output:

Result explanation:

Although your results may vary slightly, you should now have 50 (or similar) invariant target sites for both vegetation and moisture indices. The number of invariant target sites can be seen in the figure subtitles.

We were able to reduce the number of invariant sites exactly 50 vegetation and moisture sites, as a result.

Follow this workflow in the future if you need to reach a specific number of sites.

We will now proceed and actually standardise our satellite images using these invariant target sites.

Proceed to step 01I.

Step 01I: Perform vegetation and moisture standardisation using invariant target sites

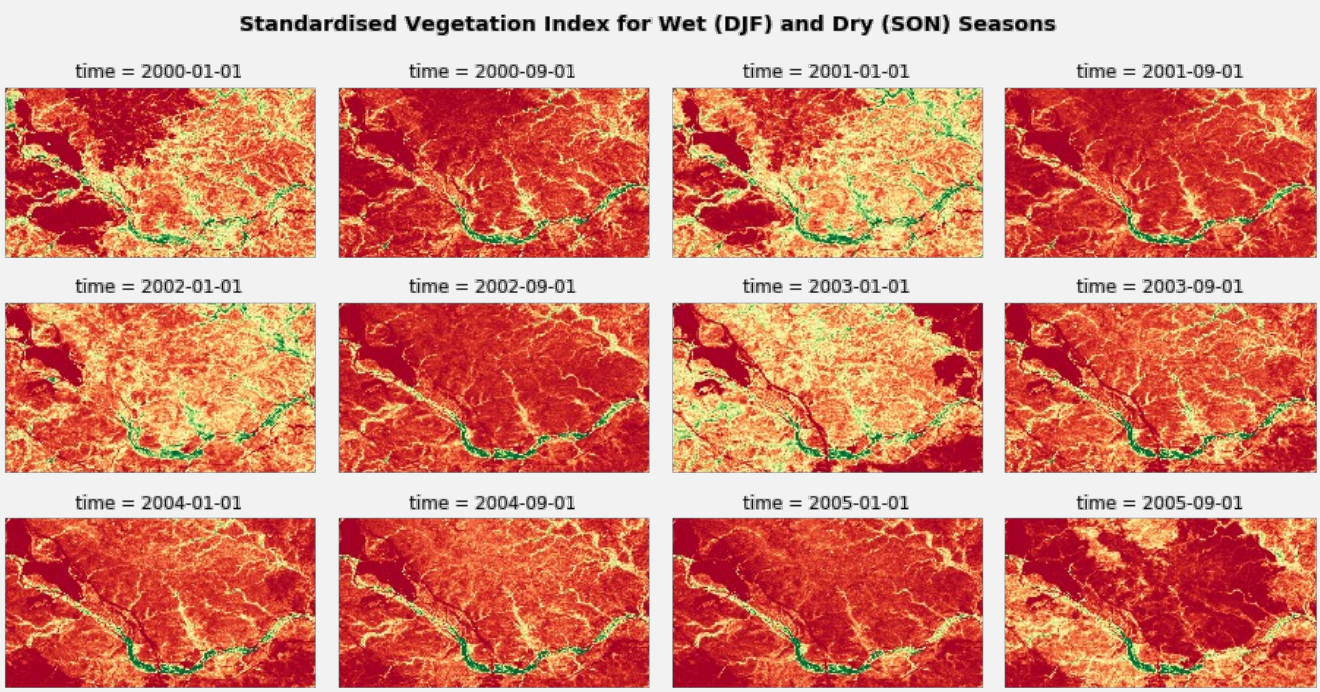
We will now apply standardisation to our vegetation and moisture images using the 'invariant target sites' generated in step 01G and 01H.

This process makes use of increasing fuzzy sigmoidal functions (Robertson et al. 2004) to re-scale images to a new range of 0-1 to ensure images commensurate across seasons and years.

Please read the tool notes for step 01I and when ready, run the code cell beneath the tool notes to begin the process. This process shouldn't take long to complete.

You should see an output similar to below:

Result explanation:



Based on this output, you can see that higher areas of vegetation appear very green on these images. Less dense vegetation appears as yellow, and no vegetation appears as red.

They should appear more consistent now that standardisation has been applied, although particularly wet or dry dates may still stand out.

Spend some time exploring the remainder of the figure.

We have no applied standardisation to our vegetation and moisture images.

When ready, proceed to step 01J.

Step 01J: Generate and standardise vegetation and moisture stability

We will now generate an additional variable known as vegetation/moisture stability (or persistence) that will be used in the spatial multicriteria evaluation modelling techniques used to derive GDV likelihood layers.

Perennial high greenness and perennial moist foliage have been found to permit the detection of GDV from remote sensing imagery (O'Grady et al. 2011; Barron et al. 2014). As we have already generated standardised vegetation and moisture images, persistent green and moisture can be obtained simply by subtracting wet season vegetation/moisture from

standardised dry season vegetation/moisture images. Stability, which improves the likelihood of a GDV-site when coupled with high vegetation and moisture layers, can be inferred from small differences (ideally 0) between these two extremes.

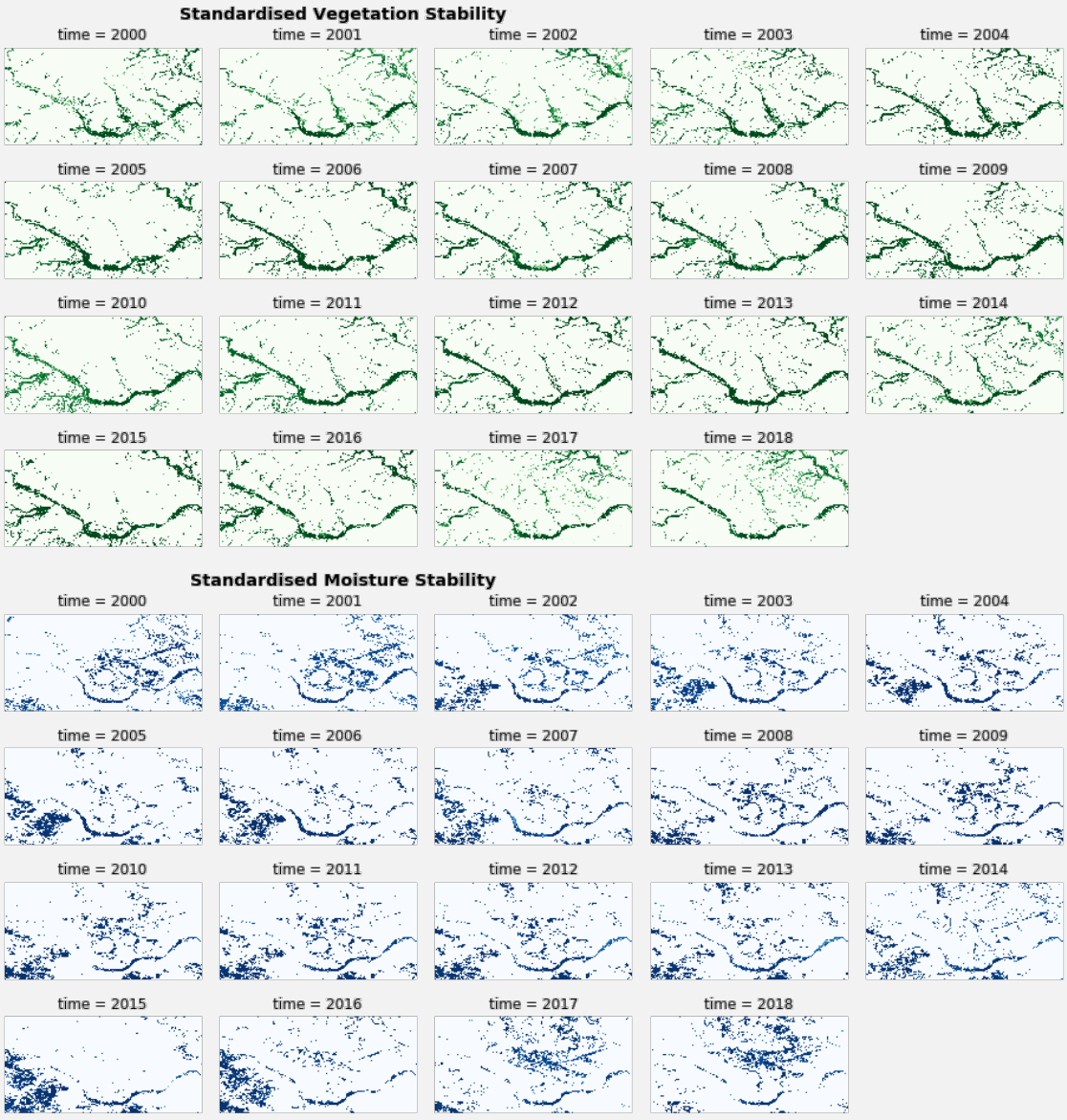
An increasing/decreasing fuzzy sigmoidal function (Robertson et al. 2004) is also applied to re-scale values from -1 to 1 to 0-1 to make stability images commensurate across years.

Please read the tool notes for step 01J and when ready, run the code cell beneath the tool notes to begin the process.

No control panel or user-inputs are required. This process shouldn't take long to complete.

You should see an output similar to below:

Result explanation:



Based on this output, areas of vegetation and moisture stability will appear as darker green and blue, respectively.

Spend some time exploring the remainder of the figure.

When ready, proceed to step 01K.

Step 01K: Weight vegetation, moisture and stability images and generate GDV likelihood

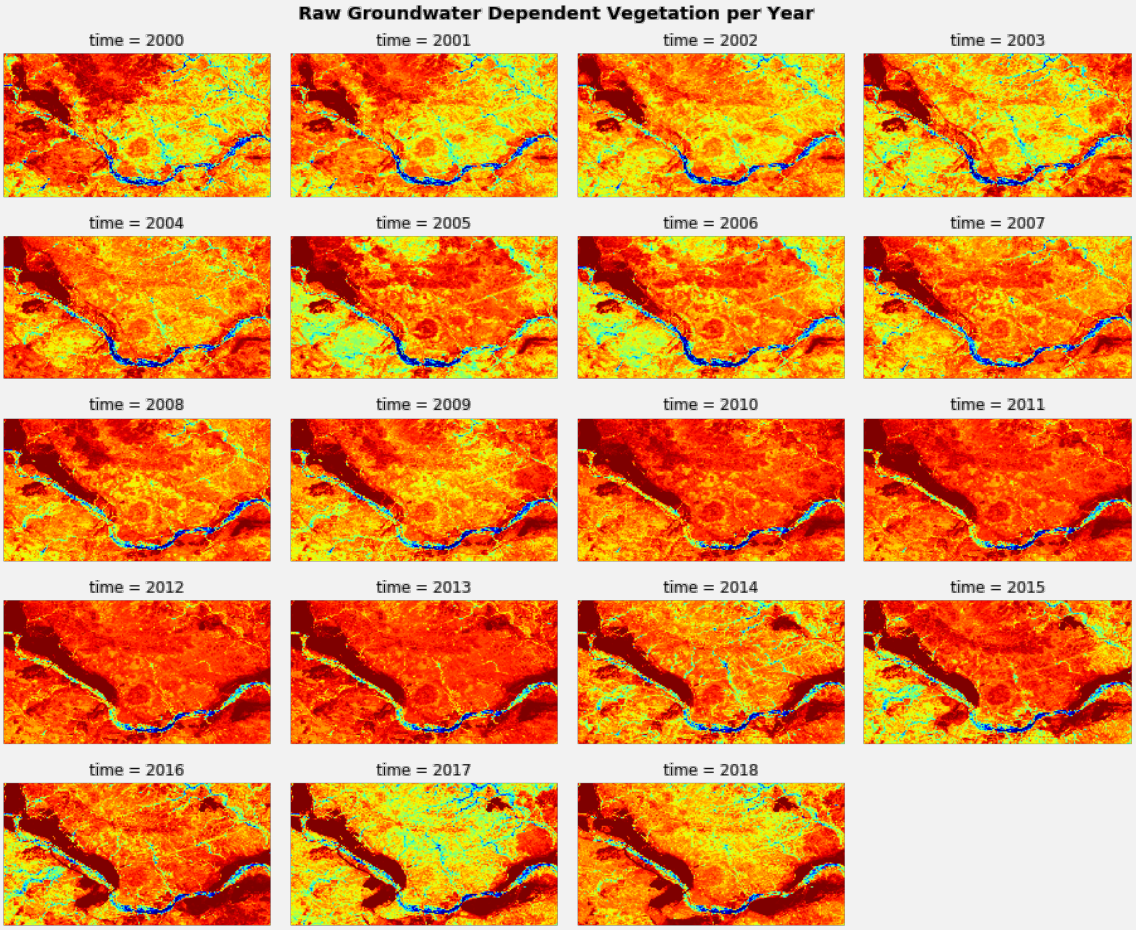
With all the imagery we've generated, we will now derive GDV likelihood layers for each year in our dataset.

This process uses a sophisticated weighting process. Weights for each vegetation, moisture and stability image have been developed using pairwise comparison in the context of the Analytical Hierarchy Process (Saatay 1980). These weights are applied to each vegetation, moisture and stability layer, and finally summed for each year. This results in annual GDV likelihood layers, each on a continuous scale of 0-1, with 0 representing locations highly unlikely to be GDV and 1 representing high likelihood.

Please read the tool notes for step 01K and when ready, run the code cell beneath the tool notes to begin the process. No control panel or user-inputs are required. This process shouldn't take long to complete.

Eventually you should see an output similar to below:

Result explanation:



The resulting images show areas of GDV likelihood for each year. High GDV likelihood areas are represented as darker blue or 'cooler' colours closer to blue (i.e. turquoise).

In contrast, 'hotter' colours like dark red or orange represent areas in the study area highly unlikely to be GDV. Those colours between blue and red (i.e. yellow) are more uncertain GDV areas.

It is worth noting that particularly high rainfall or post-bushfire greening tend to drive up GDV likelihood (see 2017, which is a particularly high rainfall year). Some of this variation can be reduced in the next step.

When ready, please proceed to step 01L.

Step 01L: Generate potential bushfire areas (optional)

GDV likelihood mapping can be influenced by bush fires (e.g. post-fire greening). This has the potential to lead to misleading GDV likelihood values for particular years when significant bush fires occurred.

Here, an optional step is provided to detect and visualise potential fire areas across GDV likelihood dates.

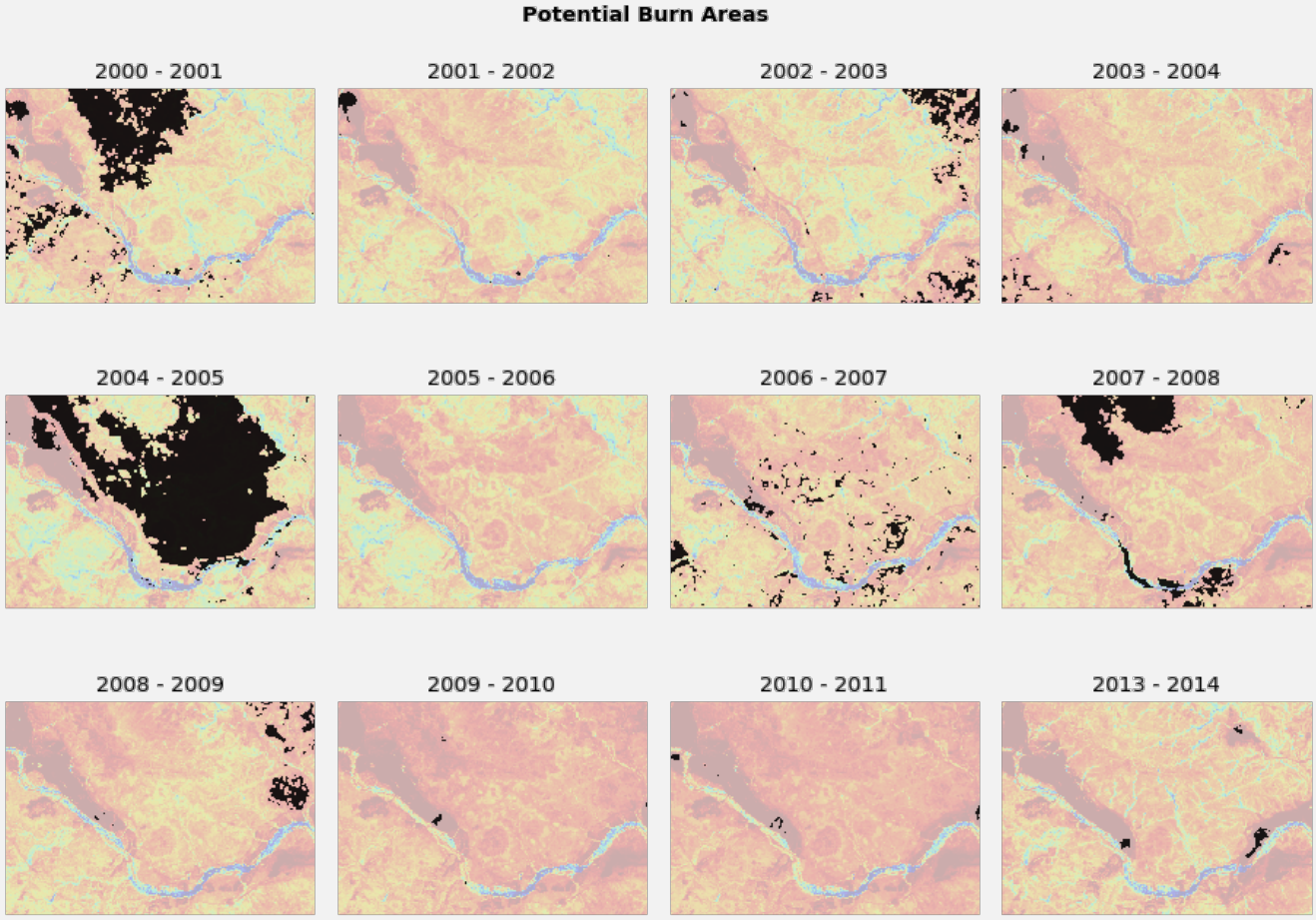
The process uses the normalised burn ratio (NBR), which returns high values for potential bushfires. Using NBR images, severity of fire can be detected by calculating delta NBR (ΔNBR) by subtracting post-fire imagery from pre-fire imagery. An overview of this process is offered by Keeley (2009).

This step does not require user input and therefore presents no control panel or control parameters.

To see potential fire areas, simply run the code cell beneath the step 01L notes. This may take some time.

Eventually, you should receive a figure that appears similar to below:

Result explanation:



High-severity burn areas are represented here as black pixels. We can see that at some point between years 2000 to 2001, a significantly large fire occurred in the northern section of the study area. Similarly, another significant fire occurred between years 2004 to 2005.

While this step is optional, it can provide context as to why a particular GDV likelihood layer may look considerably different to others. Note: areas completely cleared for infrastructure can also come back as burn areas.

When ready, please proceed to step 01M.

Step 01M: Set threshold method for GDV likelihood layers

The final steps of GDV likelihood modelling (steps 01M and 01N) involve performing GDV likelihood presence/absence thresholding. Thresholding is helpful as it dichotomises GDV likelihood layers into presence/absence, improving clarity.

This tool offers to approaches to GDV likelihood thresholding:

1. ROC-slicing (e.g. Robinson et al. 2010) using known GDV locations captured during field surveys; or
2. Standard deviation-based thresholding.

The ROC-slicing approach is limited in that it requires a shapefile of known GDV species locations (e.g. *Eucalyptus victrix* individuals).

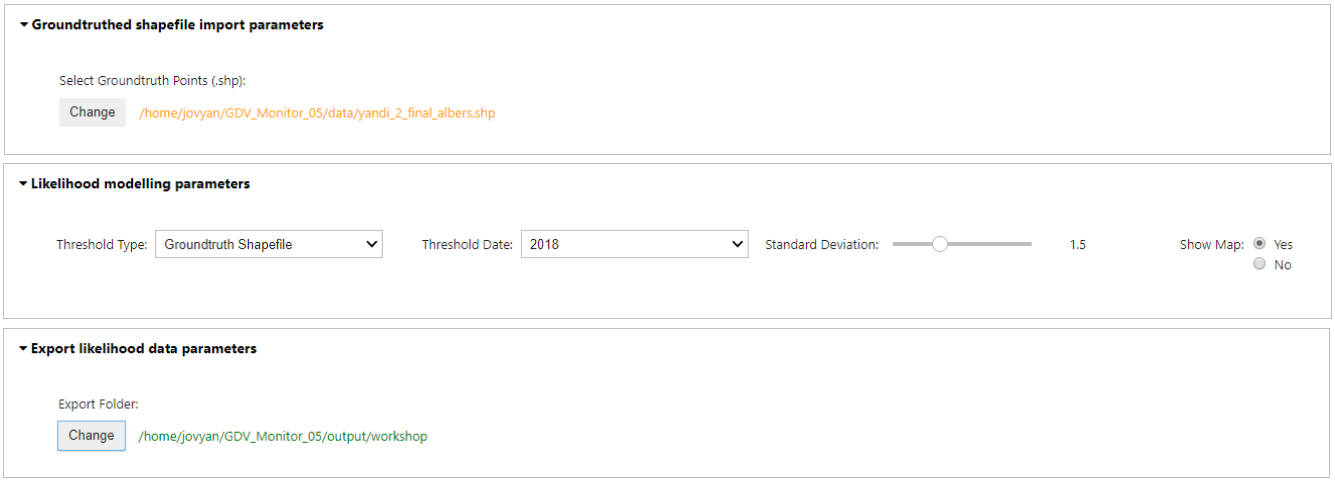
This data must be uploaded to JupyterLab in a specific format outlined in the tool notes. Regardless, this approach provides an output containing a graph on GDV likelihood model accuracy (via the area under the curve or AUC metric).

If a shapefile is unavailable, thresholding via standard deviation has worked well in current tests and offers a decent alternative.

We will go through both thresholding approaches, starting with the Groundtruth-based ROC-slicing method.

Run the code cell directly beneath the step 01M notes to reveal the control panel, then spend some time reading the tool notes for an overview of the control panel parameters.

The control panel should appear as below:



When you are ready to move on, set the control panel parameters to match the following:

- Select ground-truth points: (*select yandi_2_final_albers.shp from the data folder*)
- Threshold type: Groundtruth Shapefile
- Threshold date: 2018
- Standard deviation: 1.5 (*will be ignored when using a shapefile*)
- Show map: Yes
- Export folder: (*select the output folder and call the export file 'workshop'*)

Parameter explanation:

You have opted to use an existing shapefile of ground-truthed GDV tree locations. This data is a subset of GDV tree species presence/absence locations taken from one of the three Pilbara-based tenements visited by the research team in 2019.

These locations will be used by the tool to determine how accurate our GDV model performed and to threshold GDV likelihood layers. You have also set the threshold type to Groundtruth Shapefile to ensure the shapefile approach is used.

A threshold date of 2018 has also been set. This will specifically select the GDV likelihood layer from 2018 to use in ROC-slicing. Your GDV likelihood layer should match the year the ground-truthing was undertaken to ensure actual observations

match predicted values as closely as possible. We would use 2019 if available, but as of writing the Landsat dataset lacks the last half of 2019.

The last parameter worth mentioning is the export folder. When thresholding is finished, the GDV Tool will export a GIS-friendly netcdf file (.nc) of the GDV likelihood layers (thresholded and non-thresholded) for external/later use.

When ready, proceed to step 01N to actually perform the GDV likelihood threshold method and plot the results on an interactive map.

Step 01N: Perform GDV likelihood layer thresholding and map the result

Let's generate and visualise the thresholded GDV likelihood layer for 2018 using field-based GDV species locations.

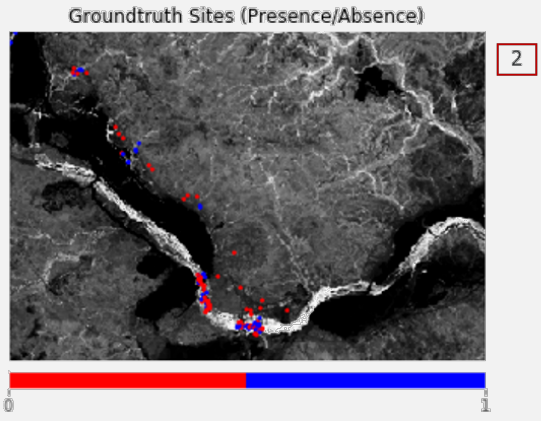
Please read the tool notes for step 01N and when ready, run the code cell beneath the tool notes to begin the process. This process shouldn't take long to complete.

Eventually you should see an output similar to below:

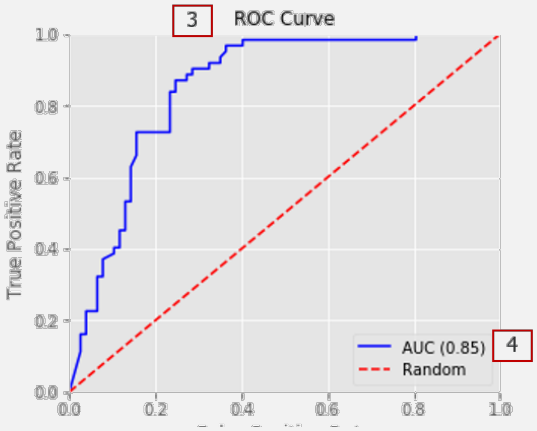
Result explanation:

```
Attempting to threshold groundwater dependent vegetation likelihood map. Please wait.

Thresholding groundwater dependent vegetation likelihood using groundtruthed shapefile. Please wait.
Groundtruth shapefile opened successfully. Beginning to extract data for ROC analysis. Please hold.
Sucessfully prepared shapefile for ROC analysis. Num of ground sites within selected study area: 139
Beginning ROC analysis. Please hold.
Successfully completed ROC analysis. Proceeding.
Successfully thresholded groundwater dependent vegetation likelihood. Proceeding.
Generating shapefile point and ROC curve, please hold.
```

1



2



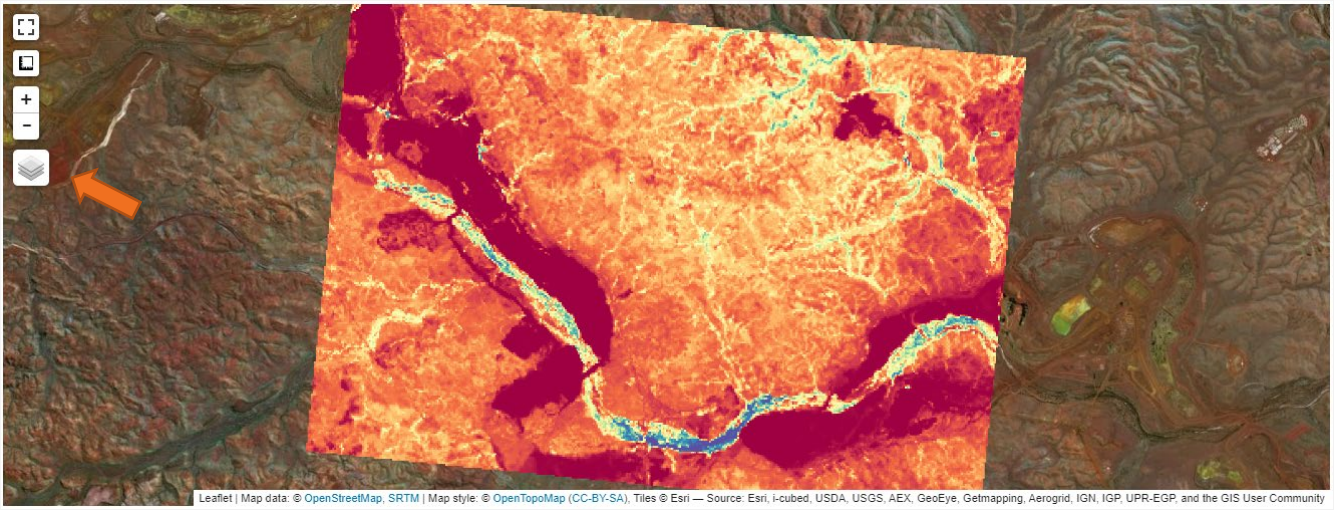
Based on this output, we are told that the provided shapefile data was successfully prepared and used to perform a ROC analysis (1). Under that output, on the left, is a grayscale image of mean vegetation with blue and red points dispersed in areas throughout it (2). These points are a subset of GDV tree locations represented as presence (blue) and absence (red) that were captured from the field in 2019.

The final output here is the ROC Curve graph (3). This graph provides us with a general indication of model performance, which is calculated using our predicted GDV likelihood values against actual observed GDV species at the same pixels. This plot plots the true positive rate (TPR) and false positive rate (FPR) for numerous classification instances. This helps us determine what optimal TPR and FPR can be achieved using the model. For example, if we use a model that obtains a TPR of ~0.82, we can expect an FPR of ~0.2. The GDV Tool will automatically find the optimal threshold value.

Also helpful is the area under the curve (AUC), which is a measure of how well the model can distinguish between GDV and non-GDV classes (4). If no difference between GDV and non-GDV can be determined, we can expect an AUC of 0.5 (random). Perfect separation would this be an AUC of 1. Our model has obtained an AUC = 0.85.

More information on ROC curves be found here: https://en.wikipedia.org/wiki/Receiver_operating_characteristic.

Further on, underneath the ROC plot, should be an interactive map, similar to below:



The GDV likelihood layer shown here is the result for the year 2018, as requested in the control panel in Step 01M.

You can pan around the map, and zoom in and out to explore the results.

When ready, click the layer control icon (see arrow above) to switch between 'Likelihood Full' (prior to thresholding) or 'Likelihood Thresholded' (post-ROC slicing) layers.

Turn off the 'Likelihood Full' layer to see just the thresholded layer and explore it too.

We will now use standard deviation to threshold GDV likelihood layers. Return to the control panel at step 01M and change the control panel parameters to match below:

- Select ground-truth points: (*leave as is*)
- Threshold type: Standard Deviation
- Threshold date: All
- Standard deviation: 1.5
- Show map: Yes
- Export folder: (*leave as is*)

Parameter explanation:

Here you have opted to threshold based on a standard deviation of 1.5 (i.e. any likelihood values > 1.5 standard deviations are considered to be very high GDV likelihood, everything else to be non-GDV).

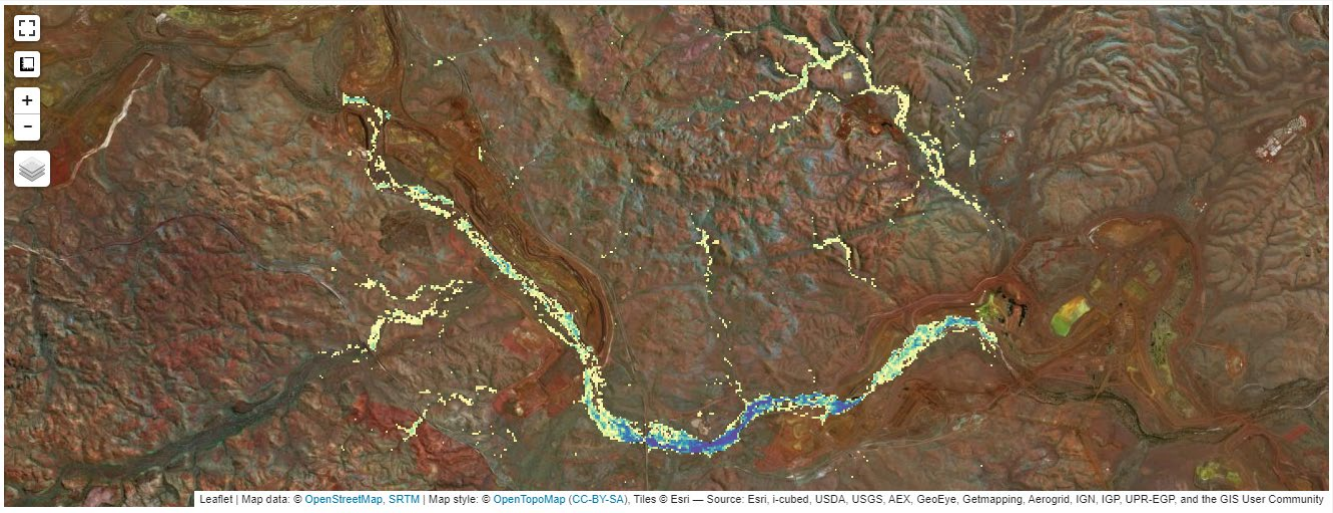
You have also chosen a threshold date of 'All', which has taken the median of all available GDV likelihood layers, instead of a specific year. This can help reduce some of the seasonal fluctuation that can occur.

As before, when ready, proceed to step 01N and re-run the code cell.

This will re-threshold the GDV likelihood data using standard deviation this time around, and will then rebuild the interactive map with the new results plotted on it.

The results should appear similar to below:

Results explanation:



Spend some time exploring the map again.

Also, feel free to go back to the control panel at step 01M and increase/decrease the standard deviation value to see what happens to the GDV likelihood areas.

When ready, proceed to the final step of Section 01.

Step 01O: Finalise GDV likelihood layer

The last step of Section 01 involves basic clean up and memory clearing procedures. It is required if proceeding to the next sections (Section 02 and/or 03).

Please read the tool notes for step 01N and when ready, run the code cell beneath the tool notes to begin the process. This process shouldn't take long to complete.

You should see the output below if successful:

```
Clearing original dataset variable (ds).
Clearing original stability variable (stability).
Clearing original mask for vegetation invariant targets (mask_inv_veg).
Clearing original mask for moisture invariant targets (mask_inv_mst).
Transferring metadata over to trend/change parameters.
Disabling likelihood dashboard. Rerun the tool if you require a different threshold.
Cleaning process complete.
```

No real need to explain these results. Either way - Success!

This concludes Section 01 of the tool. The next section (Section 02) involves generating and exploring monotonic trends over time on thresholded GDV likelihood areas using Mann-Kendall trend analysis methods.

Section 02: Monitor GDV health trends via Mann-Kendall trend analysis

The second section of this tool involves the application of several trend analyses to detect anomalous, continuously decreasing (or increasing) trends in plant health. The intent is to provide an assistive, user-driven, interactive approach to exploring GDV health trends over a specific time period and study area.

The resulting trend layers are coupled with graphing tools that facilitate exploration of vegetation health history anywhere within the study area (i.e. both inside and outside of GDV trend areas). Additionally, change point detection (e.g. Andersen 2009) is also incorporated to allow users to detect 'break points' in vegetation health history (i.e. dates when vegetation health signatures significantly shifted away from norm).

Mann-Kendall (MK) tests (Mann 1945; Hipel and McLeod 2005) are used to perform trend analysis. A MK test determines if there is a monotonic increase, monotonic decrease or no significant change over time. The MK non-parametric trend test is a function of the ranks of the observations rather than their actual values and so does not require the assumption of normally distributed data (Hamed 2008) like, for example, linear regression. Furthermore, the test allows the mapping of p-values and thus provides a confidence level for interpretation.

Change point detection is used in time-series analysis to detect abrupt shifts in time series trends (i.e. shifts in a time series' instantaneous velocity). This can be applied to a time-series of vegetation health observations to detect anomalous sequences/states in the vegetation health history, and to detect dates where sudden changes occurred (e.g. Andersen 2009).

The key processing steps applied in this section are as follows:

1. Fetch a 'cube' of Landsat or Sentinel satellite data for a user-defined study area and timespan using the DEA ODC;
2. Generate vegetation index based on user-selection;
3. Reduce data to selected season(s)/month(s) and interpolate missing data (optional);
4. Standardise vegetation using a basic invariant target site approach;
5. Perform MK test;
6. Map trend results onto interactive map; and
7. Graph vegetation history and perform change point detection on any pixel clicked by user.

Proceed to step 02A when ready.

Step 02A: Set trend analysis parameters

Now that we have generated and thresholded GDV likelihood in Section 01, the next step is to perform a MK test(s) within those GDV areas to provide insight into vegetation health trends and change points.

Two different types of MK test are included for use in this tool:

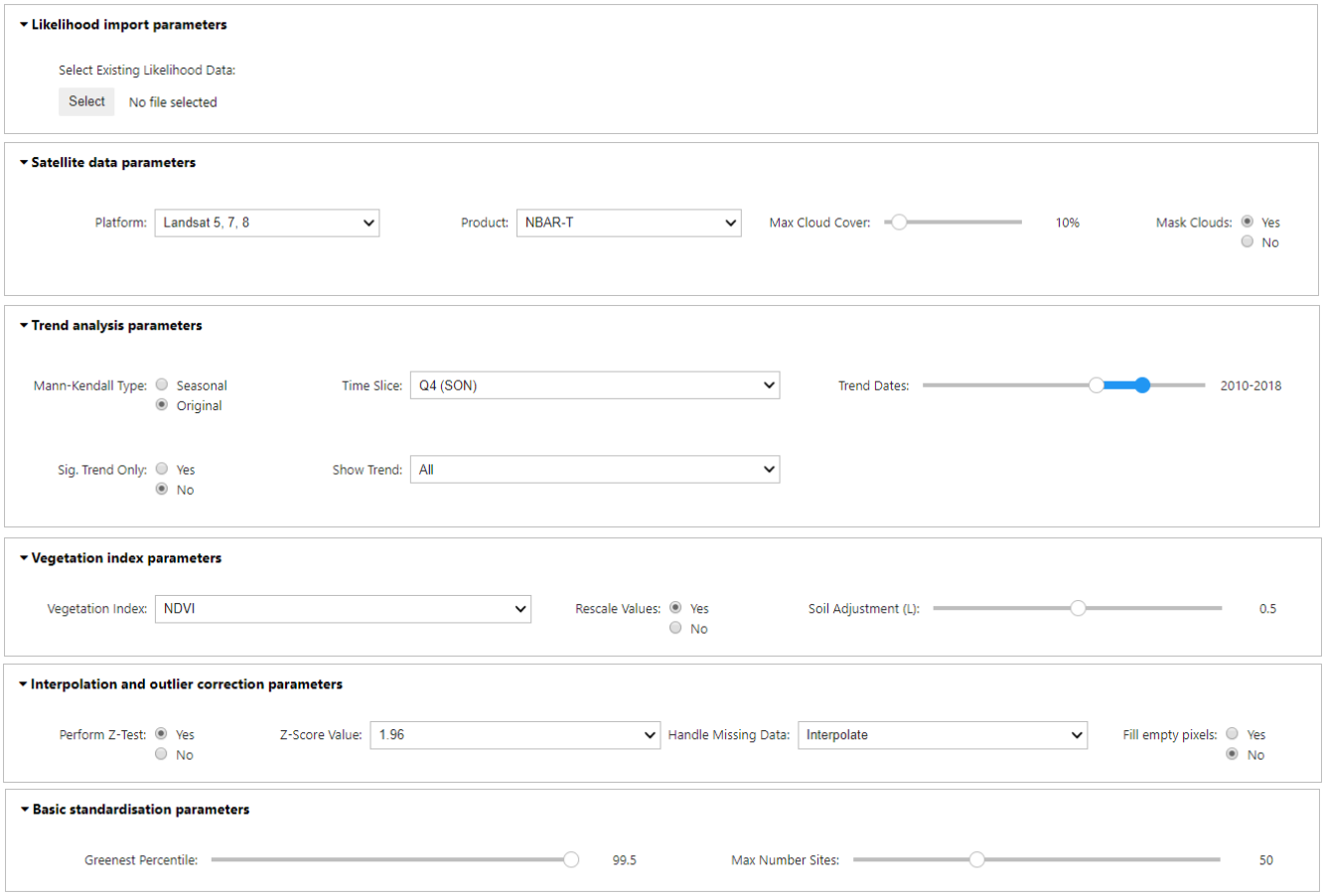
1. Original MK: is a nonparametric test which does not consider serial correlation or seasonal effects. This method is most appropriate when exploring health trends for a specific season or month across years. For example, exploring trends at a monitoring site for every January from 2009 to 2019; and

2. Seasonal MK: is a nonparametric test that considers seasonal effects and is suited for time-series data with seasonal trends (Hirsch et al. 1982). This method is most appropriate when exploring vegetation trends across every season or month across years. For example, exploring trends at a monitoring site across every season beginning 2009 and ending 2020. This approach is recommended as it allows more data to be incorporated in the MK test.

We will try both types in this workshop, starting with the original MK test.

Before we get into it, run the code cell directly beneath the step 02A notes to reveal the control panel, then spend some time reading the tool notes for an overview of the control panel parameters.

You should be familiar with most of these parameters already from Section 01. Those only new to you should be the likelihood import and trend analysis parameters. Regardless, the control panel should appear as below:



When you are ready to move on, set the control panel parameters to match the following:

- Select existing likelihood data: (*keep blank*)
- Platform: Landsat 5, 7, 8
- Product: NBART
- Max Cloud Cover: 10%
- Mask Clouds: Yes
- Mann-Kendall Type: Original
- Time Slice: Q4 (SON or September to November)
- Trend Dates: 2010-2018
- Sig. Trend Only: No
- Show Trend: All
- Vegetation Index: NDVI
- Rescale Values: Yes
- Soil Adjustment (L): 0.5 (*only applicable to SAVI*)
- Perform Z-Test: Yes
- Z-Score Value: 1.96
- Handle Missing Data: Interpolate
- Fill Empty Pixels: No
- Greenest Percentile: 99.5
- Max Number Sites: 50

Parameter explanation:

You have opted to not use the netcdf GDV likelihood file exported earlier during step 01M, as this data is currently still in memory. You have also chosen the use the original MK test. This will calculate a trend for the same Q4 (Sep-Nov) time slice

across every year (i.e. 2010 and ending 2018). In other words, vegetation health trends will be calculated for the same season across nine years at every GDV pixel.

You have also decided not to show significant trends only, meaning the MK output will show trends that may or may not be statistically significant. You will eventually able to explore the p-value layer output on an interactive map to explore significance.

You have also decided to show any type of trend on MK output layers. This means increasing, decreasing, or no trends, will be returned.

The other parameters have been covered prior and will not be re-explained. Please investigate the tool notes if you need help.

When ready, proceed to step 02B.

Step 02B: Perform trend analysis

It's now time to perform an original MK analysis.

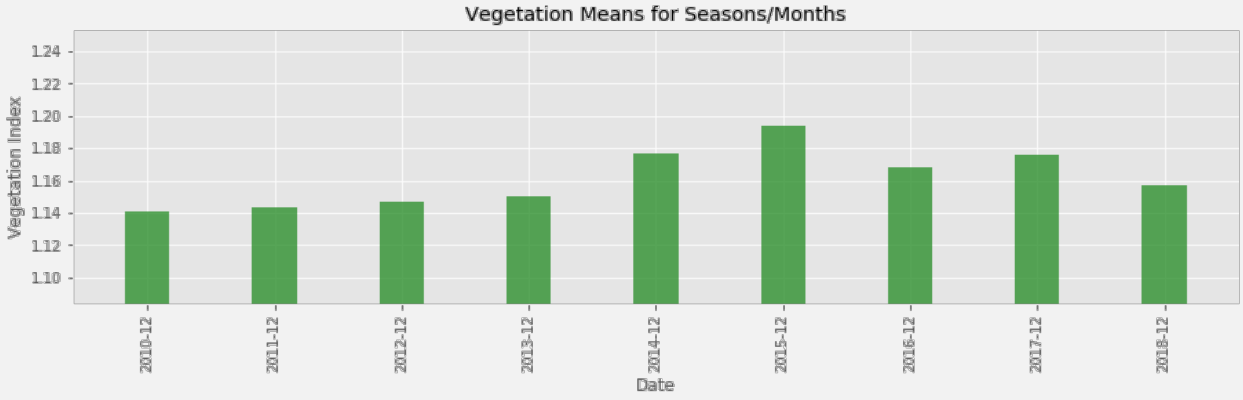
This process will create time-series pixel vectors comprised of vegetation health within our thresholded GDV dataset and perform a trend analysis on it. It is fairly processing-heavy, and thus can take a few minutes to complete. The output of this process will be three layers; the Tau, p-value, and Sen's slope layers, which will be explained shortly.

Please read the tool notes for step 02B and when ready, run the code cell beneath the tool notes to begin the process. No control panel or user-inputs are required.

Eventually you should see a page of summary information with familiar vegetation graphs. Much of this has been explained in Section 01 and won't be repeated. That said, it's important to you check this information yourself to ensure everything matches your understanding of your study area.

Result explanation:

You should see a graph of mean vegetation values similar to below:



This requires no further explanation. You should also see a progress bar beneath this graph.

This indicate that the tool is moving through every vegetation pixel vector and performing a MK test on it.

This progress bar will eventually turn green once all pixel vectors have had an MK test applied to it, and should look similar to below:

```
Basic standardisation for original mannkendall applied successfully. Proceeding.
> Preparing empty mannkendall dataset, hold on...
> Prepared mannkendall dataset successfully. Proceeding.
> Masking out non-groundwater dependent vegetation pixels to speed up processing. Please hold.
> Dataset sucessfully masked to GDV pixels. Proceeding.
> Beginning vector stacking process, please hold...
> Data stacked into vectors successfully! Mann-Kendall trend analysis now ready.
Beginning mannkendall analysis. This can take awhile. Go get a coffee!
> Original mannkendall analysis being applied, hold tight.
100% |████████████████████████████████████████| 5378/5378 [00:37<00:00, 143.79it/s]

> Successfully completed the mannkendall!
> Cleaning up some bits and pieces, please hold.
> Finished mannkendal and cleaning procedure! Proceeding.
> Transfering mannkendall result over to final dataset.
> Updating empty mannkendall result in-place using vectors, hold tight.
Mann-kendall analysis successfully completed.
```

All of this is a great sign. Your MK test completed successfully.

The next step is to plot the results of the MK test onto an interactive map.

When ready, please proceed to step 02C.

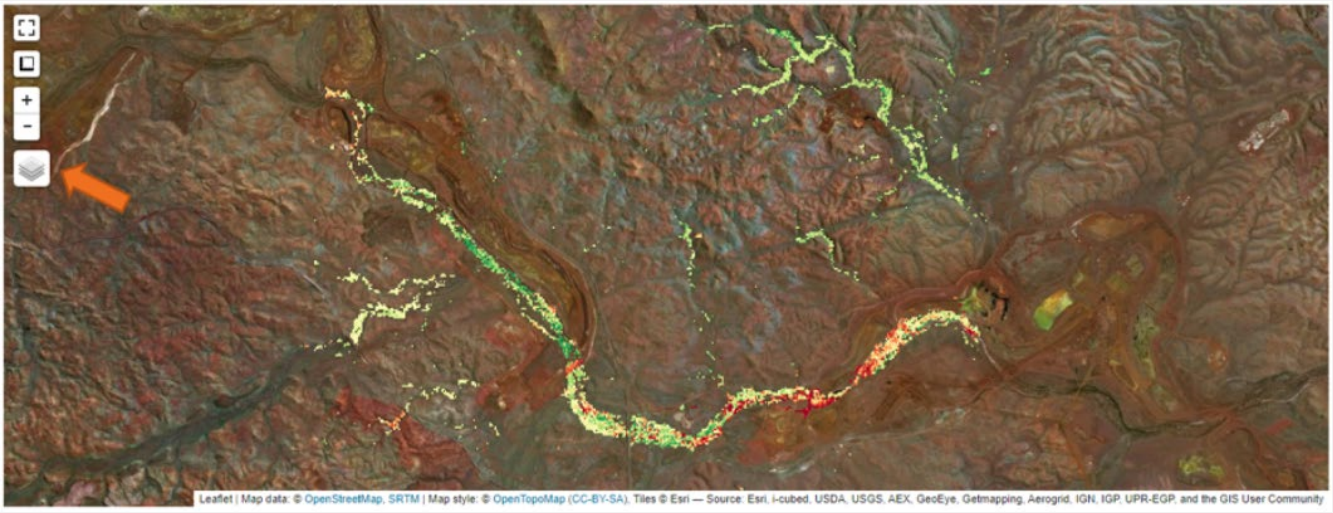
Step 02C: Visualise trend analysis results, GDV health graphing tools and change point detection controls

We will now visualise and explore the original MK test results on an interactive map and prepare the change point detection control panel.

Please read the tool notes for step 02C and when ready, run the code cell beneath the tool notes to begin the process. This process shouldn't take long to complete.

An interactive map (with a control panel underneath it) should appear, like so:

Results explanation:



This interactive map contains three layers resulting from the MK test (located in the layer control icon at orange arrow):

Tau layer: the rank correlation coefficient, which is indicative of trend direction (increasing, decreasing, no trend). A value of 1 (very green pixels) indicates a perfect, continuously increasing trend in vegetation health. A value of -1 (very red pixels) is a perfect, continuously decreasing vegetation health. A value of 0 (yellow pixels) indicates no continuous trend in any particular direction.

P-value layer: The probability of obtaining the observed trend, not unlike most statistical tests. Whiter pixels are very significant (i.e. p-value = 0.001), black/red pixels are not significant (p-value > 0.1).

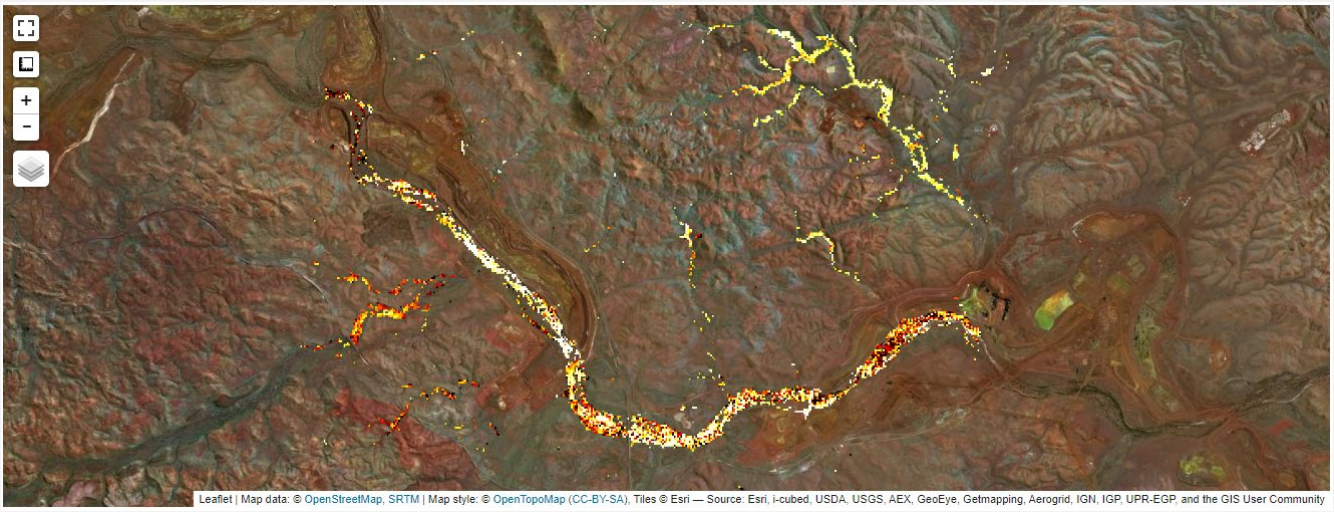
Sen's Slope layer: A more robust trend layer that ignores trends that don't occur for at least 30% of the time series. Shows robust, relatively longer-term trends and will ignore a short drought, La Nina, El-Nino cycle.

The default layer is the Tau layer. You can see that a range of increasing (green), red (decreasing), and no trend (yellow) areas have been detected.

Let's now have a look at the significance (p-value) layer. Turn the tau and Sen's slope layers off to focus on the p-values.

Your map should appear like so

Result explanation:



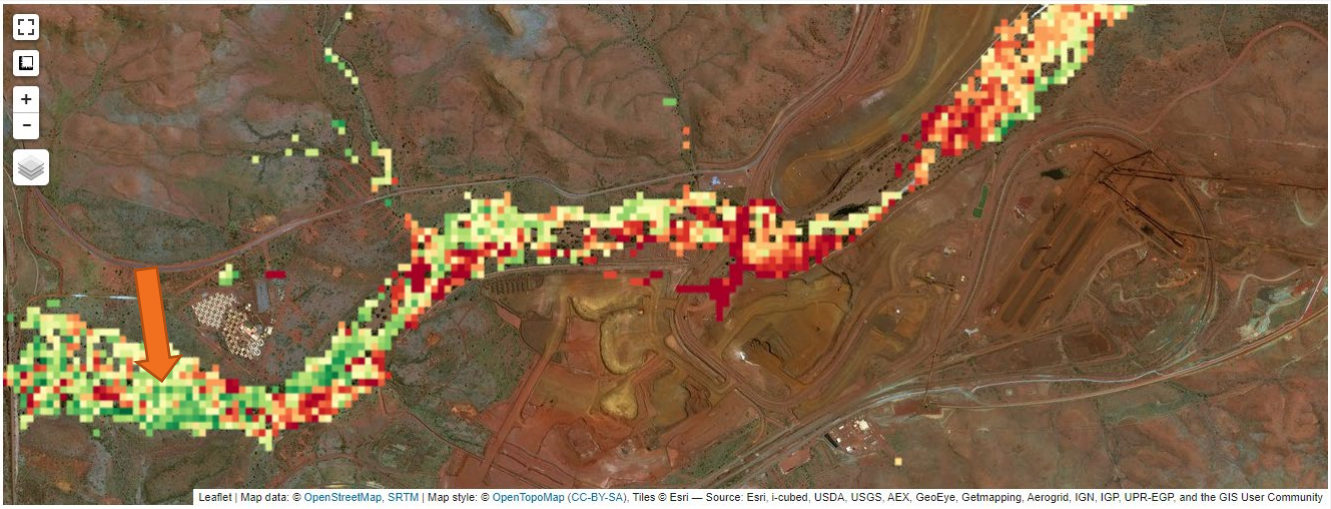
This layer presents us with the significance of trend. Colours closer to white are more statistically significant, whilst colours closer to red and black are not.

We can quickly see that not all of our Tau trends are going to be statistically significant.

Turn on the Sen's slope later now:

Sen's slope is like Tau except more robust; trend layer that ignores trends that don't occur for at least 30% of the time series. It is good at showing relatively longer-term trends and will ignore a short drought, La Nina, El-Nino cycle.

To get a more detailed look, zoom your map into the location shown at the above orange arrow and turn the Tau layer back on. Your map should appear more like below:

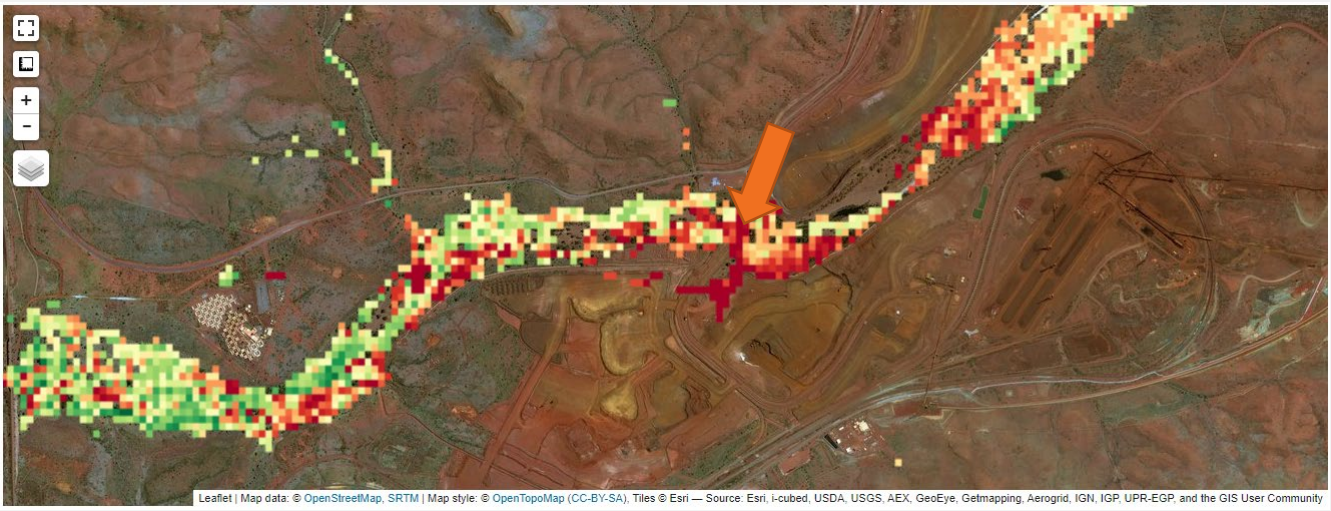


We are able to get precise information about MK results by clicking on any locations showing trend pixels. Click on one of the darker green pixels close to the orange arrow shown above. You should see a print-out similar to below appear under the map:

TAU: 0.556 | PVALUE: 0.048 | SENS SLOPE: 0.004

This tells us precisely what the Tau value is (0.556). This is a moderate increasing trend (i.e. 1 is a perfect increase, 0 is no trend). This implies that vegetation has continuously increased here from 2010-2018 during the dry season. We can also see the p-value is < 0.05, indicating this trend is statistically significant.

Let's investigate a seemingly strong decline area. Click the red pixel at the arrow below:



Again, some precise MK test results should show underneath the interactive map:

TAU: -0.667 | PVALUE: 0.016 | SENS SLOPE: -0.015

We can now see a negative Tau value of -0.667. This is a moderate to strong decreasing trend (i.e. -1 is a perfect decrease, 0 is no trend). This implies that vegetation has continuously declined here from 2010-2018 during the dry season. We can

also see the p-value is significantly lower than < 0.05, indicating this trend is statistically significant. This area has had a haul road installed within the last 5 to 8 years.

Spend some time exploring other pixels. What sort of Tau and p-values do yellow pixels have?

When you're done, you may notice a control panel underneath the interactive map called with controls for change point detection presented on it.

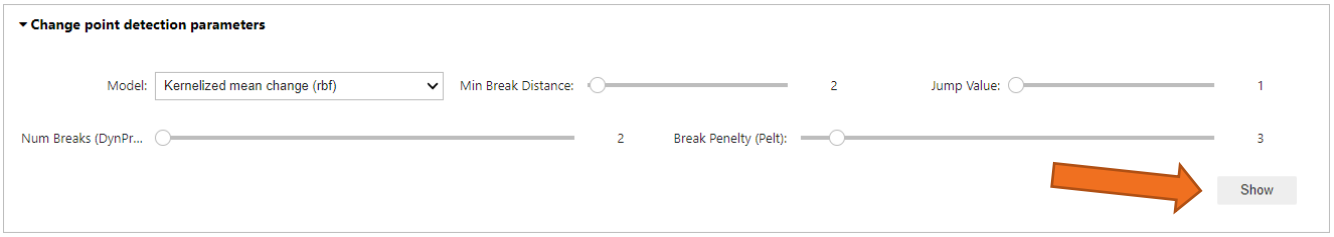
As previously mentioned, change point detection is used in time-series analysis to detect abrupt shifts in time series trends (i.e. shifts in a time series' instantaneous velocity). This can be applied to vegetation health signatures to detect precise dates of abrupt shifts in health regimes.

This tool uses two, 'exact segmentation' change point algorithms (Truong et al. 2019) to perform change point detection:

1. Pruned Exact Linear Time (PELT) search method: is an exact search method that detects break points via the minimization of costs and is generally considered to produce highly optimal and consistent results (Wambui et al. 2014). Users- determine the number of break points returned using a 'penalty' value;
2. Dynamic Programming (DynPro) search method: is also an exact method used for automated change point detection. It has a considerable computational cost of $O(Qn^2)$, where Q is the max number of change points and n is the number of data points. Generally considered less strict than PELT and, thus, returns more break points.

This control panel provides you with the ability to perform change point detection on the pixel we last clicked on and other basic vegetation health graphing tools.

It should look like below:



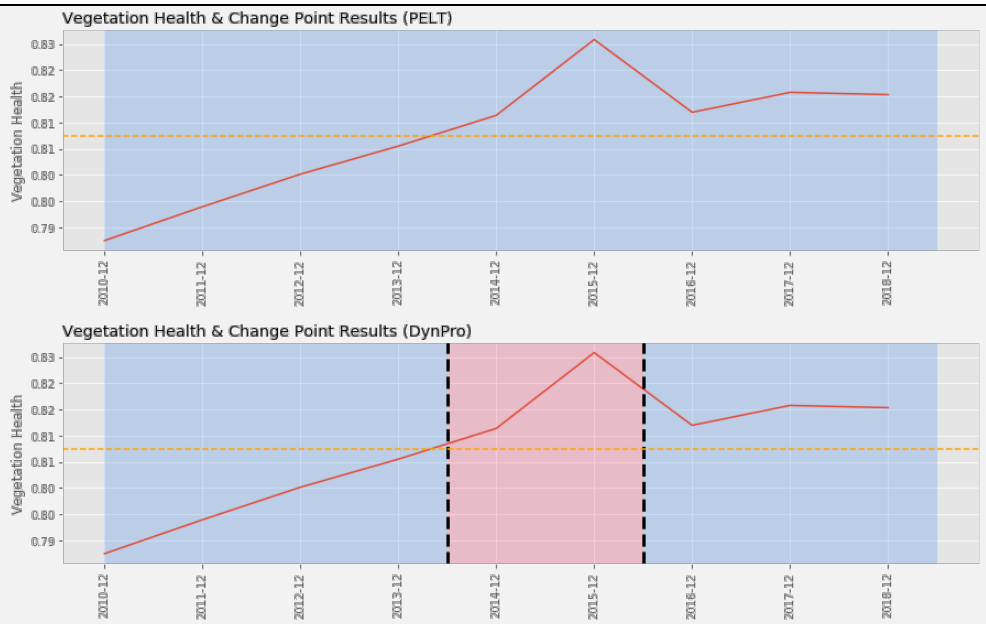
It is recommended you read the tool notes before proceeding to understand what the tool parameters mean.

When ready, click your mouse on a dark green pixel near the area you first tried above. Again, the Tau, p-value and Sen's slope values should change underneath the interactive map to indicate a new pixel has been selected.

To graph our vegetation health and perform change point detection at this pixel, click the 'Show' button as indicated at the arrow above.

The result should be graphed underneath the control panel and look something like:

Result explanation:



Both the PELT and DynPro change point detection methods (top and bottom graphs, respectively) have been applied to the same vegetation signature (the single red line).

Median vegetation across all dates is also provided (orange dashed line). At a glance, we can see vegetation has been increasing since 2010, peaking at 2015, followed by a mild decline in 2016, then a slow increase after that.

That is good – it matches the result of the original MK test (a significant moderately increasing trend).

You can see how the two methods work well together; the MK test guides you towards a trend and the vegetation signature graphs provide you with further nuance.

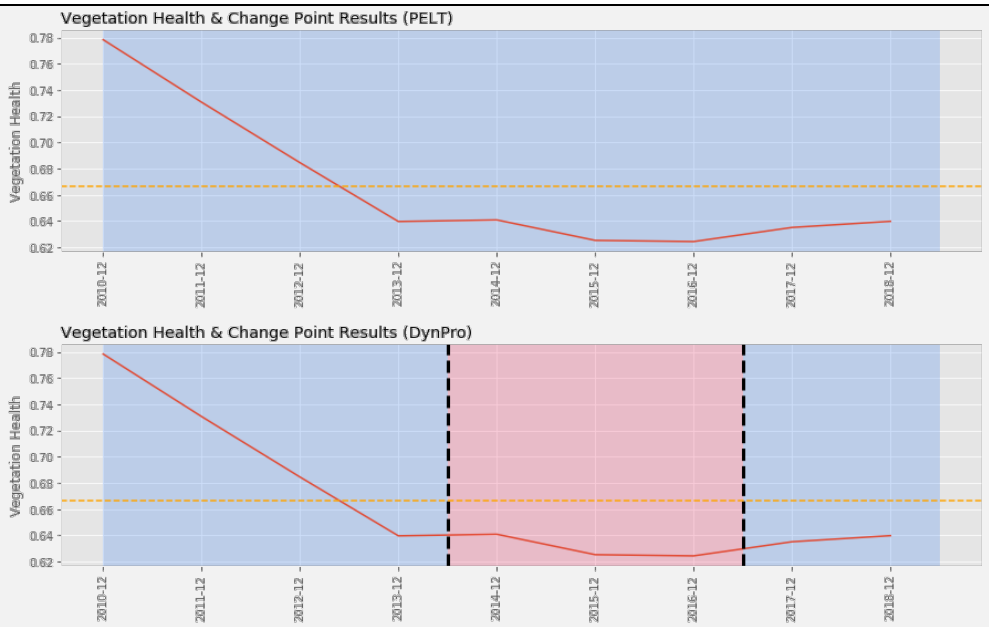
The final feature of the above graph is the break point lines (black dashed lines) and change regime areas (light red colours). You can see these on the bottom graph (DynPro) of the figure above.

We are being told that a break was detected between mid-2013 to mid-2016, which is likely due to the peak in 2015. No breaks were detected using the PELT method (top graph); PELT tends to be stricter than DynPro using our default settings (hence why both are included). The two complement each other well.

Now, try clicking on the dark red declining trend pixel from earlier and click the 'Show' button to generate a new change point detection result for the new pixel.

You may see something similar to below:

Result explanation:



As you might expect, a fairly significant negative decline signature can be seen. A break point was detected around mid-2013 using Dynamic Programming, but no breaks were returned by PELT.

We may need more data to improve the accuracy of these break points. This is where seasonal MK tests can help.

Go back up to the control panel at step 02A and make sure the trend parameters match below:

- Mann-Kendall Type: Seasonal
- Time Slice: Quarter to Quarter
- Trend Dates: 2010-2018
- Sig. Trend Only: No
- Show Trend: All

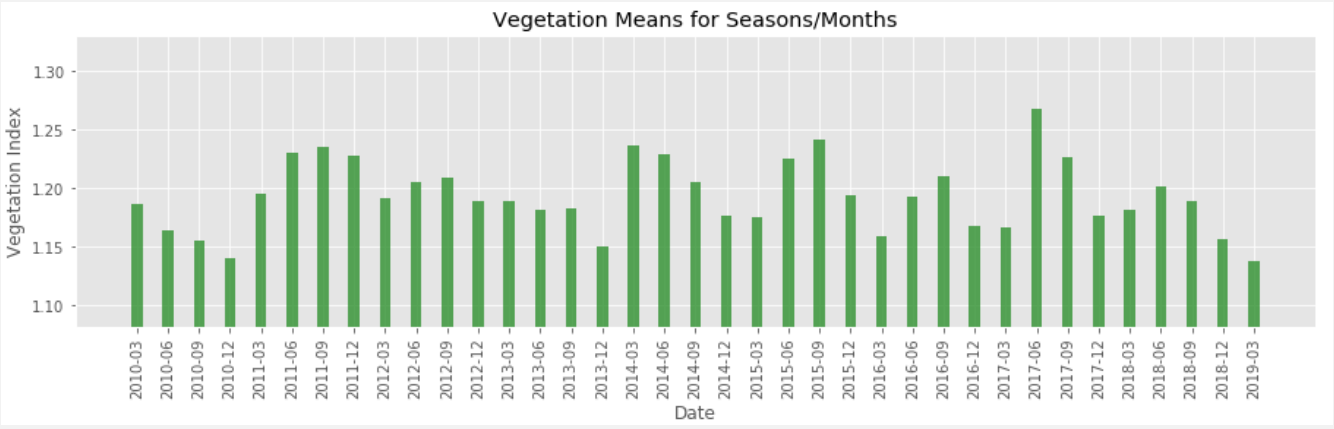
Parameter explanation:

Here, we are doing a seasonal MK test. This time, the MK test will use data from every season of every year, instead of just one specific month or quarter. This will provide the MK test and change point detectors with more data, potentially offering a more accurate trend.

Go to step 02B and re-run the code cell to reset our current datasets and perform a seasonal MK test.

The output graph should look similar to below:

Result explanation:



Notice we now have more data – four dates (seasonal medians) per year.

The date labels on the graph have months of 03, 06, 09, 12. Basically, these are the months of the quarter end (e.g. quarter 1 ends March).

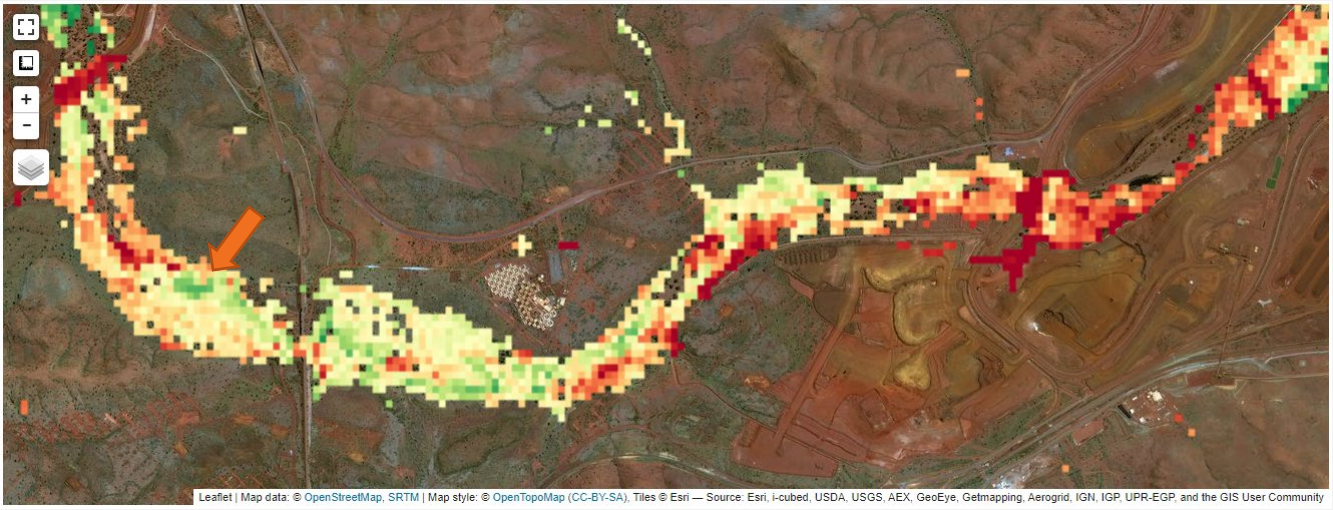
The last date is a single quarter for 2019 – this will be ignored in the MK test. Regardless, we now have more data for trend analysis and change point detection.

Go to step 02C and re-initialise the interactive map and change point detection control panel.

Your interactive map should appear like so:

Result explanation:

Let's explore some trend results at the same general areas as before. Zoom in to the location at the arrow above.

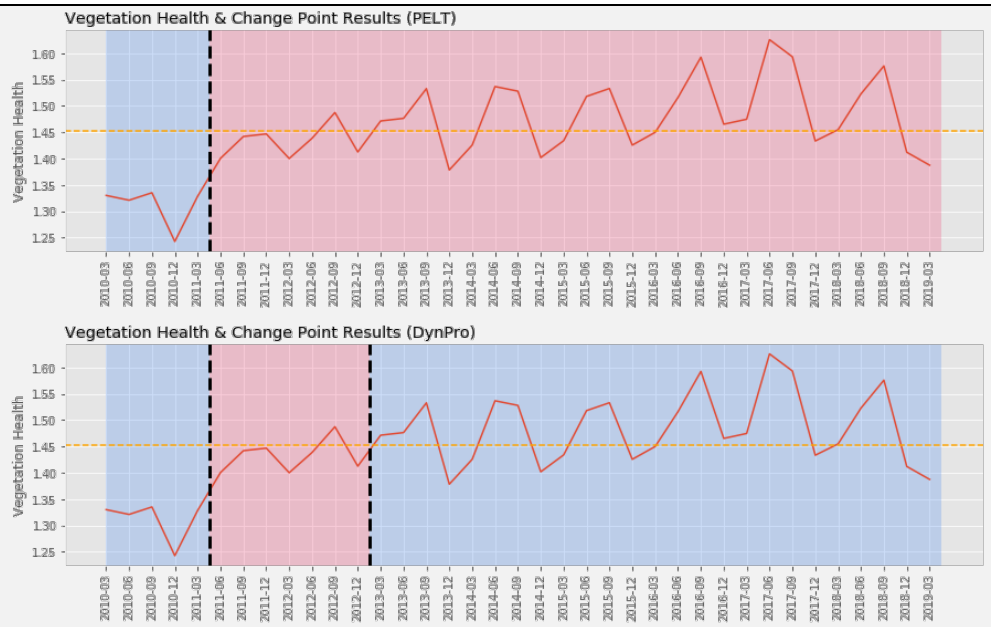


Your screen should look something like above. Click a green pixel near the arrow shown above. You should be seeing a Tau, p-value and Sen's slope value similar to below:

TAU: 0.582 | PVALUE: 0.000 | SENS SLOPE: 0.021

A moderate increasing trend with a highly statistically significant p-value (precise value is 0.0002).

Again, go to the change point detection control pane 'Show' button and click it to perform and display the graphs. They should be somewhat similar to below:



Our vegetation health signatures seem to match the MK trend test result; a moderate increasing trend over time.

The change point detections appear to both have detected a significant break around mid-2011. According to PELT, this shift has lasted onwards from 2011-06 to end of 2018. The Dynamic Programming method provides a little more detail, saying this regime changed again around early-2013.

We can squeeze more detail out of these change point detection methods by modifying some of the control panel parameters.

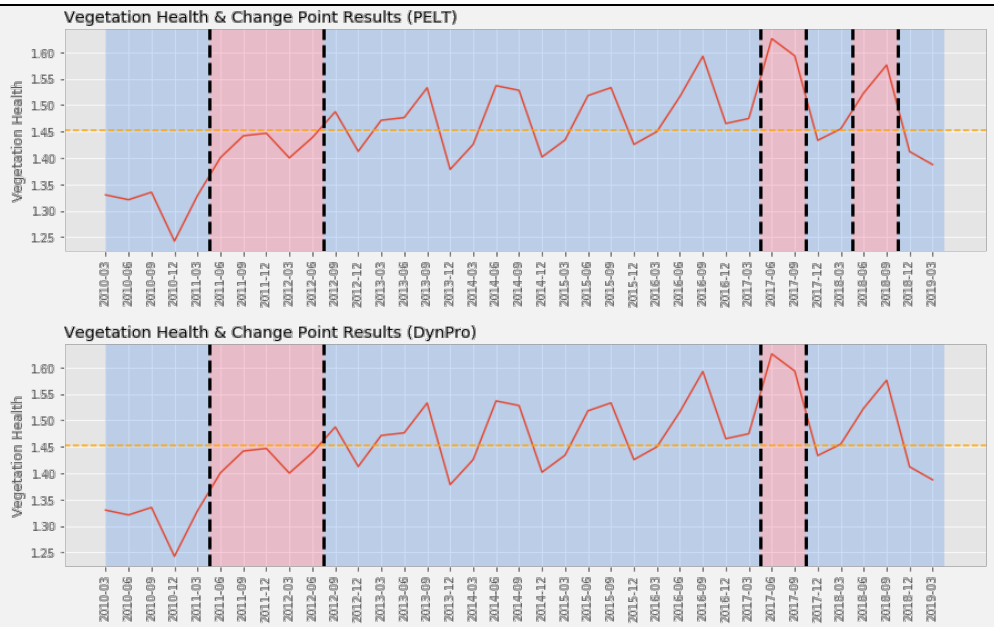
Reminder: the tool provides detailed notes on what each control parameter represents, please read them.

Can we detect more breaks? Let's try and add second set of break points to the change point detection graphs.

In the control panel, change the 'Break Penalty (PELT)' parameter to 1 and the 'Num Breaks (DynPro)' parameter to 4 and the click the 'Show' button again.

It should appear like so:

Result explanation:



We have added a bit more nuance to the change point detection methods by modifying these parameters. We essentially asked the methods to try and detect more than one break, and it could.

Spend some time playing with the other change point detection parameters, such as jump size, to further modify breaks.

One important and helpful aspect of the tool in Section 02 is the ability to visual certain aspects of the MK test results.

For instance, if you return to the control panel in Step 02A and set the 'Show Trend' option to 'Decreasing', we will only see declining trends on our interactive map. It may appear something like this:

Result explanation:

Only decreasing trends are now shown on the interactive map.

This functionality can reduce the time you will spend finding areas that may potentially warrant investigation in the field.

Feel free to spend some time exploring other trends and change points.

When ready, continue to the last section of the tool (Section 03) to perform time-series based spectral change vector analyses.

Section 03: Monitor GDV health changes via spectral change vector analysis

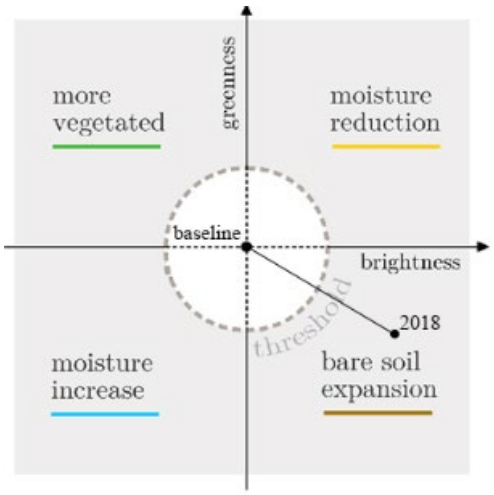
The third and final section of this tool uses time-series spectral change vector analysis to detect different types of plant health change and change magnitude. The intent is to provide an assistive, user-driven, interactive approach to exploring GDV health change over a specific time period and study area boundary.

The resulting change vector layers are coupled with graphing tools that facilitate exploration of vegetation health history anywhere within the study area (i.e. both inside and outside of GDV change areas). Finally, as with Section 02, change point detection (e.g. Andersen 2009) is incorporated to allow users to detect 'break points' in vegetation health history (i.e. dates when vegetation health signatures significantly shifted away from norm).

Spectral change vector analysis (CVA) has been successfully used to derive areas of vegetation and moisture changes, such as vegetation health increase or decrease, from satellite imagery (e.g. Malila 1980). CVA calculates angle and magnitude of change vectors derived from a two-band image, usually generated using Tasselled Cap Transformation (Crist 1985).

CVA is typically used to identify spectral changes between two identical scenes which were acquired at different times. For each pixel between the two images, CVA calculates the change vector in the two-dimensional spectral space. This results in 'angle' and 'magnitude' layer outputs, the former of which has been found to represent different types of vegetation or moisture change depending on angle range.

A figure demonstrating these angles ranges is presented below (image from Zanchetta and Bitelli 2017):



This image demonstrates the angle of change (bearing of arrow) and magnitude (distance of line) calculated from the difference between two images captured on different dates (e.g. a vegetation median baseline vs 2018 vegetation). In this example, the angle has a bearing of approx. 130°, which falls into the 'bare soil expansion' quadrant; in other words, vegetation decline. In contrast, an angle between 270°-359° would be indicative of vegetation increase.

A threshold median factor can also be set in this CVA method. This calculates a threshold magnitude for which pixels are considered stable, i.e. no change, and can be used to control the sensitivity of what magnitudes are returned.

As with Section 02, change point detection (e.g. Andersen 2009) is also incorporated into the tool here in a similar manner and will not be explained again.

The key processing steps applied in this section are as follows:

1. Fetch a 'cube' of Landsat or Sentinel satellite data for a user-defined study area and timespan using the DEA ODC;
2. Generate Tasselled Cap index;
3. Reduce data to baseline and comparison datasets for specified season(s) or month(s);
4. Standardise vegetation using a basic invariant target site approach;
5. Perform spectral CVA for every comparison year against baseline data;
6. Map trend results onto interactive map; and
7. Graph vegetation history and perform change point detection on any pixel clicked by user.

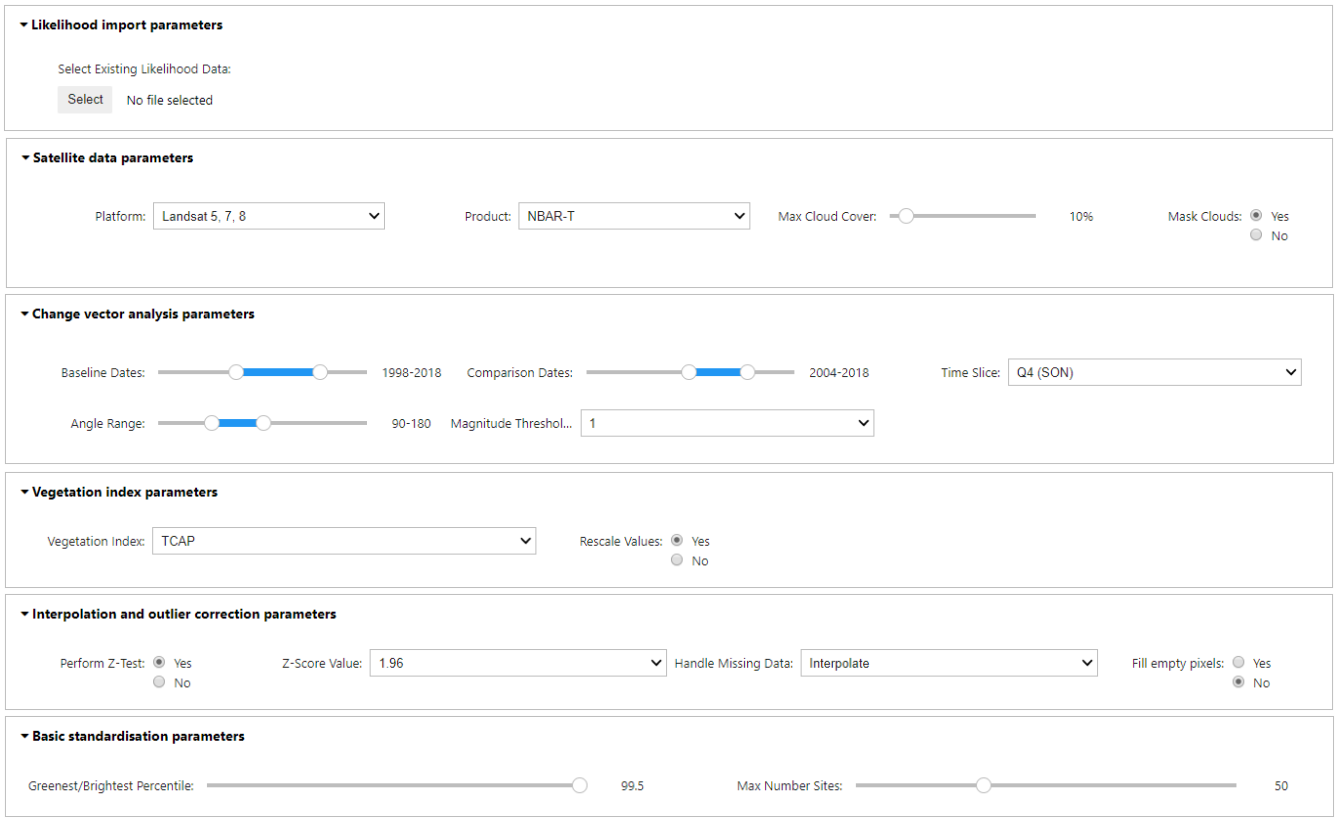
Proceed to step 03A when ready.

Step 03A: Set change analysis parameters

We will now perform spectral change vector analysis within those same thresholded GDV areas derived from Section 01. This should provide insight into vegetation health change over time in a very visual manner.

Before we get into it, run the code cell directly beneath the step 03A notes to reveal the control panel, then spend some time reading the tool notes for an overview of the control panel parameters.

You should be familiar with most of these parameters already from Section 01. Those parameters new to you should be change vector analysis parameters. The control panel should appear as below:



When you are ready to move on, set the control panel parameters to match the following:

- Select existing likelihood data: (*keep blank*)
- Platform: Landsat 5, 7, 8
- Product: NBART
- Max Cloud Cover: 10%
- Mask Clouds: Yes
- Baseline Dates: 1998-2018
- Comparison Dates: 2004-2018
- Time Slice: Q4 (SON, September to November)
- Angle Range: 90-180
- Magnitude Threshold: 1
- Vegetation Index: TCAP
- Rescale Values: Yes
- Perform Z-Test: Yes
- Z-Score Value: 1.96

- Handle Missing Data: Interpolate
- Fill Empty Pixels: No
- Greenest Percentile: 99.5
- Max Number Sites: 50

Parameter explanation:

You have chosen the use a date range of 1998 to 2018 as the vegetation/brightness (hereafter referred to as soil) baseline image used in change analysis. This baseline is essentially a median of all Q4 (Sep-Nov) vegetation and soil imagery calculated using Tasselled Cap Transformation.

You have also set the range of comparison dates to range from 2004-2018; one Q4 vegetation and soil image will be generated for each year in this range and compared vs the baseline, building a library of change maps.

Also, you've decided to only return change angles between 90°-180°, which is indicative of vegetation decline (see above). If you wanted to see only vegetation increase, you would have set the angle range to 270°-360°, or if you wanted to see all angles, 0°-360°.

The final parameter worth explaining is the magnitude threshold value, which we have set to 1. This threshold calculates magnitudes for which pixels are considered stable; i.e., no change, providing some control over the intensity of change magnitude that will be displayed on the interactive map below. As an example, a lower threshold will return more change pixels (lower magnitudes) and a higher threshold would return less (higher magnitudes only).

The other parameters have been covered prior and will not be re-explained. Please investigate the tool notes if you need help.

When ready, proceed to step 03B.

Step 03B: Perform change analysis

It is now time to perform your spectral change vector analysis. This process will compare every greenness and soil image in our comparison range to the same baseline to determine if vegetation has declined significantly from this baseline.

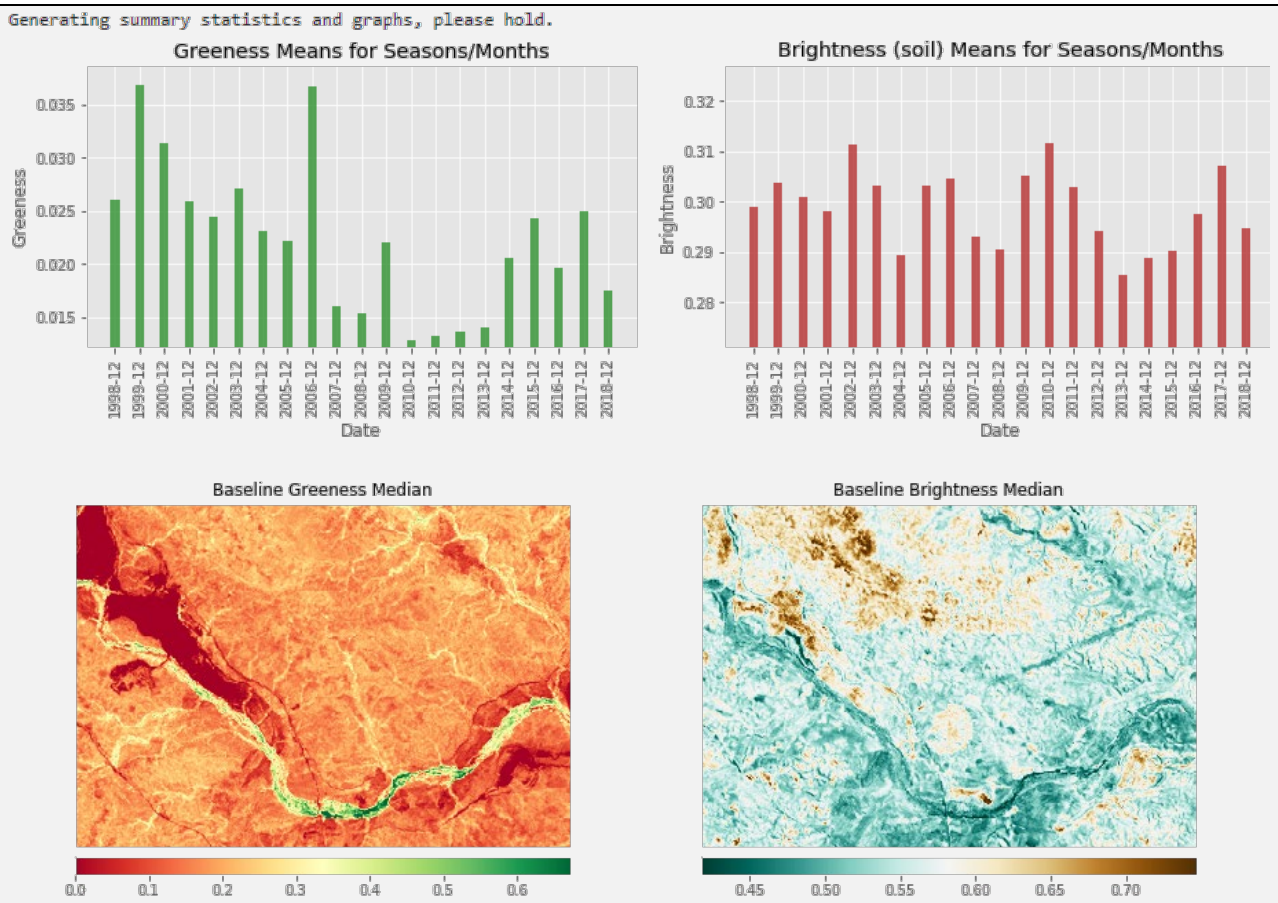
The output of this process will be two layers; the angle and magnitude layers, which will be explained shortly. This process is surprisingly fast.

Please read the tool notes for step 03B and when ready, run the code cell beneath the tool notes to begin the process. No control panel or user-inputs are required.

Eventually you should see a page of summary information with associated graphs. Much of this has been explained in Section 01 and won't be repeated.

You should see various graph, including mean vegetation and soil values and baseline images, like below:

Result explanation:



The first two graphs provide the usual mean greenness and soil values across all dates in our dataset. Under them are two baseline images (median all time) used to compare every other date against.

Your spectral change vector analysis completed successfully. We can also see vegetation and soil means and images tend to show different patterns (i.e. higher vegetation areas have lower soil values and vice versa). What we are looking for are areas where vegetation declined at the same time soil increased.

The next step is to put the results of this process onto an interactive map.

When ready, please proceed to step 03C.

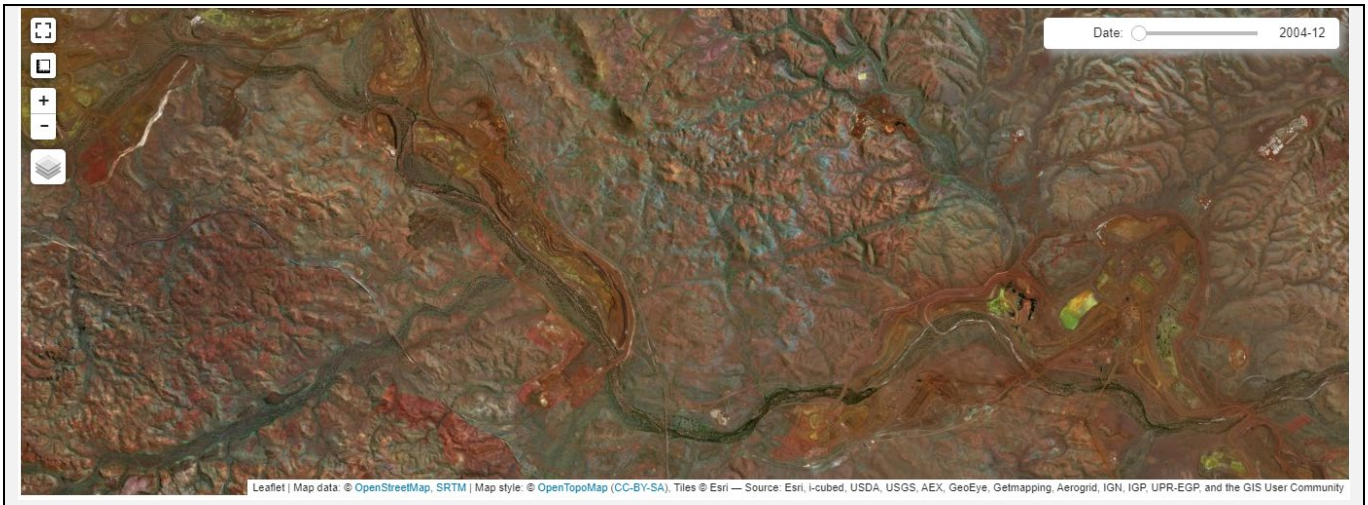
Step 03C: Visualise change analysis results, GDV health graphing tools and change point detection controls

We will now visualise and explore the spectral change vector analysis results on an interactive map and prepare the change point detection control panel.

Please read the tool notes for step 03C and when ready, run the code cell beneath the tool notes to begin the process. This process shouldn't take long to complete.

A familiar interactive map (and control panel) should appear should appear, like so:

Result explanation:



Wait… nothing seems to be showing on this map.

In order to see the result of our change vector analysis, we need to scrub to a date in the slider in the top right of the map, first.

Move the slider over once to 2005-12 to initialise the change and magnitude layers.

It should now look like below:

Results explanation:

There should be a few orange and red change magnitude areas located near the arrow above.

This map contains two layers that may be helpful to you:

Angles layer: the angles output from the change vector analysis method. This contains pixels with angle values. As you selected an angle range of 90º-180º, pixels will only contain angles within this range.

Magnitude layer: Perhaps the most important layer of the two, the magnitude layer provides a value indicating the 'distance' between the baseline and the current comparison image. Lower magnitude are orange and higher magnitude are red.

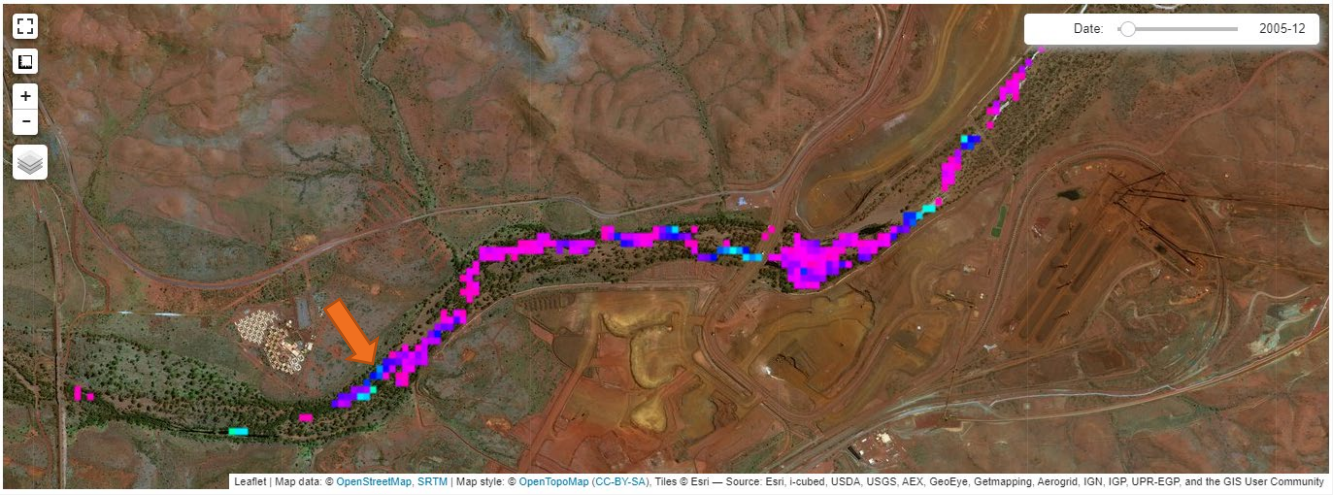
The default layer is the Magnitude layer. You can see that a range of lower magnitude (orange) vegetation decline and some higher magnitude vegetation decline (red) has occurred in 2005 when compared against the baseline.

Let's have a quick look at the angles layer. Turn the magnitude layer off to focus on the angles layer.

It may appear like so:

Results explanation:

Take a closer look by zooming into the pixels at the arrow above to match below.



As with the MK test, you can obtain precise information about change pixels by clicking on them. Try clicking on one of the blue pixels at the arrow above to get precise information.

You should see a print-out similar to below appear under the map:

ANGLE: 127.63 | MAGNITUDE: 18.05

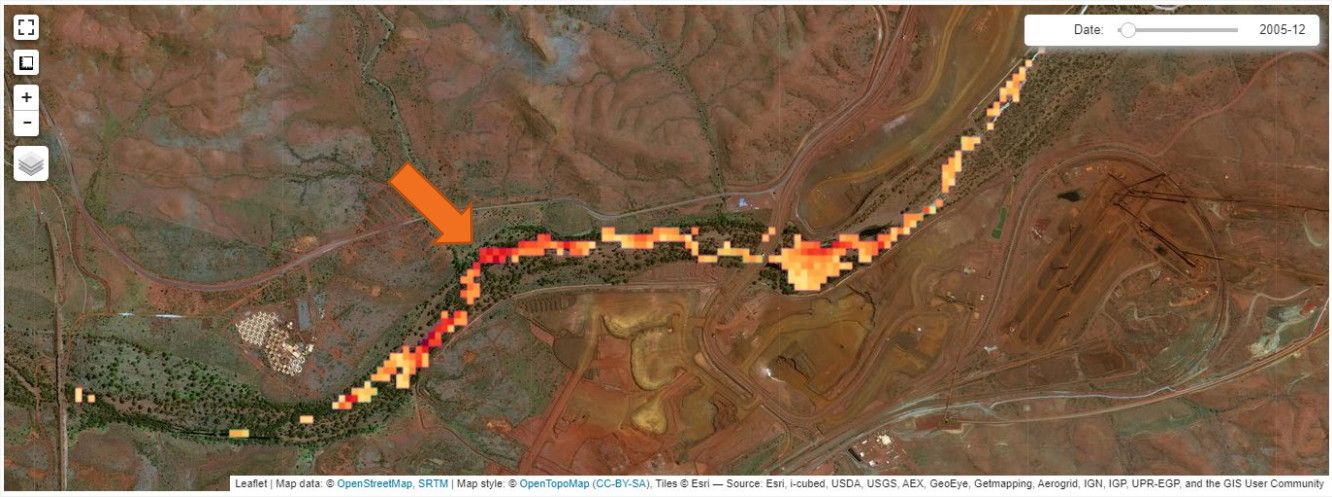
This is quite helpful – it tells us precisely what the change angle is (127.63º).

This is indicative of vegetation decline (or bare soil increase). The magnitude is also relatively high, at 18 magnitude (you rarely reach 50). This implies that vegetation at this location declined during the dry season of 2005.

How does the magnitude layer look closer up? Turn off the angle layer and turn on the angle layer.

It should appear like below:

Result explanation:



Similar to before, click on one of the higher magnitude areas at the arrow above:

ANGLE: 165.07 | MAGNITUDE: 38.27

A much higher magnitude of vegetation decline than the previous area we clicked on.

Does this vegetation decline continue on into the next year, or was it a temporary decline event?

We can answer questions like this quite visually by sliding the time slider right by a year or two.

Try sliding right to 2006, like so:

Result explanation:

Is this area still relatively dense in vegetation decline pixels of high magnitude? Not really, although a few decline areas in the vicinity can be seen:

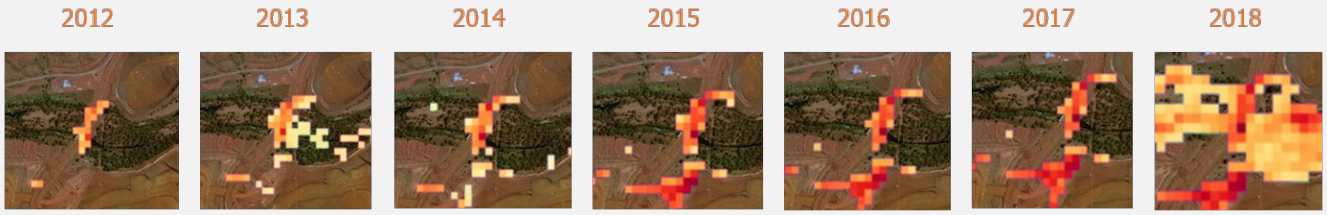
These vegetation decline areas return again in 2007, however:

Perhaps these areas would be worth monitoring via field survey.

What about the haul road to the east (see arrow)?

Can we detect when that infrastructure replaced vegetation, and if so, does that change appear consistently after clearing?

Slowly move the date slider from left to right, from 2010 to 2018, and keep an eye on the change magnitude in the vicinity of the haul road. The results should appear as so:



It appears that, in comparison to our vegetation and soil baseline, heavily cleared vegetation appears to maintain some consistency of decline magnitude.

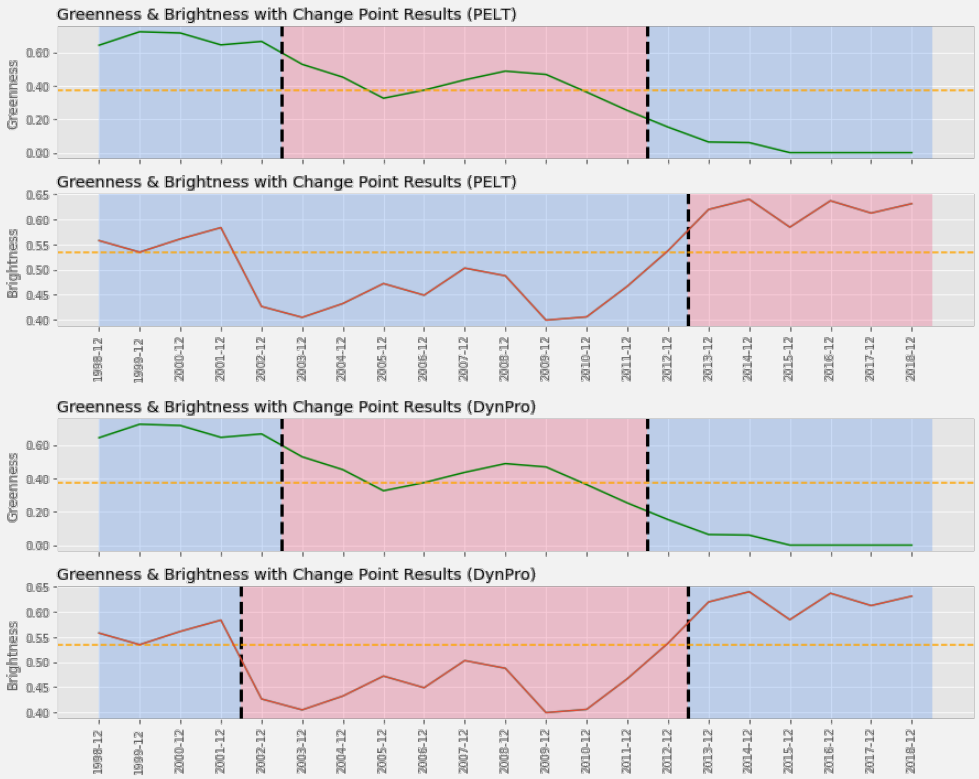
It also appears that the haul road was possibly constructed around 2012. We can use the graphing tools and change point detection to explore this deeper.

Click on one of the high magnitude pixels in the centre of the haul road.

Then, go to the change point detection control panel below the interactive map, and click the 'Show' button.

You should see an output similar to before with a few differences:

Result explanation:



We now have two additional graphs.

The top two graphs are PELT change point detection results for the Tasselled Cap-derived vegetation band (green line) and brightness (or soil) band (brown line) for each comparison year.

The bottom two graphs are the same, but show Dynamic Programming change point detection results.

Looking at the top two graphs (PELT) we can see a continuous decline in vegetation starting around mid-2002, with a minor increase around 2009, then a continuous decline from 2009 onwards.

This decline doesn't dip beneath the median (the baseline) until around 2011, and is then quite far away from the median by 2012. This matches the change vector analysis result.

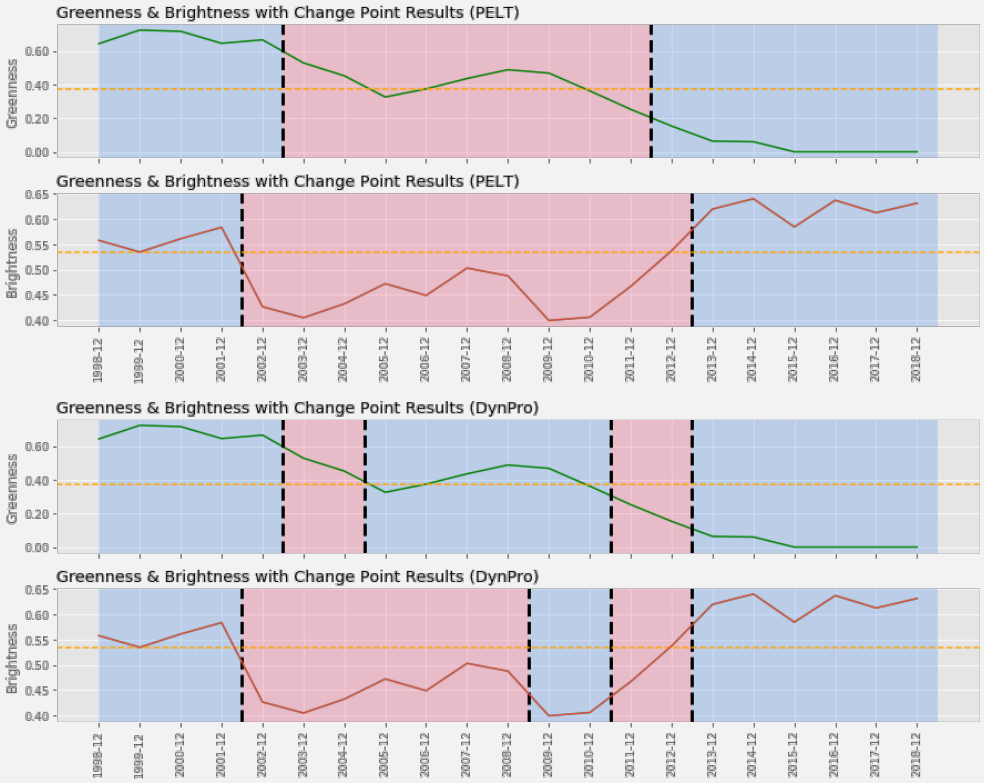
That said, vegetation doesn't tell the entire story. Notice the soil line is also increasing quite sharply from 2012 onwards? It is this declining vegetation coinciding with increase soil that appears to have been the real trigger on the CVA layers.

Either way, we can see after 2012, greenness remains down and soil remains up.

Try to reduce the PELT penalty value to 1 and the DynPro break number to 4 to add more breaks to these graphs.

Press the 'Show' button again and your plots should refresh, like so:

Result explanation:



You can see the Dynamic Programming method displays additional breaks, but PELT did not.

It appears PELT has found the most optimal breaks available to it.

We can see with the brightness breaks based on PELT that a break (black dashed line) occurred around 2012 also, ending the previous soil regime.

Take time to explore the rest of the map.

Some ideas: try different baselines ranges, such as a more recent date range (e.g. 2008-2018) and different time slice (e.g. wet season or Q1).

Likewise: increase the magnitude threshold to 1.5 to reduce the amount of magnitude pixels returned, eliminating lower magnitude and potentially insignificant vegetation decline across your comparison years.

Conclusion

This concludes the GDV Tool workshop. Thank you for your time.

References

Andersen, T., Carstensen, J., Hernandez-Garcia, E. and Duarte, C.M., (2009). Ecological thresholds and regime shifts: approaches to identification. Trends in Ecology & Evolution, 24(1), pp.49-57.

Arlot, S., Celisse, A., and Z. Harchaoui. (2012). Kernel change-point detection. arXiv preprint arXiv:1202.3878, 1(0000):1–26, 2012.

Hamed, K. (2008). Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. Journal of Hydrology, 349(3-4), pp.350-363.

Hipel, K.W. and McLeod, A.I., (2005). Time Series Modelling of Water Resources and Environmental Systems.

Hussain, M. and Mahmud, I., (2019). pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. Journal of Open Source Software, 4(39), p.1556.

Malila, W.A., (1980). Change vector analysis: an approach for detecting forest changes with Landsat. In LARS symposia (p. 385).

Mann, H.B. (1945). Nonparametric tests against trend, Econometrica, 13, 245-259.

Snedecor, G.W. (1956). Statistical methods: Applied to experiments in agriculture and biology. The Iowa State College Press.

Truong, C., Oudre, L. and Vayatis, N., (2019). Selective review of offline change point detection methods. Signal Processing, p.107299.

Zanchetta, A. and Bitelli, G., (2017). A combined change detection procedure to study desertification using opensource tools. Open Geospatial Data, Software and Standards, 2(1), p.10.