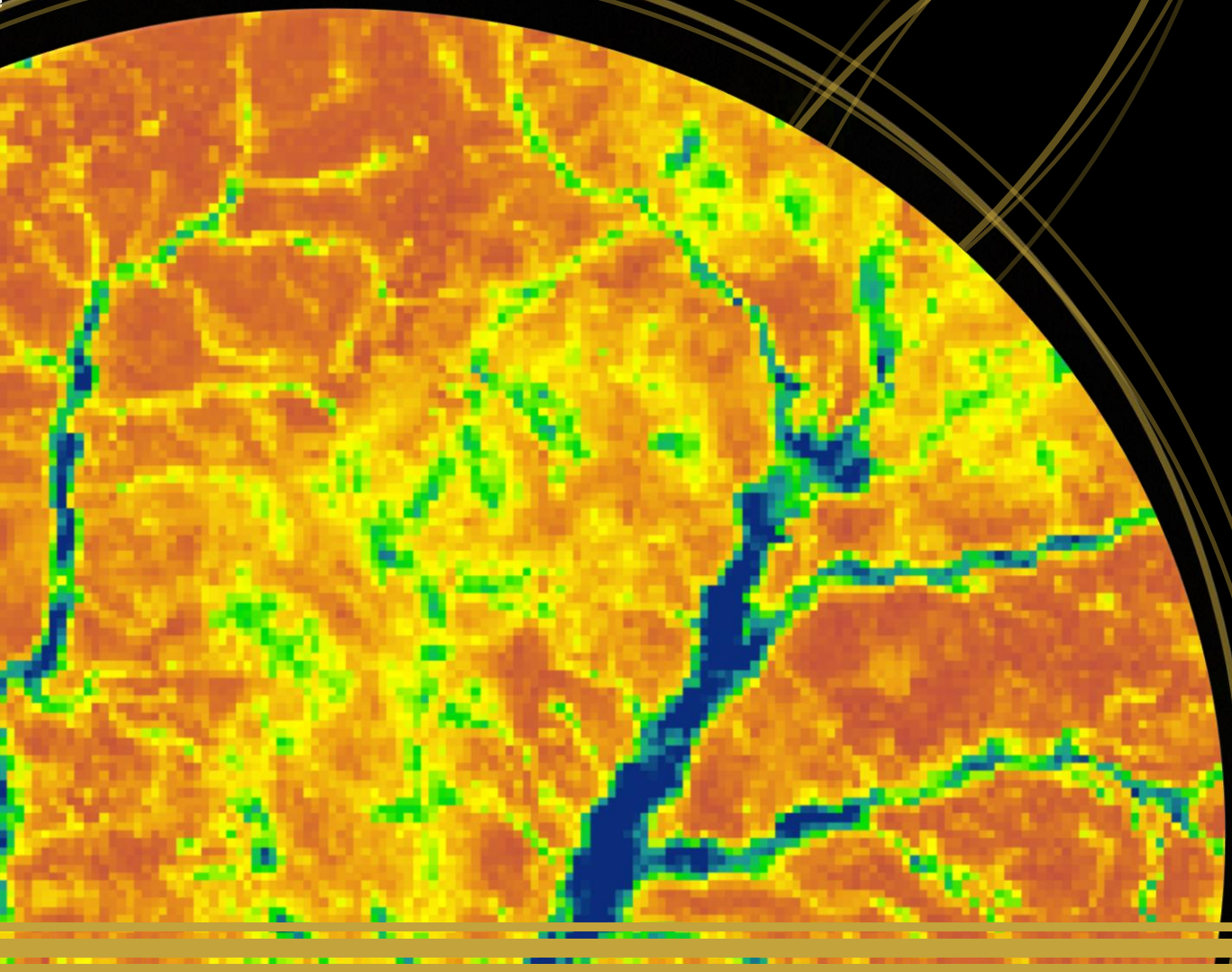
Curtin University

Ensemble Modelling for Ground Water Dependent Vegetation Detection



TECHNICAL REPORT

Technical Report

Ensemble Modelling for Groundwater Dependent Vegetation Detection

Research Associates

L. Trotter

A. Cross

Research Lead

T.P. Robinson

May 2022

Curtin University

Contents

1 Introduction 5
2 Methods and Materials 8
2.1 Datasets 8
2.2 Phenometrics 9
2.2.1 Characterisation of Pilbara GDV 9
2.2.2 Phenometrics and Variable Selection 9
2.2.3 Phenometrics and Variable Selection 10
2.3 Species Distribution Modelling 11
2.3.1 Topographic Derivatives 12
2.3.2 SDM Creation 13
2.4 Fractional Cover Mapping 14
2.5 Ensemble Model 15
2.5 Near Real-time Monitoring 18
3 Results 19
3.1 Phenometrics 19
3.1.1 Characterisation of Pilbara phreatophytes 19
3.1.2 Phenometrics and Variable Selection 19
3.2 Species Distribution Modelling 22
3.3 Fractional Cover Mapping 24
3.4 Ensemble Modelling 25
3.5 Near Real Time Monitoring 26
4 Conclusions 29
REFERENCES 30
Appendix 1 - Tutorial 33
1 Introduction 34
1.1 Requirements 35
1.2 Setup and installation 35
1.3 Tutorial data 35
1.4 Known issues 36
1.5 Tutorial preparation 36
2 COG Fetch tool 39
2.1 Introduction 39
2.2 Use the COG Fetch tool 39
2.2.1 Considerations 39
2.2.2 Initialise the tool 40
2.2.3 Download satellite data 40

3 COG Explore tool 42
3.1 Introduction 42
3.2 Use the COG Explore tool 42
3.2.1 Considerations 42
3.2.2 Initialise the tool 42
3.2.3 Visualise satellite data 43
3.2.4 Visually explore MAVI over time 43
3.2.5 Visually graph MAVI over time 44
4 GDV Spectra 46
4.1 Introduction 46
4.2 Generate GDV likelihood layers 47
4.2.1 Considerations 47
4.2.2 Initialise the GDV Likelihood tool 47
4.2.3 Generate likelihood layers (non-combined) 47
4.2.4 Explore the non-combined likelihood layer 48
4.2.5 Generate likelihood layers (combined) 50
4.2.6 Explore the combined likelihood layer 50
4.3 Threshold GDV likelihood layer 51
4.3.1 Considerations 51
4.3.2 Initialise the tool 51
4.3.3 Threshold GDV likelihood layers (non-combined) 52
4.3.4 Explore the thresholded GDV likelihood 53
5 Phenolopy 54
5.1 Introduction 54
5.2 Use the Phenolopy tool 54
5.2.1 Considerations 54
5.2.2 Initialise the tool 55
5.2.3 Generate phenological metrics 55
5.2.4 Explore the various metrics 57
6 Nicher 58
6.1 Introduction 58
6.2 Use the Nicher tool 58
6.2.1 Considerations 58
6.2.2 Add and explore provided data 59
6.2.3 Initialise the tool 61
6.2.4 Generate species distribution model 61
6.2.5 Explore the Nicher SDM 63

6.2.6 Assess model performance ... 64
7 VegFrax ... 67
7.1 Introduction ... 67
7.2 Use the VegFrax tool ... 67
7.2.1 Considerations ... 67
7.2.2 Add and explore the classified GeoTiff ... 68
7.2.3 Initialise the tool ... 69
7.2.4 Generate the class fractional covers ... 69
7.2.5 Explore the VegFrax fractional cover maps ... 71
7.2.6 Assess model performance ... 72
8 Ensemble ... 73
8.1 Introduction ... 73
8.2 Prepare evidence layer inputs ... 73
8.2.1 Considerations ... 73
8.2.2 Initialise the Ensemble Sigmoider tool ... 73
8.2.3 Rescale evidence layers ... 73
8.3 Generate ensemble model ... 74
8.3.1 Considerations ... 74
8.3.2 Initialise the Ensemble Model tool ... 75
8.3.3 Generate ensemble model ... 75
8.3.4 Explore the ensemble outputs ... 76
9 NRT Monitoring ... 78
9.1 Introduction ... 78
9.2 Switch to new study area ... 78
9.3 Preliminary vegetation change assessment ... 79
9.3.1 Initialise and run the COG Fetch tool ... 79
9.3.2 Initialise and run the COG Explore tool ... 81
9.3.3 Zoom to impact area ... 83
9.3.4 Graph the vegetation history ... 83
9.4 Detect change via Change Vector Analysis (CVA) ... 84
9.4.1 Introduction ... 84
9.4.2 Considerations ... 85
9.4.3 Initialise the tool ... 85
9.4.4 Perform Change Vector Analysis ... 86
9.4.5 Visualise vegetation decline change ... 87
9.5 Create a new monitoring project ... 88
9.5.1 Considerations ... 88

9.5.2 Initialise the tool ... 88
9.5.3 Create a new project ... 89
9.6 Create new monitoring areas ... 89
9.6.1 Considerations ... 89
9.6.2 Initialise the tool ... 89
9.6.3 Create monitoring area ... 90
9.7 Begin monitoring areas ... 94
9.7.1 Considerations: ... 94
9.7.2 Initialise the tool ... 94
9.7.3 Begin monitoring areas ... 95
9.7.4 Observe the monitoring tool process ... 95
9.8 Graph monitoring area ... 97
9.8.1 Considerations ... 97
9.8.2 Select the monitoring area ... 98
9.8.3 Explore monitor area graphs ... 99
9.9 Visualise monitoring area change ... 101
9.9.1 Introduction ... 101
9.9.2 Considerations ... 102
9.9.3 Initialise the tool ... 102
9.9.4 Visualise all-time change ... 102
9.9.5 Visualise vegetation and change densities ... 103
9.10 Extra: monitor area with dynamic method ... 105
9.10.1 Explore the dynamic change graph ... 106
10 References ... 108
11 Appendix ... 110
11.1 Summary of phenometrics ... 110
11.2 Rule one (number of consecutive and plateaus) ... 112
11.3 Rule two (minimum allowed zone) ... 113
11.4 Rule three (number of zones “jumped”) ... 114
11.5 Rulsets ... 115
11.6 Alert method ... 119
11.7 Direction of change for trigger alert ... 121

1 Introduction

Groundwater dependent vegetation (GDV) relies on a permanent and stable water table for a significant portion of their water requirements. Dewatering of mine pits to access ore can cause the water table to decline beyond their root systems, potentially resulting in mortality. Ground-based surveys of vegetation health are important in this process, and these can be assisted by earth observation (EO). Analysis ready datasets (ARD) of EO imagery can be used to map GDV at regular intervals to detect change and direct ground-based surveys to investigate potential areas of concern.

Eucalyptus victrix, E. camaldulensis and Melaleuca argentea are three common tree species referred to as phreatophytes (or more generally "GDV species") in the Pilbara region of Western Australia. However, they generally have differing levels of reliance on groundwater. *E. victrix* is a facultative phreatophyte that utilises stored soil water reserves and may utilise groundwater to satisfy part of its water requirements. *E. camaldulensis* and *M. argentea* are both obligate phreatophytes, meaning access to groundwater is critical for survival. *M. argentea* have shallower root systems than both *Eucalyptus* spp. and so are usually the first to show water stress if groundwater depth increases.

Healthy specimens of these GDV-species are perennially green with moist foliage relative to coexisting species (O'Grady et al., 2011; Barron et al., 2014). It is this characteristic of "green persistence" that can be capitalised on for detection and mapping from EO imagery. Earlier research by Trotter et al. (2020) using a knowledge driven spatial multicriteria approach has shown that the concept of green persistence could narrow the search space for GDV in the Pilbara bioregion of Western Australia. Their approach used just two EO images per year based on the premise that if vegetation was vigorous, green, and moist in the wet season and in the dry season then it was likely to be GDV. Other vegetation is likely to be less vigorous all year round (e.g., mulga) or very green in the wet season but senescent in the dry season (e.g., grass species).

Given the discrimination potential provided by the green persistence approach, it is hypothesised that an entire annual stack of EO imagery could further discriminate GDV from coexisting non-GDV. Repeat observations through time are powerful for characterising and quantifying ever-changing processes such as phenological events

(Reed et al. 2009). These include greening, senescence, flowering, leaf fall and leaf development. Phenological metrics (or phenometrics as used herein) are metrics that can be related back to many of these events in a plant's life cycle.

Much of the early (and still current) use of phenometrics were paddock-scale applications such as crop type mapping and productivity estimates (e.g., Sakamoto et al., 2013; Araya et al. 2015; and Brown et al., 2013). Kramer et al. (2000) used observations of later than usual commencement of seasonal green-up to provide greater insight into changing climatic patterns. Van Leeuwen et al. (2010) used productivity to discriminate between different vegetation communities.

Most studies utilise the 250 m spatial resolution of MODIS prepared using 8 or 16-day maximum value composite (MVC) imagery to mitigate cloud cover thereby ensuring no bias from missed, poor quality, observations. The desirable features of higher resolution for phenology mapping (e.g. Melaas et al., 2013) has led some researchers to downscale MODIS imagery via fusion with lower frequency but higher spatial resolution Landsat and Sentinel imagery (e.g. Gao et al. 2017, Onojeghuo et al. 2018; Schreier et al, 2021). However, ideally higher spatial resolution imagery could be used without fusion techniques. The Open Data Cube (ODC) is an innovation by Geoscience Australia and partners to harness the power of satellite imagery that is sun-synchronous – captured at regular intervals. The ODC includes a set of programmatic libraries as well as a backend database of Landsat, Sentinel and MODIS imagery that are corrected for various attenuations such as atmospheric and topographic effects. Here we test the potential of Sentinel imagery from the ODC for computing phenometrics. This imagery is processed to ensure an image is available consistently every 7 days and is supplied with a spatial resolution of 10 m in the visible and near infrared regions of the spectrum

Species distribution models (SDM) correlate spatial predictors (e.g., slope of terrain, incoming solar radiation) with species occurrences (Guisan and Thuiller, 2005). SDMs, by themselves, are powerful tools for mapping a species or a set of species ecological niches in locations not yet fully explored. However, their aim in this research is to add context to the phenometrics – for example, if a highly productive species is also found in GDV habitat then it improves the likelihood of it being GDV. In contrast, highly

productive vegetation (e.g., irrigated fields) may be found outside of GDV habitat, which can then be down weighted in the final model.

This study was undertaken in collaboration with several prominent mining companies and government departments. The steering committee consisted of members from Rio Tinto, BHP, Roy Hill, FMG, Atlas Iron, DWER and DBCA. All mining companies have mining tenements in the vicinity of the southeast Pilbara region study area. Most mining tenements have been mapped in whole or in part using fine scale imagery like aerial photographs, other airborne imagery (e.g., DMSI) or more recently from satellite imagery (e.g. WorldView 2/3). This finer scale mapping can be exploited using coarser imagery with larger extents. Fractional cover mapping is one method that makes use of an image that has higher resolution relative to a second image. We use it in this research to extrapolate existing mapping to derive a layer representing the fraction or percentage or GDV in each pixel at the tenement scale. Ensemble modelling is proposed to integrate these three sources of evidence (phenometrics, SDM's and fractional cover) and to provide a measure of model uncertainty.

Once GDV has been detected it requires monitoring for deviations from a baseline representative of a pre-impact period. Change detection methods have progressed to incorporate time series of imagery by fitting harmonic regressions to vegetation index transformations. Breakpoints in the harmonic are used to trigger an alert that something has changed. The larger and more prolonged the deviation from the initial harmonic model, the greater the change at that pixel (Brooks et al., 2014).

This project aims to develop validated remote sensing-based methods for detecting and monitoring GDV in the Pilbara bioregion of Western Australia. It has the following objectives:

1. Summarise temporal sequences of imagery using phenometrics based on calibrated 10 m resolution Sentinel imagery.
2. Identify the suitable GDV habitat.
3. Extrapolate existing layers of mapped GDV into a fractional cover map.
4. Integrate all layers into an ensemble model that incorporates uncertainty.
5. Monitor GDV locations for change over time relative to a benchmark.

2 Methods and Materials

2.1 Datasets

We assembled 5865 data points of the three key GDV species from an alliance of five mining stakeholders in the Pilbara region (Figure 1). As each survey was taken at different times, by different people and with differing equipment, we inspected each point and aligned to tree crowns shown on the basemap imagery provided in ArcGIS. Where there was ambiguity or duplicates of the same plant, we deleted the record from the master database. We separated the points into eight discrete groups using OPTICS density-based clustering (Kriegel et al., 2011) with minimum number of features set to 30 and sensitivity set to 5 to emphasise fewer but bigger clusters (Figure 1). This allowed for us to test for differences between suitable phenometrics across the study area.

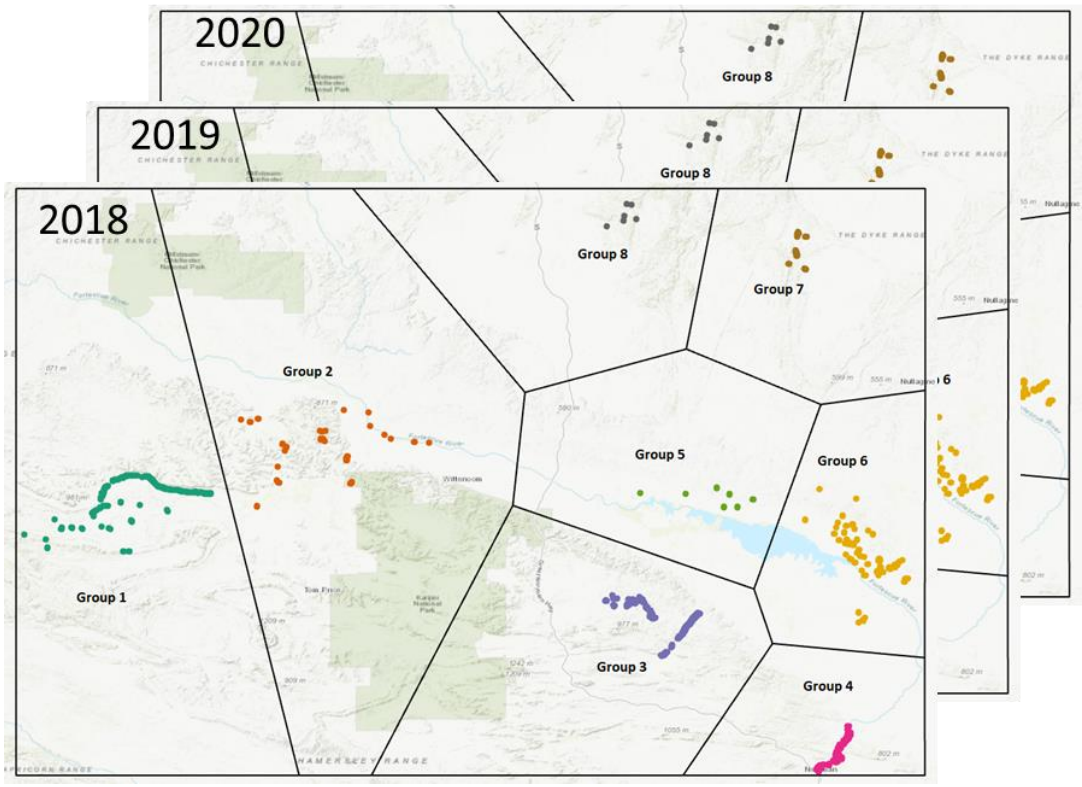


Figure 1 – Data points (5865) collected over the Pilbara region of Western Australia were used for training and testing model development. To determine if model inputs were spatially and temporally stable, the points were separated into eight groups and tested using imagery acquired over three years (2018-2020).

2.2 Phenometrics

To identify a phenometric-based model for GDV prediction that is stable across space and time we first characterised their phenology, extracted a suite of phenometrics, and then explored discrimination potential by comparing area under the curve (AUC) statistics from receiver operating characteristic (ROC) curves for each. The ROC curve is a graph of the false positive rate (x-axis) against the true positive rate (y-axis). and the AUC is a summary performance measurement that is calculated using the trapezoidal rule of the area under the ROC (Pontius and Schneider, 2001). In our implementation, the AUC measures how much separability there is between GDV and background samples (other land covers). The higher the AUC, the better the separability, with 1 representing perfect separation, with 0.5 suggesting separation is no better than random (Fielding and Bell, 1997). We hypothesised that healthy GDV would remain greener for longer periods of the year relative to coexisting shrubs and grasses and thus phenometrics related to annual productivity would best assist in their discrimination.

2.2.1 Characterisation of Pilbara GDV

The moisture adjusted vegetation index (MAVI) proposed by Zhu et al. (2014) was computed from Sentinel 2A/B imagery acquired from the Open Data Cube for 2018, 2019 and 2020. To characterise and explore the differences between GDV and background (a random sample within each group equalling the presence point sample size)over space and time we extracted MAVI for pixels coinciding with tree observations (Figure 1) over the three years and plotted the mean for each day-of-year sampled against the study area mean. This enabled observation of the behaviour of GDV over both space and time and with respect to the regional mean allowing quick determination of locations that were underperforming.

2.2.2 Phenometrics and Variable Selection

We used custom written python code to generate 18 phenometrics, which are illustrated in Figure 2. The MAVI index was resampled to a weekly frequency using a rolling median approach consisting of three images. Each pixel was sampled over the

year and smoothed by fitting a Savitzky-Golay smoother (Figure 2). Outliers were filtered using a median threshold with a factor of 2, which limits removal to only the most significant. Both clouds and outliers are post-filled using an interpolation approach. The start of season (SOS) and end of season (EOS) were based on the seasonal amplitude (AOS), which is calculated from the base level (BSE) and the maximum value for each season (POS). The SOS occurs when the left part of the smoothed curve reaches 50% of the AOS.The EOS is the same but defined for the right side of the curve. Other options could be implemented and are documented in Eklundh and Jönsson (2017). Definitions for all other metrics are provided in Table 1.

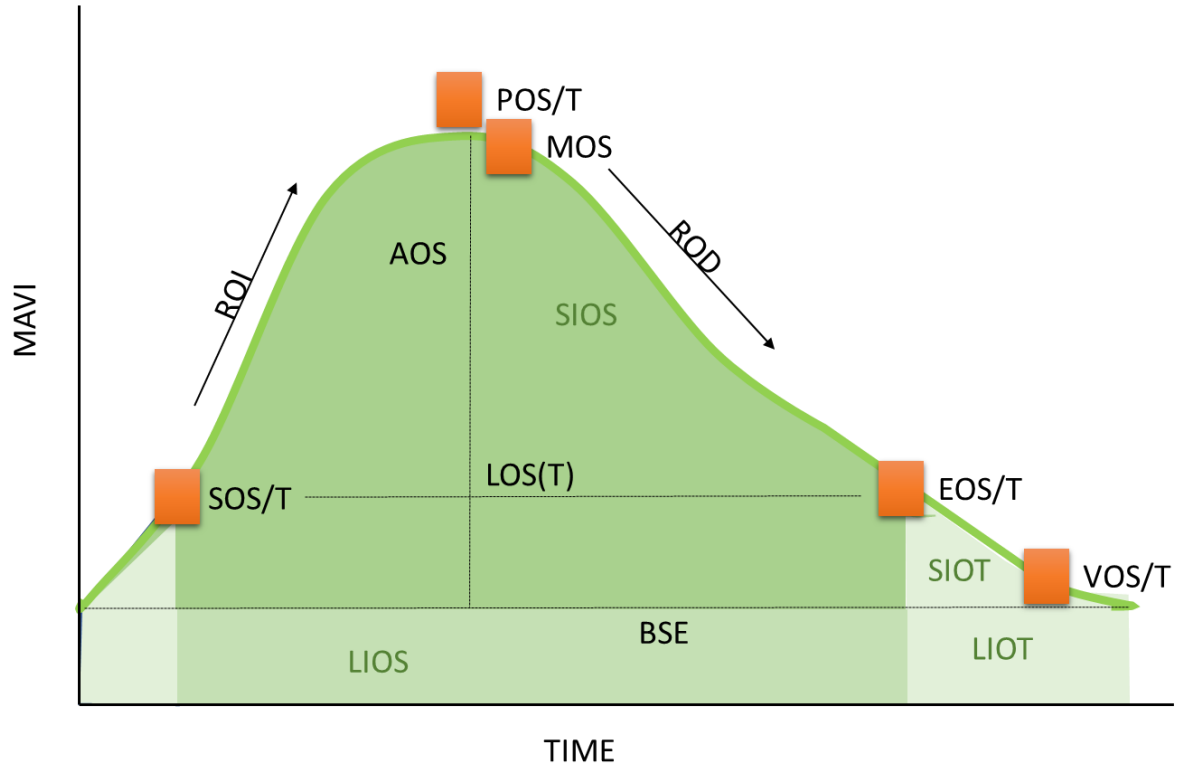


Figure 2 A smoothed time-series of one hypothetical pixel showing a typical MAVI response over one year of samples. Annotated are the 18 phenometrics extracted. See Table 2.1 for a definition of the abbreviations.

2.2.3 Phenometrics and Variable Selection

To determine the most useful variable or variables for predicting GDV habitat we partitioned the dataset into 80% training and 20% testing. The testing dataset was used to perform a jackknife of the AUC, which is used to identify variables that have important individual effects, i.e., not in combination with other variables (Elith et al.,

2010). We used clustering analysis to determine which variables offered similar levels of discrimination over all groups over the three study periods (2018-2020).

Table 1. Summary of each phenological metric generated provided in alphabetical order. Phenometrics with /T have both a MAVI value and a time value variant. Those with (T) are only measured in time.

Code	Name	Description
AOS	Amplitude of Season	The amplitude of VI values for season.
BSE	Base	Mean of the lowest VI values on the left and right of peak value.
EOS/T	End of season	VI at the end of season and time this occurred in day of year
LIOS	Long Integral of Season	Represents the total productivity of vegetation when in season.
LIOT	Long Integral of Total	Represents the total productivity of vegetation throughout the season.
LOS(T)	Length of Season	Length of time (number of days) between the start and end of season.
MOS	Middle of Season	Mean vegetation value and time of values in top 80% of season.
POS/T	Peak of Season	Highest VI value and time this occurred in day of year.
ROD	Rate of Decrease	The rate of vegetation "green down" at the end of season.
ROI	Rate of Increase	The rate of vegetation "green up" at the beginning of season.
SIOS	Short Integral of Season	Represents the seasonally active vegetation.
SIOT	Short Integral of Total	Represents total vegetation productivity throughout the season.
SOS/T	Start of Season	VI value at the start of the season and time this occurred in day of year.
VOS/T	Valley of Season	Lowest vegetation value and time this occurred in day of year.

2.3 Species Distribution Modelling

Habitat modelling, also known as Species Distribution Modelling (SDM), and Ecological Niche Modelling (ENM), attempts to identify habitats over broad areas based on exemplars (e.g. points of the target) and environmental correlates. A variety of machine learning techniques have been applied (see Lorena et al. (2011) for an overview of nine techniques), including the widely used Random Forest regression technique. The Random Forest algorithm is a supervised learning method that can be used for regression or for classification. Under regression mode the algorithm creates multiple decision trees and averages them resulting in a layer from 0 to 1, where 1 is high likelihood of preferential habitat and 0 is not. Under classification it takes the mode of all decision trees, resulting in a layer showing either presence (1) or absence (0) of the target. Both methods improve the accuracy and mitigate overfitting issues found by using singular decision trees (Friedman et al., 2001). Here, we utilise regression mode. Each decision tree constructed during model training is derived from the input species occurrence locations and the coinciding environmental variables values at each location.

2.3.1 Topographic Derivatives

Topographic derivatives representing environmental data can be generated from remotely sensed digital elevation model (DEM) data provided from platforms such as the Shuttle Radar Topographic Mission (SRTM) or airborne light detection and ranging (LiDAR) sensors and used as input variables. Correlates between these variables and known GDV locations can be used to model the potential habitats of these species at the landscape scale.

To describe the topographic characteristics of GDV species, LiDAR data for various mining tenements in the Pilbara region were obtained from tenement owners. Ground (last) returns were interpolated into a 10 m raster grid using Binning interpolation with mean cell aggregation to create a digital terrain model (DTM), which matched the resolution of the Sentinel imagery being used. Raster surfaces were generated for nine preselected topographic variables from the LiDAR-based DTM based largely on Robinson et al. (2019). These nine variables were aspect-eastness, aspect-northness, curvature, dissection, slope, solar radiation, topographic position index (TPI), topographic wetness index (TWI) and topographic ruggedness index (TRI). These variables were chosen for their influence on the distribution of moisture. In addition, the raw DTM was included as a variable (i.e., elevation above sea level).

Aspect, curvature, slope and solar radiation were calculated using tools available in the Spatial Analyst Toolbox of ArcGIS Pro 2.7 (ESRI, 2021). Curvature values of 0 suggest flat terrain, whereas negative and positive curvature values indicate convex and concave terrain, respectively. Slope represents the inclination of slope in degrees ranging from 0 to 90. Aspect-northness and aspect-eastness were calculated by first transforming aspect degrees into radians and then taking the sine and cosine, respectively. Values for aspect-eastness and aspect-northness range from -1 to 1, representing west to east and north to south facing pixels, respectively (Samis & Eckert, 2009). Dissection, which is a measure of how incised the landscape is at each cell, was calculated based on Evans (1972). Solar radiation represents the amount of solar radiation received from the sun, considering the target geographic area, atmosphere, site latitude, elevation, sun angle and shadow of the landscape (Fu & Rich 2002). The output raster represents the total amount of incoming solar radiation in watt hours per square metre (WH/m^2) per annum.

Calculation of the TPI, TRI and TWI were achieved using python within ArcGIS Pro. The TPI compares the elevation of each cell in a DEM to the mean elevation within a specified neighbourhood around that cell and was calculated using a 30 × 30 m neighbourhood size. Positive TPI values represent locations that are higher than the average of their neighbourhood window (e.g., ridges), negative values are lower (e.g., valleys), and flat areas are close to 0 (Guisan, Weiss & Weiss, 1999). The TRI was used to quantify terrain heterogeneity and was based on Riley, De Gloria and Elliott (1999). The TWI was used to quantify topographic control on hydrological processes and was based on Gessler et al. (1995).

Finally, a canopy height model (CHM) was derived from the LiDAR data to demonstrate its potential application as a vegetation mask and SDM model variable. First returns were interpolated into an DSM (digital surface model) via the same method used to derive the DTM described earlier. The CHM was calculated by subtracting the DTM from the DSM (Lisein et al., 2013) and canopy heights <1 m were used to mask out non-GDV.

2.3.2 SDM Creation

We wrote custom python code (collectively known as “Nicher”) to generate Random Forest-based SDMs from user-provided plant locations and topographic derivatives for GDV characterisation in the Pilbara bioregion of Western Australia. Nicher employs the Scikit-learn machine learning library (Pedregosa et al., 2011) and is designed to accept topographic derivatives that have been pre-derived from DEM data, but currently it does not generate such variables itself. This is to cater for more broad use than just GDV species, where users can consider their target species and choose variables that are mostly likely predictors of its niche.

To demonstrate Nicher, SDM was performed using 173 occurrence locations representing three key groundwater dependent species (*Eucalyptus victrix*, *E. camaldulensis* and *Melaleuca argentea*) captured from the Pilbara region of Western Australia. As per most correlative SDM techniques, pseudo-absence occurrence points are required for model fitting. A maximum of 5000 pseudo-absence points were generated randomly within the study area boundary. As these values were computed randomly, it is possible that some were generated at the same point as putative GDV.

Given the size of the sample it is unlikely that this would affect accuracy greatly, although this is dependent on the density of GDV in the study area. The topographic derivative values were extracted at each location, and the presence and pseudo-absence occurrence locations were combined and then split into training and testing sets, with 10% of the processed data designated as the testing set. Nicher was then created using 500 estimators and was set to undertake five cross-validations using the training dataset. Finally, the model was fit, output classes were converted into class probabilities, and probabilities were mapped onto a raster grid. This process was repeated with the raw canopy height model added as a model variable.

2.4 Fractional Cover Mapping

Fractional cover mapping (FCM) is a technique that takes as input a lower resolution image and first determines the percentage of pixels that represent the landcovers of a higher resolution image. As shown in Figure 3, using a high-resolution image that is four times smaller than a lower resolution image, the percentage of pixels that are GDV or not GDV are translated into new layers.

To determine GDV pixels, we used prior mapping that had been conducted by external parties as input and set the rest of the classes to “Not GDV”. The GDV and Not GDV layers transformed into percentages (Figure 3) are then used as evidence in a random forest regression against the image bands of the lower resolution image. The output is then a prediction of the percentage of pixels making up the input land covers, here shown as just ‘GDV’ and ‘Not GDV’ but any number of classes (or land covers) are permissible.

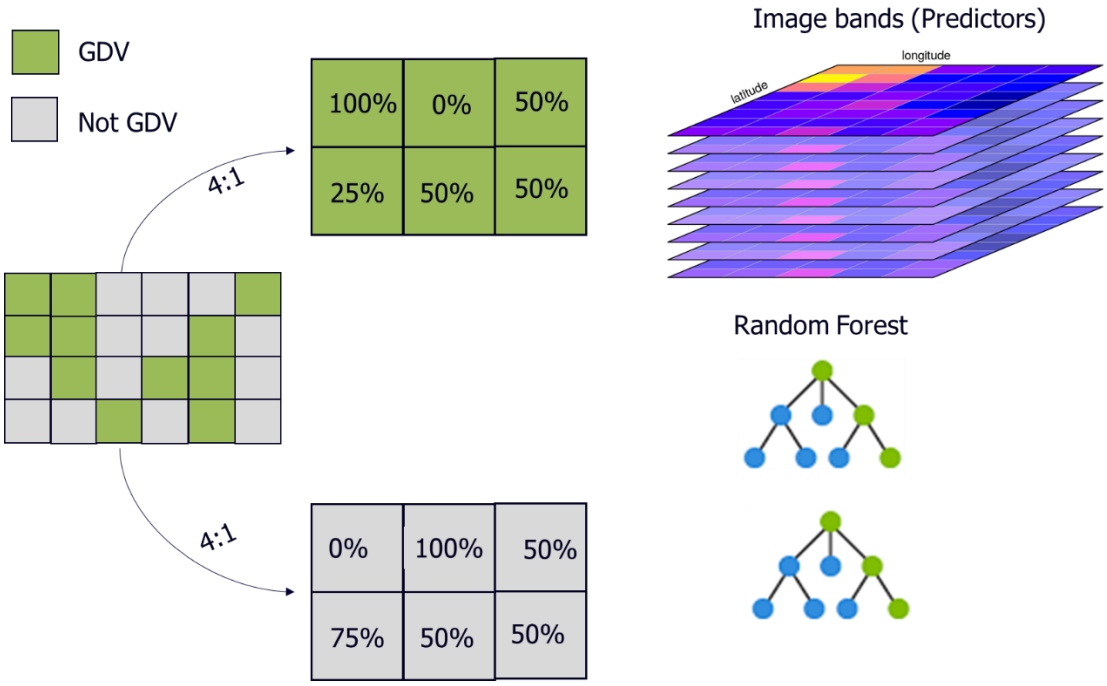


Figure 3 – Depiction of the algorithm used for fractional cover mapping. The percentage of pixels of a lower resolution image are determined on a higher resolution image before being used in a random forest regression against image bands. The output provides predicted percentages of the input classes (here GDV or not GDV).

2.5 Ensemble Model

Dempster-Shafer Belief Modelling (DSBM) was used to combine the three outputs (phenometrics, SDM and FCM). DSBM requires a frame of discernment (aka as a universe of discourse) where all possible hypotheses and combinations of hypotheses are disclosed. We had two singleton hypotheses [IS GDV] and [IS NOT GDV] with a third non-singleton hypothesis [IS GDV, IS NOT GDV] used to represent uncertainty (Figure 4). For example, if a pixel had a value of 0.6 (on a scale of 0 – 1) representing the basic probability of belonging to the set [IS GDV] then the complement (1-0.6 = 0.4) is given to the uncertainty hypothesis [IS GDV, IS NOT GDV] and no value is given to the opposing hypothesis [IS NOT GDV]. Likewise, if a pixel has a value of 0.7 in favour of the hypothesis [IS NOT GDV] then 0.3 is given to the uncertainty hypothesis and not to the opposing hypothesis [IS GDV].

Deciding which hypothesis each layer of evidence supports is not always clear cut. The evidence used to support the hypothesis [IS GDV] were the phenometrics and the SDM layers (Figure 4). Whilst these layers could certainly be inverted and used for the [NOT GDV] hypothesis we felt that the phenometrics showing highly productive vegetation represented our best evidence in favour of GDV being present. Likewise, we felt that highly suitable habitat was excellent, yet indirect, evidence of GDV. The evidence used to support the hypothesis [NOT GDV] was an inverted layer of the fractional cover mapping where a value of 1 represented land covers like stony plains and a value of 0 represented a high percentage likelihood of GDV (Figure 4). The idea being that by producing evidence images that support the negative hypothesis, we can further refine where GDV is likely to occur by applying evidence that reduces the likelihood that it will not be there.

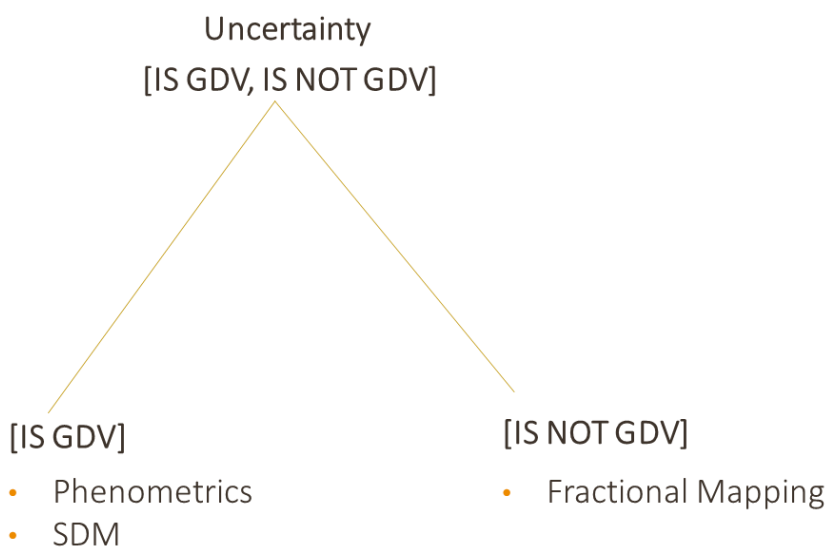


Figure 4 – The frame of discernment showing the two singleton hypotheses and one non-singleton hypothesis representing uncertainty. The phenometrics and SDM layers were used in support of the [IS GDV] hypothesis and an inverted fractional cover map was used to support the hypothesis [IS NOT GDV].

We used Dempster-Shaffer's rule of combination to combine our layers in support of their respective hypotheses. The mathematically inquisitive should see Equation 4 of Jøsang & Pope (2012). Essentially the logic is as follows, where support for the non-

singleton hypothesis is determined by using the complement of the images (e.g. 1-SDM_site represents the degree of support for [ISGDV, IS NOT GDV]:

- #combine all site information
- bpa_site = ("phenometrics_site" * "SDM_site") + ((1 - "SDM_site") * ("phenometrics_site")) + ((1 - "phenometrics_site") * ("SDM_site"))
- #combine output of above with the non-site imagery (FCM)
- IS GDV = ("bpa_site" * (1 - "FCM")) / (1 - ("FCM" * "%bpa_site%"))
- IS NOT GDV = ("FCM%" * (1 - "%bpa_site%")) / (1 - ("%FCM%" * "%bpa_site%"))
- PLAUSIBILITY = 1 – “IS NOT GDV”
- BELINT = “PLAUSIBILITY” – “IS GDV”

PLAUSIBILITY denotes the degree to which we cannot reject our hypothesis [IS GDV], but where we require more evidence. To put it in context, there’s some evidence to suggest GDV could occupy that pixel, but not enough to be certain, i.e., there is uncertainty. Such uncertainty can only be overcome by additional data or sampling and provides an excellent opportunity for feedback loops to improve the modelling.

The difference between the IS GDV and PLAUSIBILITY layers is the BELIEF INTERVAL. The higher the BELINT the more uncertain we are of confirming the hypothesis and is thus a useful map for directing field crews to locations of plausible, but uncertain, GDV locations.

It is not necessary to use these exact layers and users can also add additional layers to the universe of discourse if they wish. However, as a guide, please note:

- It is necessary to have at least one layer supporting each of the singleton hypotheses, otherwise, it is not possible to compute the plausibility, or belief interval; and
- If only one layer is available for a hypothesis (e.g. phenometrics supporting “IS GDV”) then that will be the full support for that hypothesis and no combination of layers will be necessary.

2.5 Near Real-time Monitoring

Near real-time monitoring was achieved using the Exponentially Weighted Moving Average Change Detection (EWMACD) method of Brooks et al., (2014). This method works by computing a harmonic model for a training period of imagery representing the baseline (or pre-impact period) and comparing subsequent imagery to identify the residuals. Here, we test the methodology using a study area north of Ophthalmia Dam northeast of the town of in Newman (Figure 5). Constructed in 1981, it is suspected that a lack of flood events north of the dam wall contributed to vegetation degradation, along with overstocking and a prolonged period of drought (Fox et al., 2000). This area was targeted for this exemplar because of the dynamism in vegetation health since its construction.

Figure 5 – Study area north of the Ophthalmia Dam.

3 Results

3.1 Phenometrics

3.1.1 Characterisation of Pilbara phreatophytes

GDV species had considerably higher MAVI values over the course of the year at all sites, relative to background (pseudo-absence) points.. Average MAVI values peaked around Julian Day 70 (early March) and again around Julian Day 203 (late July) before declining monotonically (Figure 2.3). Despite a lower response in 2019, the cycles appear to be consistent across the Pilbara. Relative to the Pilbara average, Groups 1-4 and 7 had consistently higher MAVI averages, whereas Groups 5, 6, and 8 were below the regional average (Figure 6).

3.1.2 Phenometrics and Variable Selection

Table 2.2 presents the AUC values for each of the eight areas based on the 18 different phenometrics trialled. We identified eight phenometrics that averaged over 0.9 from Table 2.2 (BSE, EOS, LIOS, LIOT, MOS, POS, SOS, VOS). These were consistently high across all groups and study periods. Notably, Groups 4 and 7 had almost perfect discrimination (0.98-0.99). The AUC values from all groups and study periods were used to construct a dendrogram and partition cluster plot for better visualisation of their discriminatory power and grouping into discriminatory “families”. Figure 8 identifies four families of phenometrics based on the GDV data. The eight phenometrics identified made up a single family (cluster 4) suggesting there is likely to be minimal differences between the discriminatory power and hence any of them could be used. We have chosen to use the LIOT.

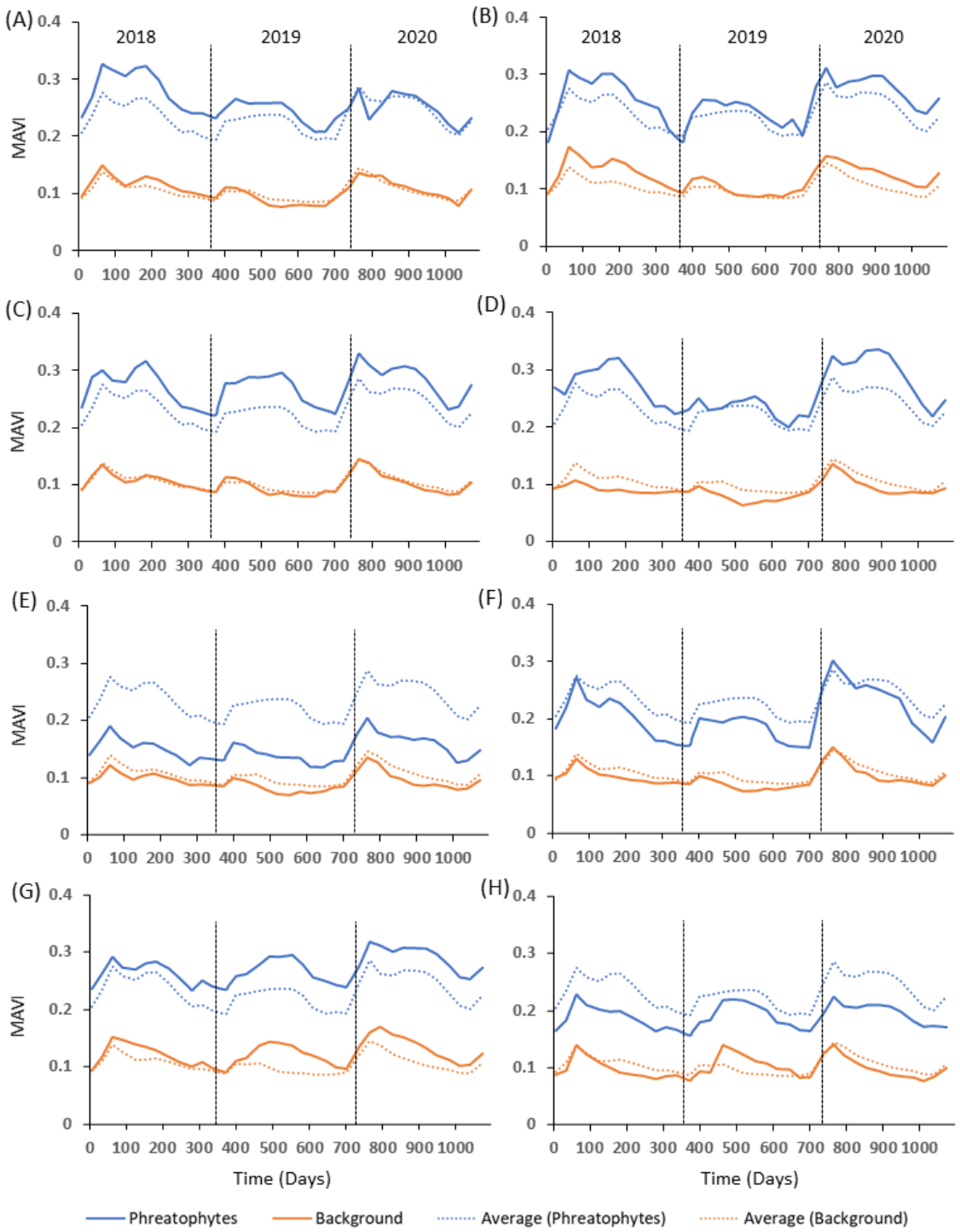


Figure 6 – MAVI response of phreatophytes versus background vegetation over three years at 7-day intervals at eight study sites in the Pilbara: A) Group 1, B) Group 2, C) Group 3, D) Group 4, E) Group 5, F) Group 6, G) Group 7 and H) Group 8 (see Figure 2.1).

Table 2. Areas under the curve for each of the eight groups generated using jack-knife validation based on a random 20% partition for A. 2018, B. 2019, and C. 2020.

A	Metric	Grp 1	Grp 2	Grp 3	Grp 4	Grp 5	Grp 6	Grp 7	Grp 8	Mean	S.Dev
	AOS	0.81	0.84	0.87	0.94	0.73	0.90	0.86	0.71	0.83	0.09
	BSE	0.93	0.93	0.93	0.98	0.86	0.95	0.99	0.90	0.93	0.05
	EOS_T	0.66	0.67	0.64	0.79	0.75	0.76	0.48	0.83	0.70	0.12
	EOS	0.93	0.93	0.94	0.98	0.84	0.95	0.99	0.89	0.93	0.05
	LIOS	0.91	0.91	0.90	0.97	0.84	0.92	0.99	0.87	0.91	0.05
	LIOT	0.93	0.93	0.94	0.98	0.85	0.95	0.99	0.88	0.93	0.05
	LOS	0.70	0.67	0.61	0.77	0.66	0.76	0.57	0.78	0.69	0.08
	MOS	0.93	0.93	0.94	0.97	0.84	0.95	0.99	0.84	0.92	0.06
	POS_T	0.76	0.77	0.81	0.92	0.64	0.54	0.95	0.65	0.75	0.15
	POS	0.93	0.93	0.94	0.98	0.84	0.95	0.99	0.85	0.93	0.06
	ROD	0.82	0.83	0.87	0.92	0.54	0.79	0.90	0.75	0.80	0.13
	ROI	0.64	0.67	0.68	0.76	0.78	0.92	0.50	0.64	0.70	0.13
	SIOS	0.81	0.83	0.85	0.94	0.77	0.93	0.93	0.68	0.84	0.10
	SIOT	0.83	0.83	0.88	0.95	0.73	0.56	0.90	0.68	0.79	0.14
	SOS_T	0.68	0.68	0.66	0.69	0.69	0.96	0.63	0.68	0.71	0.11
	SOS	0.93	0.92	0.94	0.98	0.87	0.73	0.99	0.89	0.90	0.09
	VOS_T	0.66	0.71	0.74	0.83	0.84	0.94	0.59	0.61	0.74	0.13
	VOS	0.93	0.93	0.94	0.98	0.83	0.90	0.99	0.90	0.92	0.05
B	Metric	Grp 1	Grp 2	Grp 3	Grp 4	Grp 5	Grp 6	Grp 7	Grp 8	Mean	S.Dev
	AOS	0.82	0.84	0.88	0.90	0.83	0.88	0.80	0.62	0.82	0.09
	BSE	0.94	0.94	0.94	0.98	0.90	0.95	1.00	0.90	0.94	0.04
	EOS_T	0.81	0.83	0.86	0.91	0.77	0.90	0.75	0.81	0.83	0.06
	EOS	0.94	0.94	0.94	0.98	0.90	0.96	0.99	0.87	0.94	0.04
	LIOS	0.93	0.93	0.93	0.93	0.90	0.93	0.93	0.87	0.92	0.02
	LIOT	0.95	0.94	0.94	0.98	0.92	0.96	0.99	0.85	0.94	0.05
	LOS	0.79	0.82	0.85	0.85	0.79	0.88	0.76	0.74	0.81	0.05
	MOS	0.94	0.94	0.94	0.98	0.92	0.95	0.98	0.83	0.93	0.05
	POS_T	0.79	0.81	0.87	0.90	0.65	0.82	0.69	0.65	0.77	0.10
	POS	0.94	0.94	0.94	0.98	0.92	0.95	0.98	0.82	0.93	0.05
	ROD	0.66	0.74	0.82	0.83	0.62	0.77	0.91	0.62	0.75	0.11
	ROI	0.57	0.61	0.59	0.71	0.83	0.72	0.75	0.48	0.66	0.12
	SIOS	0.88	0.87	0.90	0.91	0.89	0.90	0.73	0.64	0.84	0.11
	SIOT	0.89	0.88	0.91	0.90	0.90	0.78	0.77	0.64	0.83	0.10
	SOS_T	0.62	0.71	0.78	0.66	0.69	0.95	0.68	0.64	0.72	0.11
	SOS	0.93	0.93	0.94	0.98	0.90	0.89	0.99	0.87	0.93	0.05
	VOS_T	0.79	0.79	0.81	0.92	0.82	0.95	0.61	0.58	0.78	0.14
	VOS	0.93	0.93	0.94	0.98	0.90	0.88	1.00	0.88	0.93	0.05
C	Metric	Grp 1	Grp 2	Grp 3	Grp 4	Grp 5	Grp 6	Grp 7	Grp 8	Mean	S.Dev
	AOS	0.77	0.78	0.82	0.86	0.71	0.81	0.55	0.75	0.75	0.10
	BSE	0.91	0.93	0.95	0.98	0.78	0.94	0.95	0.84	0.92	0.07
	EOS_T	0.71	0.80	0.83	0.83	0.95	0.82	0.91	0.91	0.85	0.08
	EOS	0.93	0.93	0.95	0.98	0.79	0.95	0.97	0.89	0.92	0.07
	LIOS	0.83	0.89	0.94	0.95	0.84	0.94	0.97	0.88	0.91	0.06
	LIOT	0.91	0.93	0.95	0.98	0.82	0.95	0.96	0.87	0.92	0.06
	LOS	0.71	0.77	0.80	0.76	0.83	0.80	0.89	0.84	0.80	0.06
	MOS	0.91	0.92	0.95	0.97	0.74	0.90	0.93	0.86	0.89	0.08
	POS_T	0.85	0.78	0.73	0.81	0.67	0.62	0.81	0.76	0.75	0.08
	POS	0.91	0.92	0.95	0.97	0.74	0.90	0.94	0.86	0.90	0.08
	ROD	0.81	0.75	0.76	0.82	0.49	0.67	0.54	0.57	0.67	0.14
	ROI	0.68	0.66	0.68	0.66	0.44	0.65	0.56	0.62	0.61	0.09
	SIOS	0.69	0.78	0.87	0.92	0.75	0.91	0.58	0.78	0.79	0.12
	SIOT	0.77	0.82	0.89	0.95	0.77	0.90	0.58	0.75	0.80	0.12
	SOS_T	0.65	0.60	0.54	0.70	0.63	0.54	0.61	0.66	0.62	0.06
	SOS	0.90	0.91	0.95	0.98	0.76	0.91	0.92	0.83	0.89	0.07
	VOS_T	0.77	0.71	0.63	0.82	0.71	0.78	0.60	0.78	0.73	0.08
	VOS	0.92	0.93	0.95	0.98	0.81	0.95	0.97	0.89	0.92	0.06

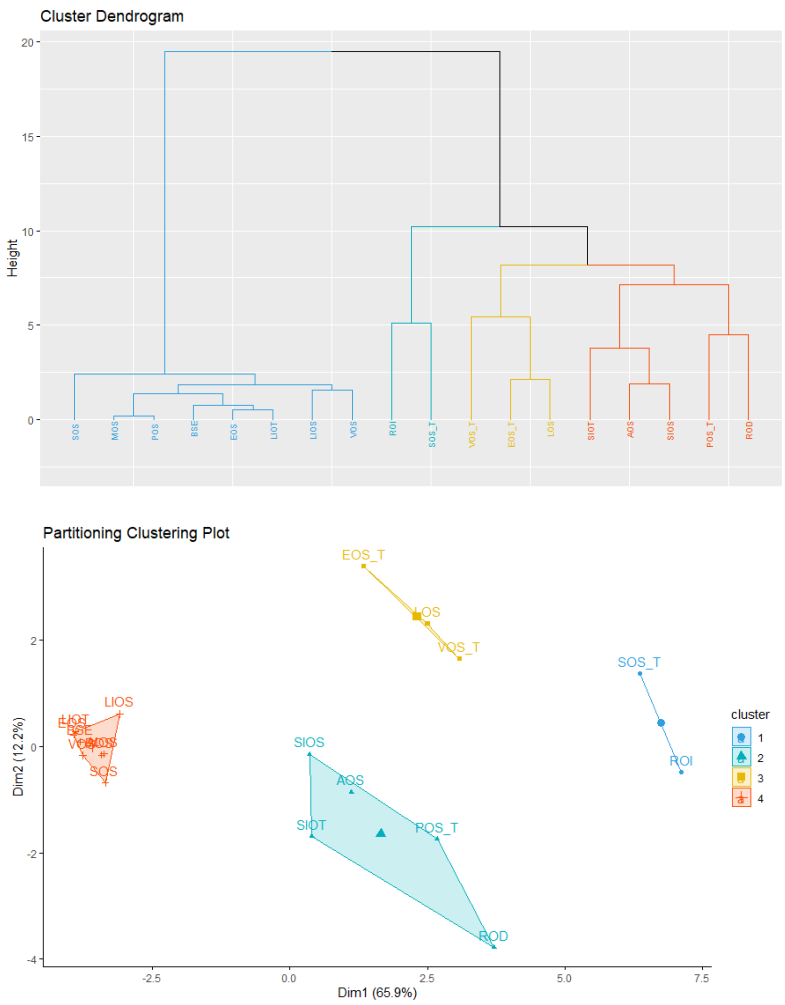


Figure 7. The clustering dendrogram and partitioning cluster plot identified four “families” of phenometrics based on AUC values.

3.2 Species Distribution Modelling

SDM results from the Nicher module are provided in Figure 8. Figure 8c presents a satellite image of an area with known GDV tree locations along a major drainage channel and vehicle track in the Pilbara. This vegetation, which was mostly restricted to the banks of the drainage line, were captured by Nicher using the nine derivatives (Figure 8b). As some mining infrastructure (such as cleared vehicle tracks) were returned by Nicher, the CHM was used to mask out any vegetation < 1 m in height (Figure 8c). This process successfully removed the vehicle track (Figure 8c). Finally, addition of the CHM as a model variable mapped the same general GDV extent along the drainage line, with the exception that more vegetation appears to have been captured on the drainage line margins as well as in additional drainage lines and stony

flat areas to the north of the GDV occurrence points (Figure 9a, Figure 9d) in the northern drainage lines and stony flats.

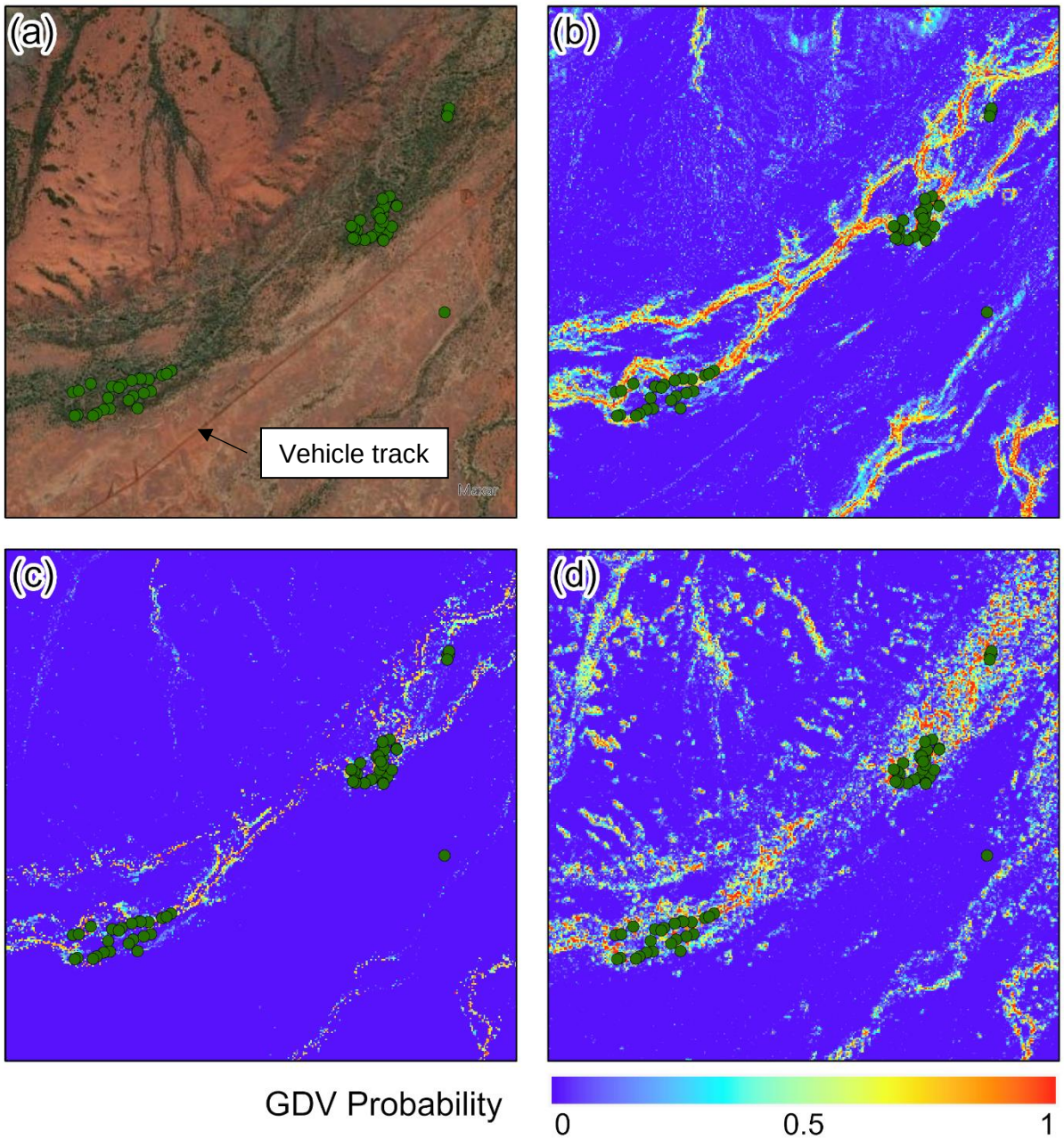


Figure 8 Several versions of the same Nicher SDM. Green points are known GDV species field occurrences. Coloured pixels represent GDV probability: blue is highly improbable and red is highly probable. Displayed: a) satellite image basemap; b) SDM based on nine derivative variables (and DEM); c) same SDM with vegetation < 1 m height masked out; and d) SDM based on nine derivative variables (and DEM) in addition to the CHM included as a model variable.

Receiver operating characteristic (ROC) curves were constructed and the area under the curve (AUC; Fawcett, 2006) was calculated for both models (with and without the CHM variable), resulting in an AUC of 0.91 and 0.93, respectively. The balanced accuracy score was also generated for each model, providing an indication of the fraction of correct predictions that were made by the model when compared to the testing set. For the models with and without CHM, the balanced accuracy of classification returned was 0.79 and 0.87, respectively. This suggests that while both models performed strongly, the inclusion of the CHM variable improved model accuracy.

3.3 Fractional Cover Mapping

Figure 9 shows previous mapping of several land covers in the top left-hand portion of the image. The orange class contained GDV with mixed understory and was used to train the image on the GDV class. All other classes were converted into one absence class. The output shows green pixels as having a high fraction of GDV in them whereas white to brown pixels have a low fraction of GDV. This image was then inverted so that the whitest areas represented the “disbelief” hypothesis (“Is Not GDV”; see next) in the ensemble modelling.

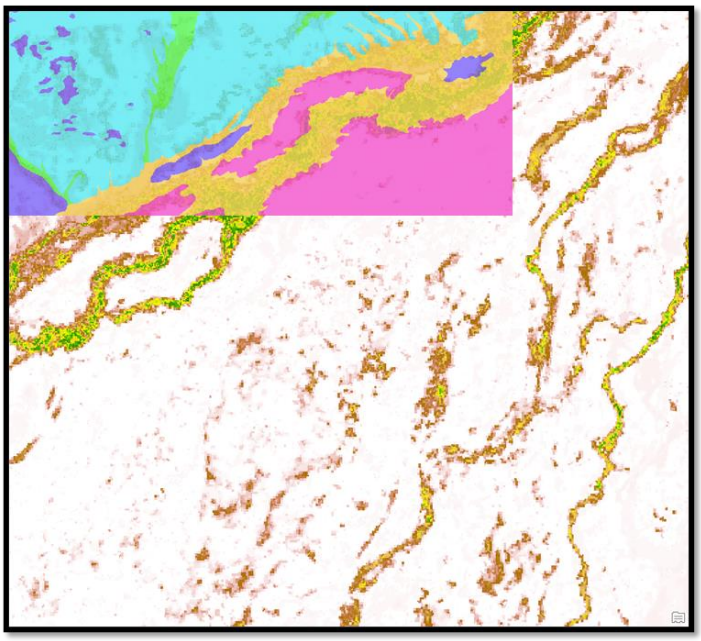


Figure 9 Fractional cover map of GDV species extrapolated from partial mapping of the area.

3.4 Ensemble Modelling

Figure 10 show the outputs of the Dempster-Shafer Belief ensemble model. Belief in the hypothesis [IS GDV] shows up most riparian channels in the area as likely containing GDV (Figure 10A). The greatest disbelief is on the stony plains away from the riparian channels (Figure 10B). Plausibility softens the belief map by providing extra pixels in proximity to high belief GDV pixels (Figure 10C) and the Uncertainty map (Figure 10D) shows the difference between the Plausibility and the Belief layers. Areas with the highest uncertainty would benefit most from additional sampling and/or new evidential layers in the ensemble.

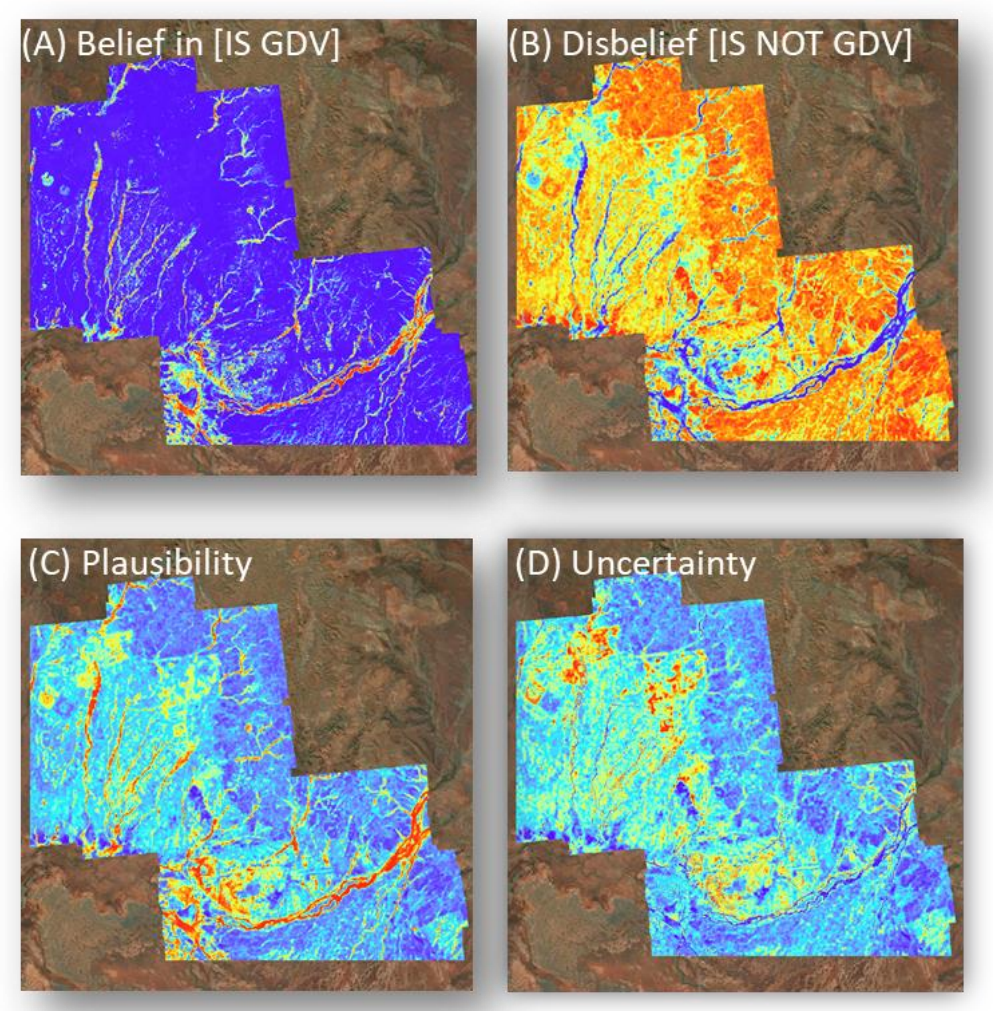


Figure 10 Outputs of the Dempster-Shafer Belief ensemble model. (A) Belief that the pixel contains GDV. (B) Belief that the pixel does not contain GDV (disbelief). (C) Plausibility that a pixel may contain GDV, but more evidence is needed. (D) Locations shown in red are those where more evidence would be most beneficial in reducing uncertainty.

3.5 Near Real Time Monitoring

An undisclosed area north of the dam was selected and given the unique ID – A1 (Figure 11). It was chosen because we identified historical degradation of *Eucalyptus* spp during ground-truthing in July 2019.

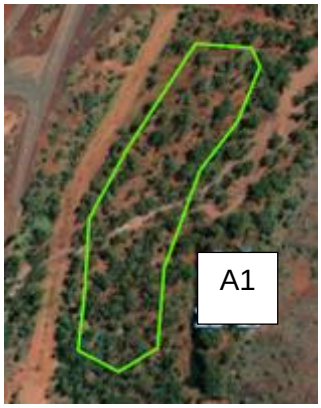


Figure 11 – Study area used to test the NRT algorithm

The baseline period is shown in purple (Figure 11) and corresponds to 1/1/1999 to 1/1/2001 and chosen because of its apparent stability and was several years before obvious degradation occurred. A harmonic model was created for the baseline period, which fit the vegetation index closely (R^2 = 0.9) for all periods up to 2006. Subsequent to that, the vegetation index deviated significantly from the harmonic, breaking the lower bounding confidence limit in 2006 (Figure 11).

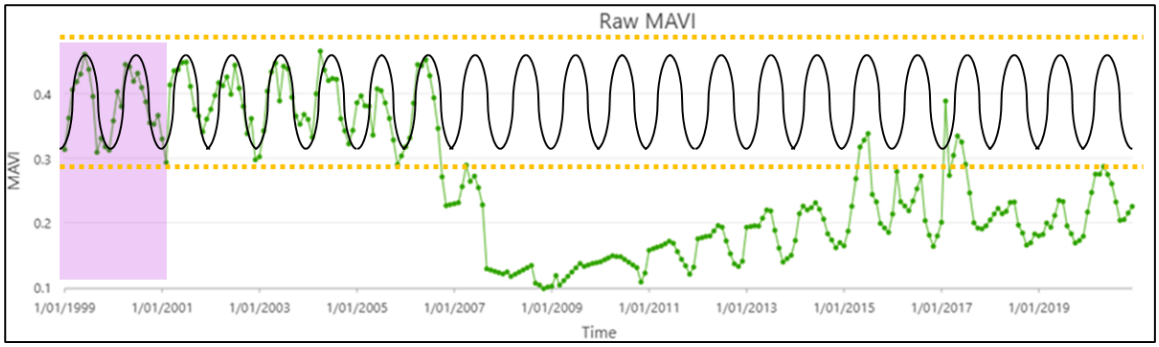


Figure 11 – Training of the time series of MAVI imagery over one pixel with a harmonic model. The harmonic breaks beyond the lower control limit in late 2006.

Significant changes can also be observed using the change deviation chart (Figure 12). This picks up the apparent decline around 2006 and onwards and provides a

magnitude of how far the pixel has deviated from the harmonic model. Change deviations are grouped by zone, which can be interpreted as shown in Figure 13. For example, Zone -11 represents the most significant decline, which is 19 deviations below the baseline vegetation. A change deviation of 0, represents similar vegetation vigour to the baseline in terms of seasonal cycles and MAVI. Figure 14A shows map representation of the change deviation magnitude for each pixel and summarised into the zone classes as a polygon average (Figure 14B).

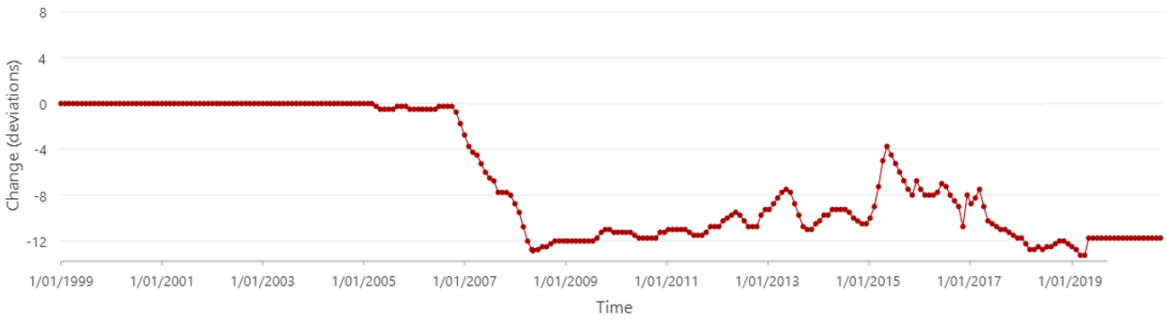


Figure 12 – The change deviations resulting from the EWMCD method. Vegetation decline begins to register in 2006 and continues to be compared to the training period into the future. The larger the change deviation the more different it is from the pre-impact period (1999-2001).

Zone Legend

Symbology	Zone	Description
	-11	Change deviation is below -19. Decline.
	-10	Change deviation is between -17 and -19. Decline.
	-9	Change deviation is between -15 and -17. Decline.
	-8	Change deviation is between -13 and -15. Decline.
	-7	Change deviation is between -11 and -13. Decline.
	-6	Change deviation is between -9 and -11. Decline.
	-5	Change deviation is between -7 and -9. Decline.
	-4	Change deviation is between -5 and -7. Decline.
	-3	Change deviation is between -3 and -5. Decline.
	-2	Change deviation is between -1 and -3. Decline.
	-1	Change deviation is between 0 and -1. Decline.
	0	No change in either direction.
	1	Change deviation is between 0 and 1. Growth.
	2	Change deviation is between 1 and 3. Growth.
	3	Change deviation is between 3 and 5. Growth.
	4	Change deviation is between 5 and 7. Growth.
	5	Change deviation is between 7 and 9. Growth.
	6	Change deviation is between 9 and 11. Growth.
	7	Change deviation is between 11 and 13. Growth.
	8	Change deviation is between 13 and 15. Growth.
	9	Change deviation is between 15 and 17. Growth.
	10	Change deviation is between 17 and 19. Growth.

Figure 13 – Interpreting change deviation zones and colours.

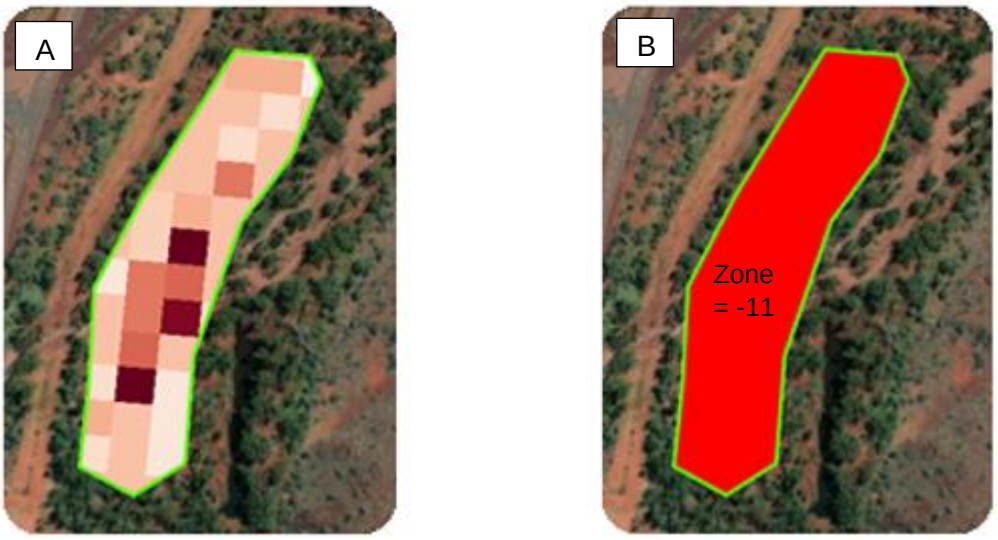


Figure 14 – (A) Representation of change deviation at the pixel level vs (B) alert system representation using the change deviation zone classification method.

4 Conclusions

Phenometrics provide the ability to harness the power of temporal resolution available in sun-synchronous remote sensing imagery. The power of Sentinel imagery, with 5-day revisit (7-day rolling median) and 10 m spatial resolution, can open new doors not previously available using 250 m MODIS imagery. We identified at least eight phenometrics that could discriminate between GDV and non-GDV, including LIOT, which we hypothesised would provide separation based on year-round productivity.

Interpolated Lidar data provides fine scale topographic data that can be used to derive a series of inputs for habitat modelling. Random forests coupled with these derivatives and known GDV locations can be used to create habitat models with demonstrated high accuracy in the Pilbara bioregion, where most GDV is restricted to riparian zones.

GDV can be mixed with other land covers, even at 10 m resolution. Fractional mapping is a way to extract the fraction of GDV within a pixel and express it as a percentage. It is also a suitable method for scaling up from fine scale mapping and extrapolating over entire tenements.

Ensemble modelling using Dempster-Shafer Belief Theory allows the integration of layers of information that are indirect evidence of GDV to produce maps relating to the belief of GDV being present, the disbelief of presence as well as providing a level of uncertainty in the prediction. The latter is useful for model refinement through new knowledge and/or data layers.

Near-real-time monitoring allows the user to track important changes in vegetation over time to detect anomalies that may be indicative of GDV loss (or regrowth). Its ability to provide alerts can be used for early detection of decline, potentially from dewatering.

REFERENCES

Brooks, E.B., Wynne, R.H., Thomas, V.A., Blinn, C.E. and Coulston, J.W., 2014. On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6), pp.3316-3332

ESRI, 2021. ArcGIS Pro: Release 2.7. Redlands, CA: Environmental Systems Research Institute.

Elith, J., Phillips, S., Hastie, T., Dudík, M., Chee, Y. and Yates, C., 2010. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, Vol. 17(1), pp.43-57.

Eklundh, L., and Jönsson, P., 2017. TIMESAT 3.3 with seasonal trend decomposition and parallel processing - Software Manual.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd (Vol. 96, No. 34, pp. 226-231).

Evans, I. S., 1972. General geomorphometry, derivatives of altitude, and descriptive statistics. Spatial Analysis in Geomorphology New York: Harper & Row pp.17 – 90.

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

Fox, J., Burrows, C. and Hopkins, M., 2022. Proceedings of the 3rd Queensland Environmental Conference: Sustainable Environmental Solutions for Industry and Government; a Focus of Sound, Practical and Economically Viable Solutions for Industry and Government - Monitoring Revegetation of a Severely Degraded Rangeland, Western Australia (Engineering Collection) - Informit. [online] Qatest.informit.com.au. Available at: <http://qatest.informit.com.au/documentSummary;dn=721076761468586;res=IELENG> [Accessed 13 July 2022].

Friedman, J., Hastie, T. and Tibshirani, R., 2001. The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics, pp. 587-588.

Fu, P. and Rich, P.M., 2002. A geometric solar radiation model with applications in agriculture and forestry. Computers and electronics in agriculture, 37(1-3), pp.25-35.

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M. and Prueger, J.H. (2017) Toward Mapping Crop Progress at Field Scales through Fusion of Landsat and MODIS Imagery. *Remote Sensing of Environment*, 188: 9–25., doi:10.1016/j.rse.2016.11.004.

Guisan, A., Weiss, S.B. and Weiss, A.D., 1999. GLM versus CCA spatial modelling of plant species distribution. Plant ecology, 143(1), pp.107-122.

Fielding, A. H. and Bell, J. F. (1997) A Review of Methods for the Assessment of PredictionErrors in Conservation Presence/Absence Models, *Environmental Conservation*, Vol. 24, pp.38-49.

Jøsang, A., & Pope, S. (2012). DEMPSTER'S RULE AS SEEN BY LITTLE COLORED BALLS. Computational Intelligence, 28(4), 453-474. doi: 10.1111/j.1467-8640.2012.00421.x

Kramer, K., Leinonen, I. & Loustau, D. The importance of phenology for the evaluation of impact of climate change on growth of boreal, temperate and Mediterranean forests ecosystems: an overview. International Journal of Biometeorology 44, 67–75 (2000). https://doi.org/10.1007/s004840000066

Lisein, J., Pierrot-Deseilligny, M., Bonnet, S. and Lejeune, P., 2013. A photogrammetric workflow for the creation of a forest canopy height model from small unmanned aerial system imagery. Forests, 4(4), pp.922-944.

Lorena, A.C., Jacintho, L.F., Siqueira, M.F., De Giovanni, R., Lohmann, L.G., De Carvalho, A.C. and Yamamoto, M., 2011. Comparing machine learning classifiers in potential distribution modelling. Expert Systems with Applications, 38(5), pp.5268-5275.

Melaas, E.K., Friedl, M.A., and Zhu, Z. (2013) Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data, Remote Sensing of Environment, 132: 176-185.

Onojeghuo, A., Blackburn, G.A., Wang, Q., Atkinson, P.M., Kindred, D., and Miao, Y. (2018) Rice Crop Phenology Mapping at High Spatial and Temporal Resolution

Using Downscaled MODIS Time-Series. *GIScience & Remote Sensing*, 55: 659–677., doi:10.1080/15481603.2018.1423725.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, pp.2825-2830.

Pontius, R. G. Jr., and Schneider, L. C. (2001) Land-Cover Change Model Validation by anROC method for the Ipswich Watershed, Massachusetts, USA, Agriculture, *Ecosystems and Environment*, Vol. 85, pp. 239-248.

Reed, B.C., Schwartz, M.D., Xiao, X., 2009. Remote Sensing Phenology: Status and the Way Forward, in: Noormets, A. (Ed.), Phenology of Ecosystem Processes. Springer New York, pp. 231-246.

Riley, S.J., De Gloria, S.D., Elliot, R., 1999. A Terrain Ruggedness that Quantifies Topographic Heterogeneity. Intermountain Journal of Science, Vol.5, No.1-4, pp.23-27.

Robinson, T., Di Virgilio, G., Temple-Smith, D., Hesford, J., & Wardell-Johnson, G. (2019). Characterisation of range restriction amongst the rare flora of Banded Ironstone Formation ranges in semiarid south-western Australia. Australian Journal Of Botany, 67(3), 234. doi: 10.1071/bt18111

Samis, K.E. and Eckert, C.G., 2009. Ecological correlates of fitness across the northern geographic range limit of a Pacific Coast dune plant. Ecology, 90(11), pp.3051-3061.

Schreier, J., Ghazaryan, G. & Dubovyk, O. (2021) Crop-specific phenomapping by fusing Landsat and Sentinel data with MODIS time series, European Journal of Remote Sensing, 54:sup1, 47-58, DOI: 10.1080/22797254.2020.1831969

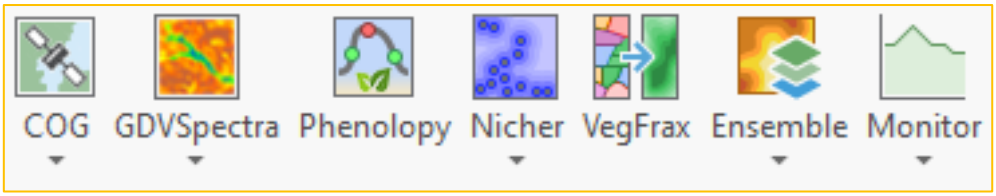
Appendix 1 - Tutorial

1 Introduction

Removal of unwanted groundwater from mine pit dewatering and other associated operations can affect the health of GDV, which relies on a stable water table for its water requirements. Monitoring GDV is thus a requirement around mining leases but is poorly targeted. Narrowing the search space using remote sensing imagery is the first step towards a time saving, cost efficient, and comprehensive monitoring program that can provide peace of mind that dewatering is being conducted appropriately.

Tenement Tools is a free and open-source ArcGIS Pro 2.8 plug-in that offers a suite of tools to assist users in detecting and monitoring potential GDV areas within the Pilbara region of Western Australia at a tenement-scale.

A brief overview of the tools currently on offer include:



COG	Query, download and explore a 'cube' of analysis ready satellite data (Landsat 5, 7, 8 or Sentinel 2) from the Digital Earth Australia (DEA) for any given study area and time period in Australia.
GDVSpectra	Utilise a field-validated, seasonally-weighted model to detect potential GDV for a given study area and time period. Also offers tools to detect vegetation greenness trends and change over time.
Phenolopy	Generate metrics from a year or more of satellite imagery depicting various characteristics of vegetation life cycle (phenology). Notable metrics include peak of season (POS), start of season (SOS) and productivity (LIOT, SIOT).
Nicher	Generates a species distribution model (SDM) for a given set of elevation-based topographic or climatic variables and known field-observed occurrences points of a target species. Used to detect suitable habitat (i.e., niche) for a given species in a landscape.
VegFrax	Performs Fractional Cover Mapping (FCA), given a high-resolution pre-classified vegetation map and a 'cube' of lower-resolution satellite imagery. Useful for extrapolating vegetation type mapping over larger spatial extents.
Ensemble	Combine two or more maps of evidence (from above tools or other) into one ensemble map via Dempster-Shafer theory. Useful to refine potential GDV areas by combining relevant evidence. Also provides outputs that help highlight areas that may be under-sampled or low in model confidence.

Monitor	Perform on-going, near real-time vegetation monitoring via modern change detection algorithms for one or more monitoring areas for any area in Australia. Provides tools to create monitoring areas, set pre-impact periods and model parameters, and result graphing.

This document intends to offer a generic walkthrough of the release version (1.0.1) of the Tenement Tool suite.

1.1 Requirements

- ArcGIS Pro version 2.8 or 2.9
- At least 8gb of ram (>=16gb recommended)
- Fast and stable internet connection
- Access to the following websites (IT may need to open):
    - https://conda.anaconda.org
    - https://repo.anaconda.com
    - https://conda-forge.org/
    - https://data.dea.ga.gov.au
    - https://explorer.sandbox.dea.ga.gov.au/stac
    - https://explorer.sandbox.dea.ga.gov.au/stac/search

1.2 Setup and installation

Installation instructions are provided on the official Tenement Tools GitHub page located at https://github.com/frontiersi/tenement-tools. Please complete the installation process before proceeding with this tutorial.

1.3 Tutorial data

Data has been provided for this tutorial and is located on the GitHub page on the releases page at https://github.com/frontiersi/tenement-tools/releases. Download the Tutorial.zip file for your relevant version before proceeding with this tutorial. Please note, most of this data is simulated and is simply for demonstration purposes.

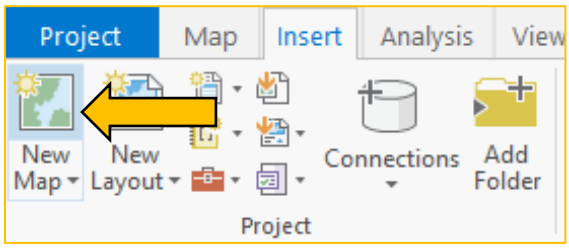
1.4 Known issues

The Tenement Tools plugin has been thoroughly bug checked for ArcGIS Pro versions 2.8 and 2.9 and can now be considered stable. Please see the known bugs section of the GitHub page at https://github.com/frontiersi/tenement-tools for any rare issues that remain.

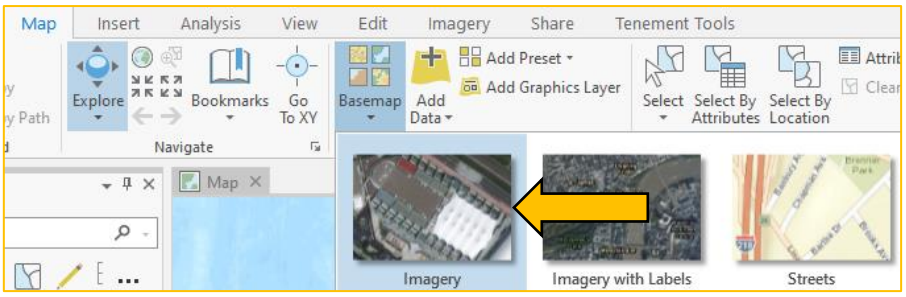
1.5 Tutorial preparation

Let's setup ArcGIS Pro for the tutorial. First, ensure you have followed the installation instructions on the GitHub page and have a fresh ArcGIS project running on your screen. Also ensure you have added the tenement tools toolbox to your current project.

If no map is currently visible on your screen, click the Insert tab on the top menu and create a new project map. Otherwise, use the default map that already exists:

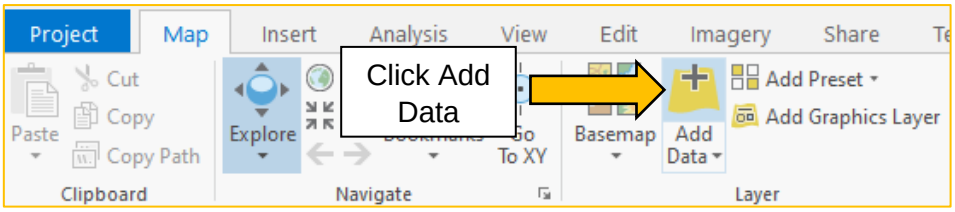


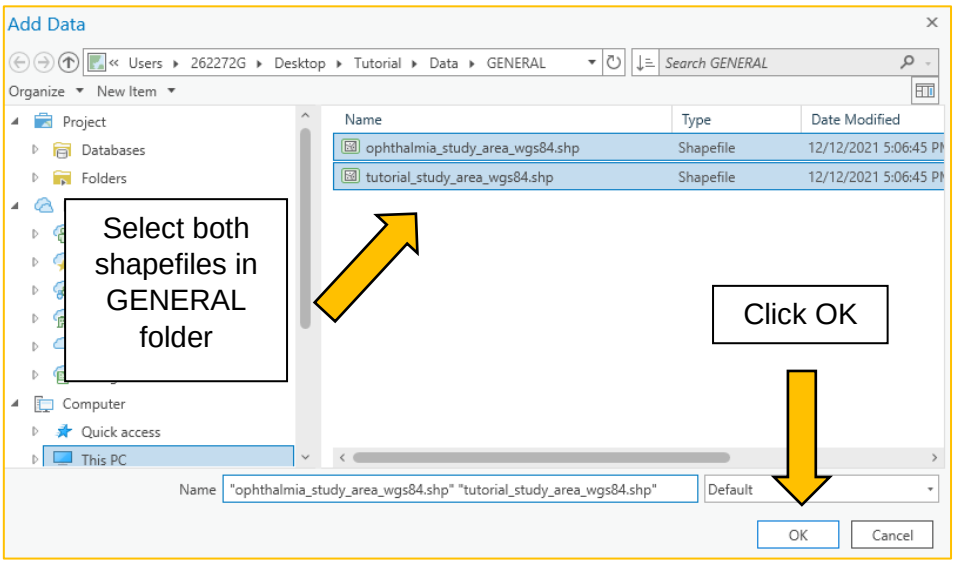
The default map will likely have a basic looking topographic basemap of Australia. If you prefer seeing an aerial image, click the Map tab and change the basemap layer to Imagery:



Finally, we need to add our study area boundaries to the currently active map. If you haven't already, unzip the tutorial data contained with the Tutorial.zip anywhere on your computer.

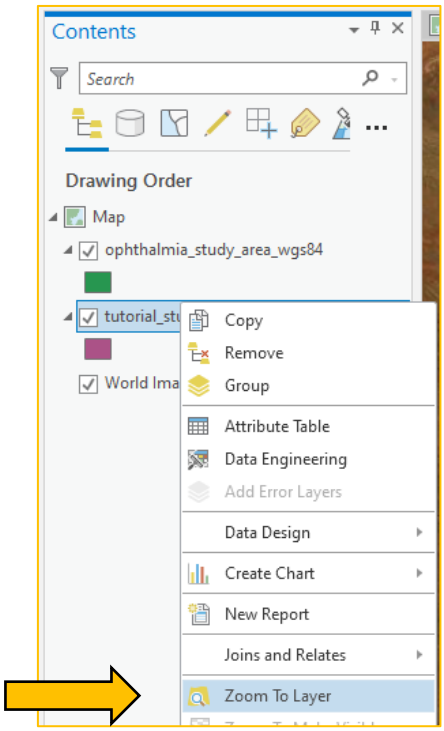
Then, click the Add Data button on the top menu, navigate to the GENERAL folder within, and add the two study area shapefiles *tutorial_study_area_wgs84.shp* and *ophthalmia_study_area_wgs84.shp* to the current map, like so:





You should see the two layers appear on the contents pane, as well as on the main map view area.

We will focus on the tutorial area first. Right click the layer called *tutorial_study_area_wgs84* in the contents pane and click Zoom to Layer:



The map view will zoom to the selected area and will look similar to this (likely with different colouring):

Spend some time changing the symbology of the study area feature (e.g., remove the fill colour and set the border to black). Likewise, have a look around the study area and the underlying aerial imagery to get a sense of the area.

This particular area is in the Hamersley Range in the Pilbara, Western Australia. It is located approximately 60km south-east of Pannawonica. Based on relatively recent aerial imagery it appears to be a relatively untouched by disturbance with the exception of significant bushfires.

The majority of the area appears to be plateaus, mesas and hills of Triodia grasslands with scattered shrubs and the occasional moderate to large drainage area containing *Eucalyptus* species.

2 COG Fetch tool

2.1 Introduction

The COG Fetch tool is used to query and download analysis ready satellite images for any area in Australia. This is achieved using Digital Earth Australia (DEA) Amazon Web Services (AWS) public data bucket (https://data.dea.ga.gov.au/).

Satellite data offered by Tenement Tools includes imagery from the Landsat 5, 7, 8 and Sentinel 2A, 2B satellite platforms. Landsat provides an archive of 30m resolution imagery starting 1986 to today (and on-going), whereas Sentinel provides imagery starting in 2016 and on-going today at a higher resolution of 10m.

The specific Landsat and Sentinel products available include:

- Landsat 5 TM Surface Reflectance NBART ARD (ga_ls5t_ard_3)
- Landsat 7 ETM+ Surface Reflectance NBART ARD (ga_ls7e_ard_3)
- Landsat 8 OLI Surface Reflectance NBART ARD (ga_ls8c_ard_3)
- Sentinel 2A Surface Reflectance NBART ARD (s2a_ard_granule)
- Sentinel 2B Surface Reflectance NBART ARD (s2b_ard_granule)

These products are all analysis ready and have been geometrically, atmospherically, and terrain corrected using a Nadir-corrected BRDF Adjusted Reflectance with Terrain illumination correction (NBART) method. They also come with a 'pixel quality' band that allows us to remove cloud, shadow and other invalid pixels.

A deeper overview of these DEA products is provided on these products here: https://docs.dea.ga.gov.au/notebooks/Beginners_guide/02_DEA.html

2.2 Use the COG Fetch tool

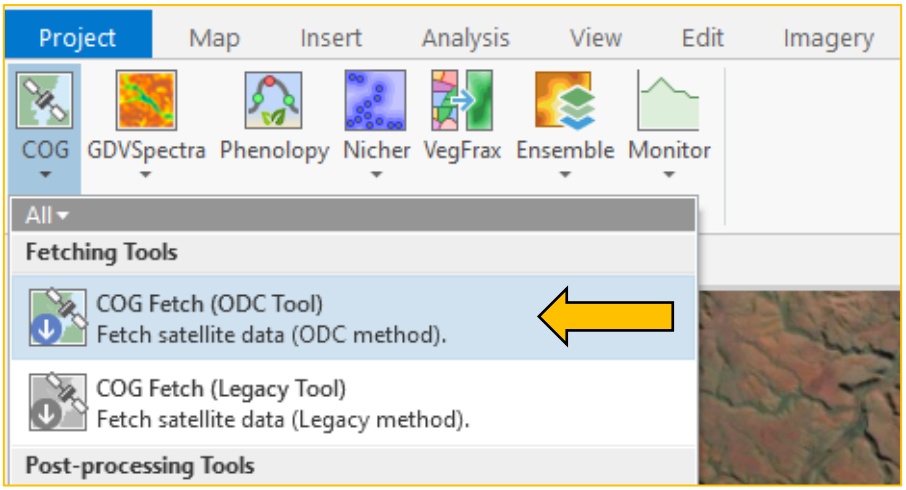
2.2.1 Considerations

- The input study area feature should be in the WGS84 (EPSG: 4326) coordinate system.
- If one (or many) polygons are selected, the minimum bounding box of the selected polygon(s) will be used, else the minimum bounding box of all features will be used.

- The cloud mask data provided by DEA is not perfect and sometimes misses a clouded image.

2.2.2 Initialise the tool

Click the COG Fetch (ODC) tool from the Tenement Tools top menu tab, like so:

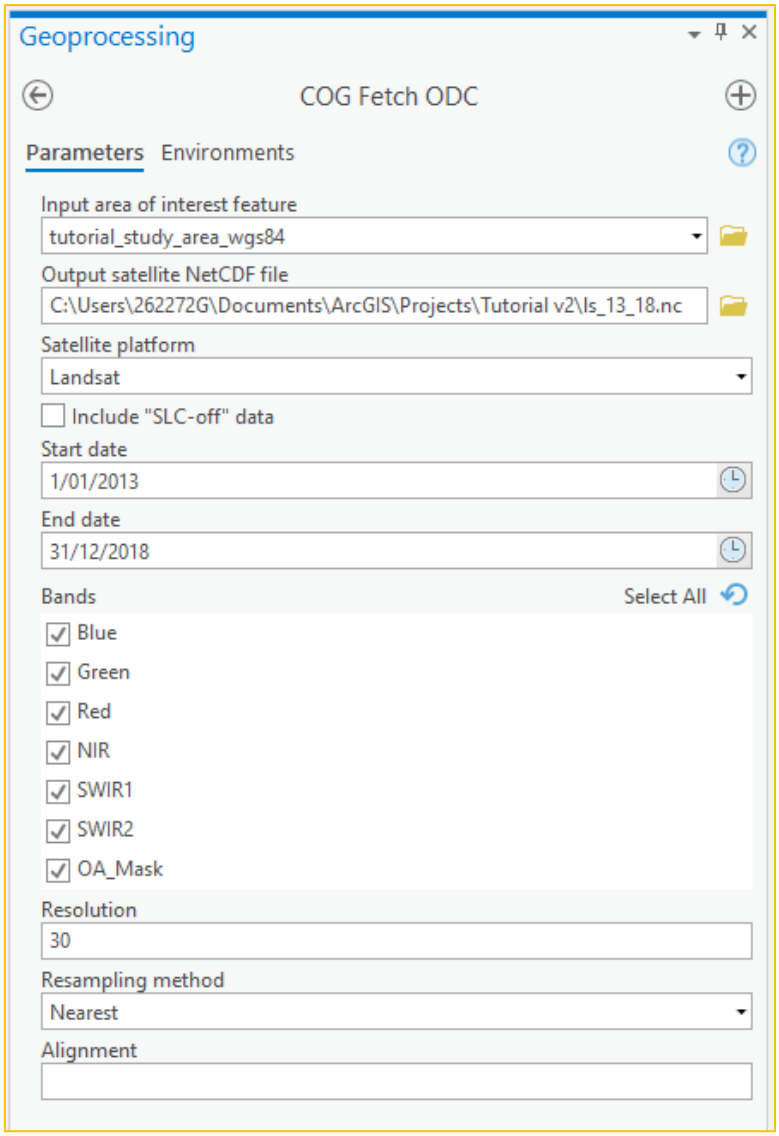


The COG Fetch (ODC) geoprocessing panel will appear on the right side of the screen.

2.2.3 Download satellite data

For the first exercise, we will download several years' worth of Landsat data for our current tutorial study area.

Set the parameters as follows:



Select the **tutorial_study_area_wgs84** feature layer as the input area of interest.

The above parameters are requesting all available **Landsat** satellite data from DEA for the period **01/01/2013** to **31/12/2018**. We've requested more satellite data than we need, but having a range of years offers some flexibility in upcoming analyses.

Save the output NetCDF file anywhere on your computer and call it **ls_13_18.nc**.

Leave the remaining parameters as default (though hover over their headings to see an explanation of what each does).

When ready, click the **Run** button at the bottom of the panel. The download should take approximately 5 - 10 minutes on a decent network.

When this process is complete, you should see a single **.nc** file approxmately 700mb in size at the output location set above.

3 COG Explore tool

3.1 Introduction

The COG Explore tool is used to visualise, explore and graph satellite imagery NetCDFs downloaded via the COG Fetch tool. Additionally, the tool converts the data into a format (a **.crf** file) that can now be used in any of ArcGIS Pro's powerful multi-dimensional analyses, greatly expanding the use of the Tenement Tools plugin.

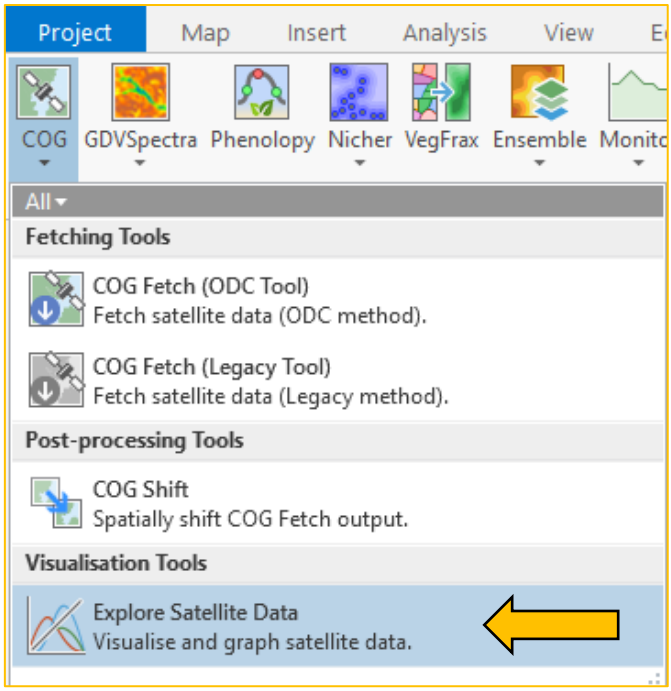
3.2 Use the COG Explore tool

3.2.1 Considerations

- The input must be a NetCDF file of Landsat or Sentinel data obtained specifically from the COG Fetch (ODC) or COG Fetch (Legacy) tools.
- Currently the tool only supports the visualisation of the NetCDF data as a vegetation index.

3.2.2 Initialise the tool

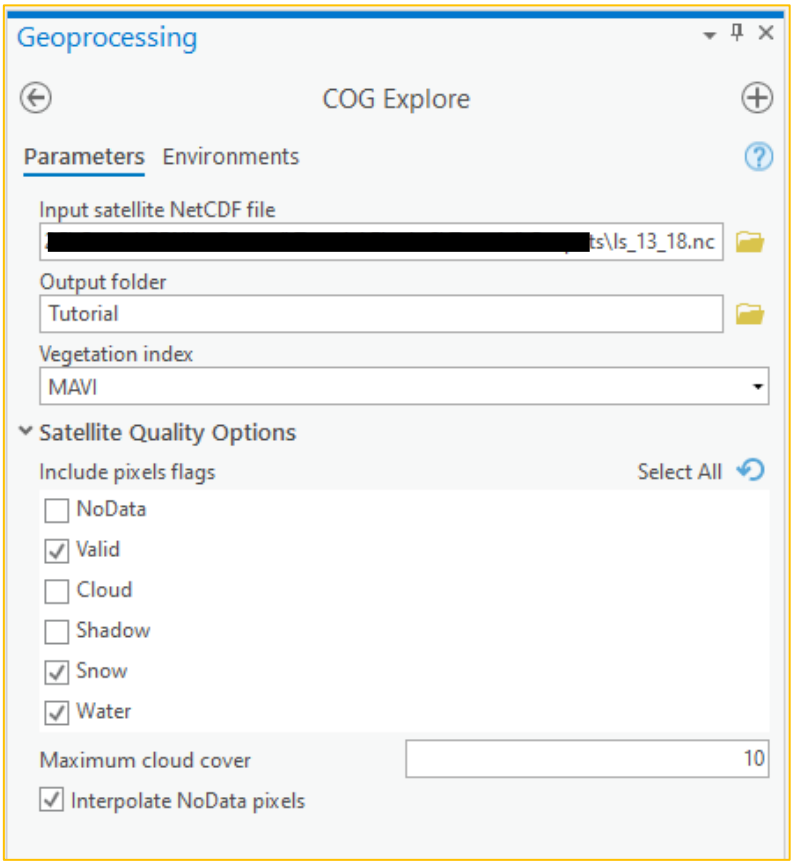
Click the COG Explore tool from the Tenement Tools top menu tab, like so:



The COG Explore geoprocessing panel will appear on the right side of the screen.

3.2.3 Visualise satellite data

It's worth spending some time visualising and exploring the vegetation of the study area over time. Set the parameters as follows:



The above parameters will generate an image displaying the vegetation index known as MAVI (Moisture-adjusted Vegetation Index; Zhu et al. 2014) for every date captured within the input raw satellite NetCDF **ls_13_18.nc**.

Set the output folder to any appropriate **folder** on your computer for the outputs of the tool (a **.crf** file). Finally, set the vegetation index to **MAVI** (though feel free to use any),

Leave the remaining parameters as default. These defaults will remove any pixels with clouds, shadows or invalid data and remove dates where cloud cover is over 10% of the study area.

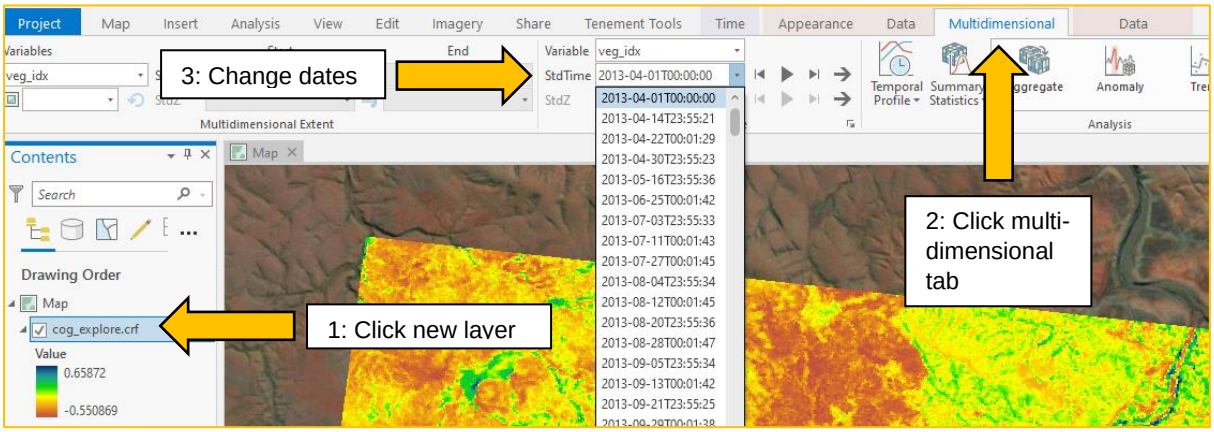
When ready, click the **Run** button at the bottom of the panel. When finished, the satellite data represented as MAVI will automatically appear on your map.

3.2.4 Visually explore MAVI over time

The output shows MAVI values. The higher the value, the "greener" the vegetation is in the landscape. Based on the output symbology, the "greenest" vegetation is coloured blue.

The first image shown on the map is the first valid date in original NetCDF, which is early April 2013. This indicates no cloud-free images existed between January and March in 2013. Additionally, this image appears to be showing a large bush fire in the north-west (this area is prone to them).

Let's see how other images look. Click on the **cog_explore.crf** layer in the contents panel (1) to enable to **Multidimensional tab** (2). Browse different images at different dates using the **StdTime drop-down** control (3):

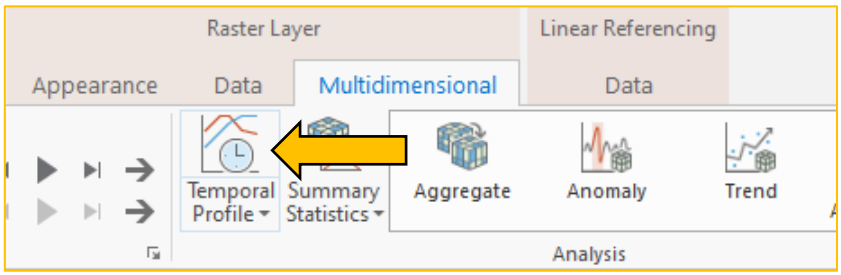


Take your time scrolling through some of the different dates to get a better idea of how the vegetation changes throughout the years.

Most of the "greennest" vegetation appears to occur in the riparian areas associated with drainage lines and deeply incised valleys.

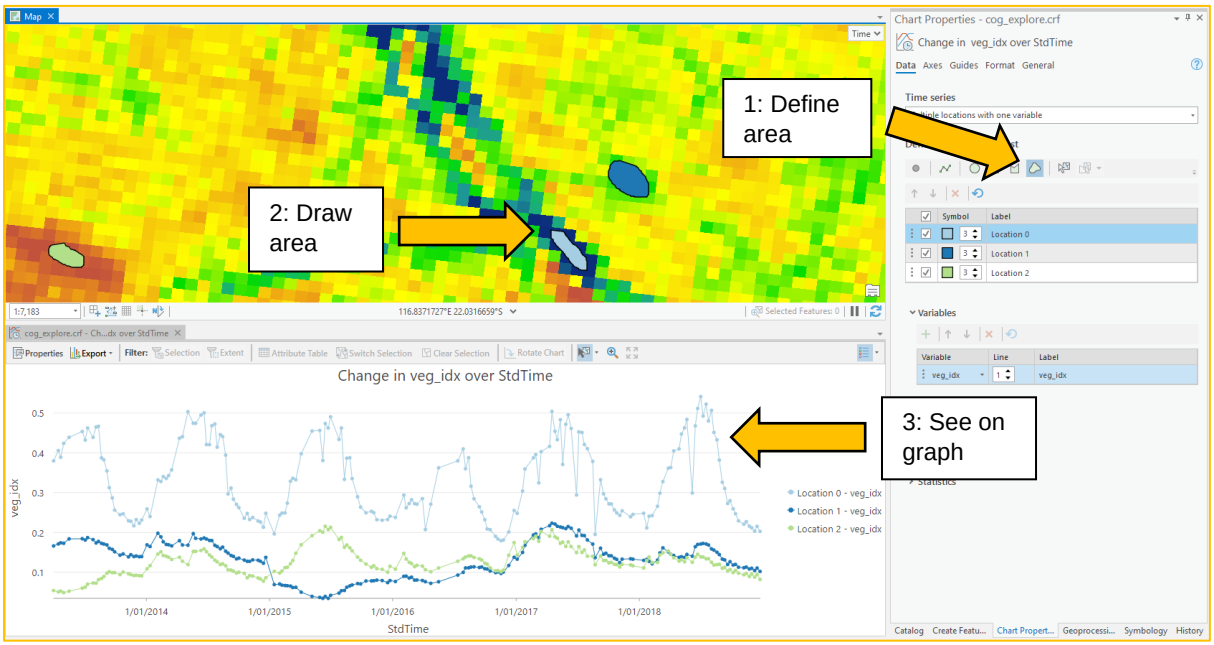
3.2.5 Visually graph MAVI over time

Another tool in our toolkit is the ability to graph MAVI over time using the temporal profile tool. On the same **Multidimensional tab** show in (2) above, click on the **Temporal Profile** button:



An empty graphing panel should appear at the bottom half of your screen along with some controls in a panel on the right.

We can now start drawing areas around different vegetation types on our current map and see the vegetation history as a graph. To create an area, select an area of interest tool (1), draw an area around an area of vegetation (2) and watch the vegetation history for that area appear in the graph (3).

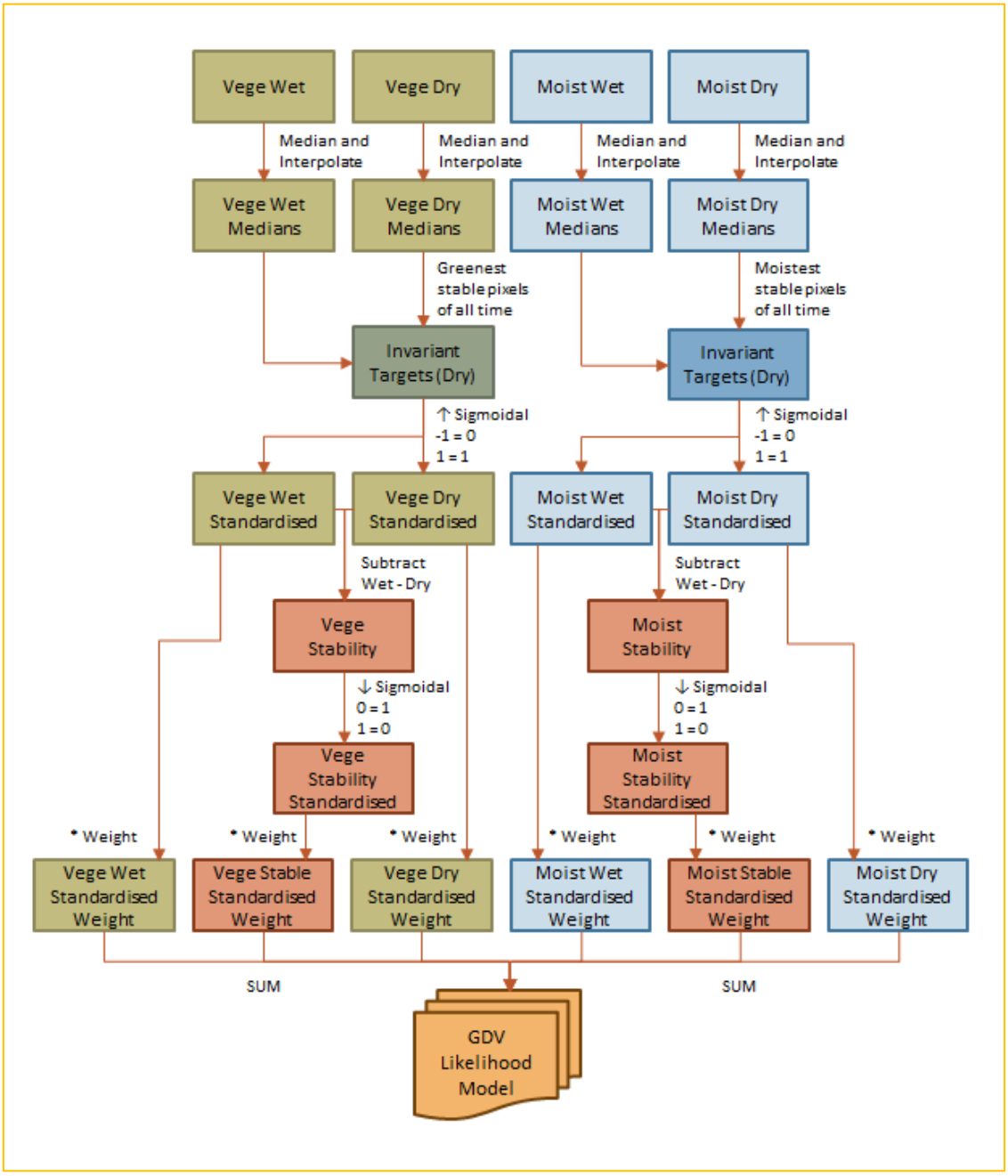


You can draw as many areas as you want. Spend some time graphing different types of vegetation. Do different types of vegetation have different seasonal cycles? What does the 2013 bush fire area look like?

4 GDV Spectra

4.1 Introduction

The GDVSpectra tool is used to detect areas of potential groundwater dependent (GDV) vegetation using a time-series of satellite data. This is achieved using seasonally-weighted vegetation and moisture information derived from satellite images that are put through an Analytical Hierarchy Process (AHP) weighting process:



The output is a series (or an aggregate if requested) of annualised GDV likelihood layers which highlight areas within a landscape that may or may not have a high likelihood of containing persistently green and moist vegetation (i.e., potential GDV).

The GDVSpectra toolset also provides several additional tools to assist in subsequent monitoring of these areas using long-term trend analysis (e.g., Mann-Kendall trend analysis) or Change Vector Analysis (CVA).

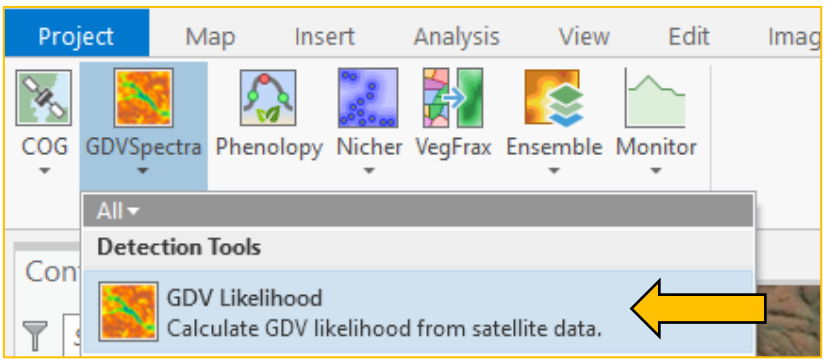
4.2 Generate GDV likelihood layers

4.2.1 Considerations

- The input must be a NetCDF file of Landsat or Sentinel data obtained specifically from the COG Fetch (ODC) or COG Fetch (Legacy) tools.
- Three or more years of data is required in order to run the tool.
- The tool is prone to seasonal fluctuation, so the combine outputs option is recommended to reduce variation.

4.2.2 Initialise the GDV Likelihood tool

Click the GDVSpectra Likelihood tool from the Tenement Tools top menu tab, like so:

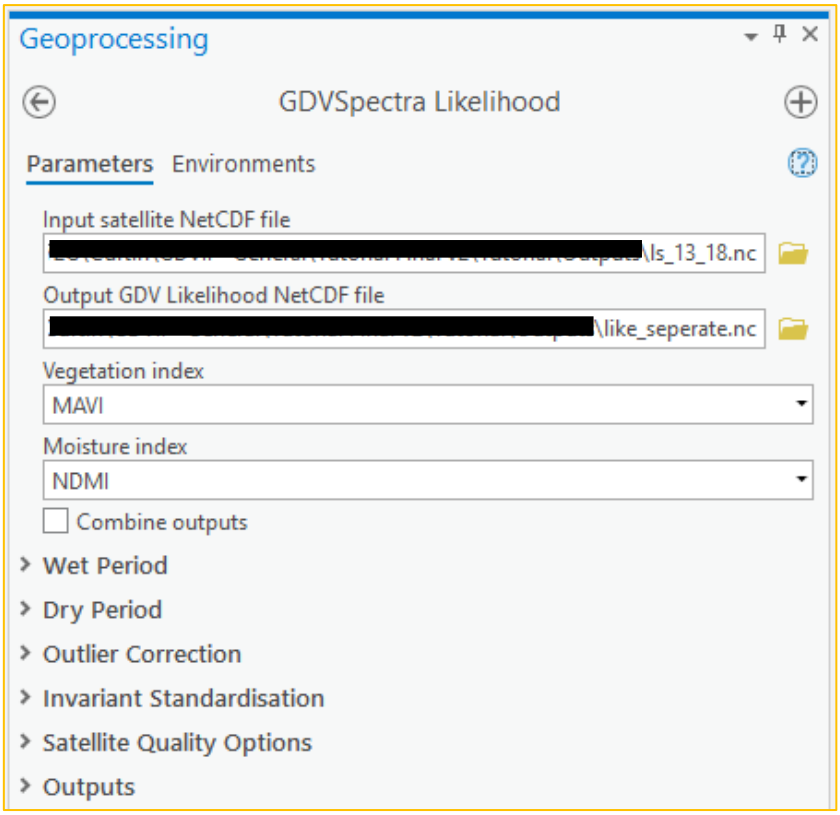


The GDVSpectra Likelihood geoprocessing panel will appear on the right side of the screen.

4.2.3 Generate likelihood layers (non-combined)

We will run this tool twice; first where we keep output years separate, and a second run where we combine all likelihood into a single image (i.e., an aggregate).

Set the parameters as follows:



Select the previous downloaded **ls_13_18.nc** as the input NetCDF file and save the output anywhere on your computer as **like_seperate.nc**.

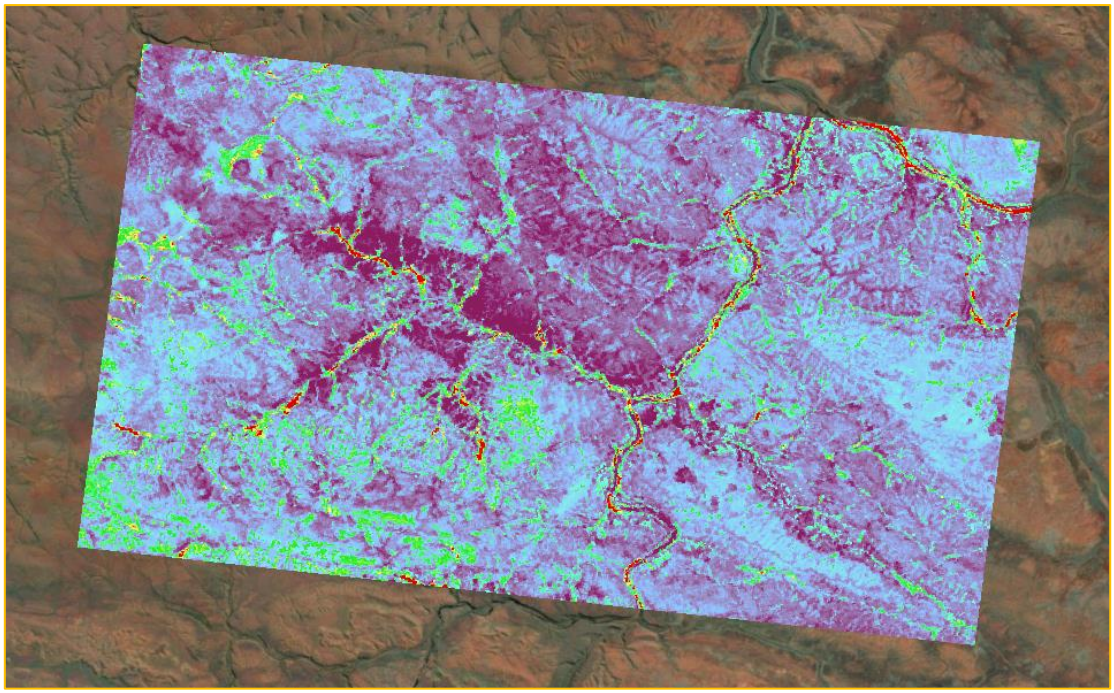
Ensure the **combine outputs** option is uncheked. This will ensure a likelihood layer is retained for each valid year within the input NetCDF (i.e., one layer for every year 2013 – 2018).

Leave the remaining parameters as **default**.

When ready, click the **Run** button at the bottom of the panel. When finished, a single **likelihood.crf** layer will be added to the contents panel on the left as well as your current map.

4.2.4 Explore the non-combined likelihood layer

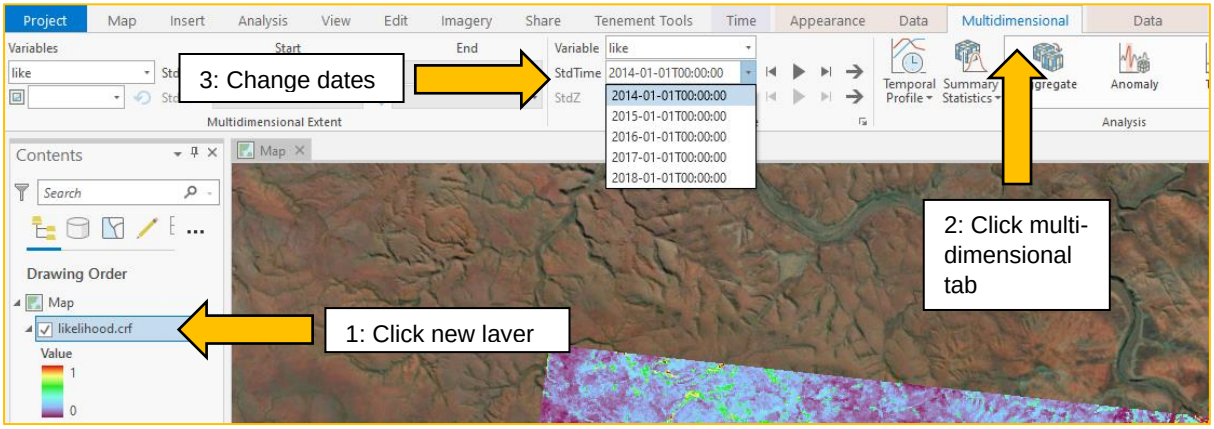
The non-combined likelihood.crf layer should appear similar to below:



The output shows likelihood values. The higher likelihood values appear red. These red areas represent areas in the landscape that were most green, moist and stable for the currently visualised year.

The first image shown on the map is the first year in the non-combined likelihood output, which is 2014 (2013 lacked sufficient seasonal information).

As before, you can visualise different years by clicking on the **likelihood.crf** layer in the contents panel (1) to enable to **Multidimensional tab** (2). Browse different years using the **StdTime drop-down** control (3):

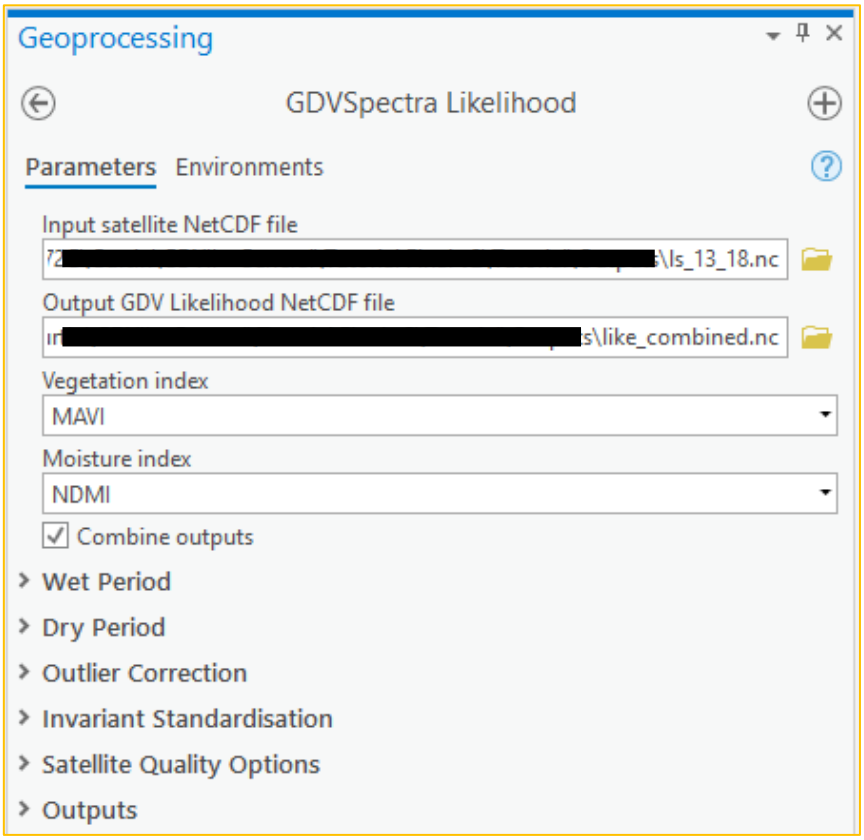


Take some time to scroll between different years. The highest likelihood areas are unlikely to change too dramatically across years, though you may see some fluctuation for the lower values (i.e., *Triodia* grasslands or other shrublands).

4.2.5 Generate likelihood layers (combined)

Run the tool again, except this time let's combine the likelihood outputs. This can produce a more refined GDV likelihood layer.

Set the parameters as follows:



As before, select the previous downloaded **ls_13_18.nc** as the input NetCDF file and save the output anywhere on your computer as **like_combined.nc**.

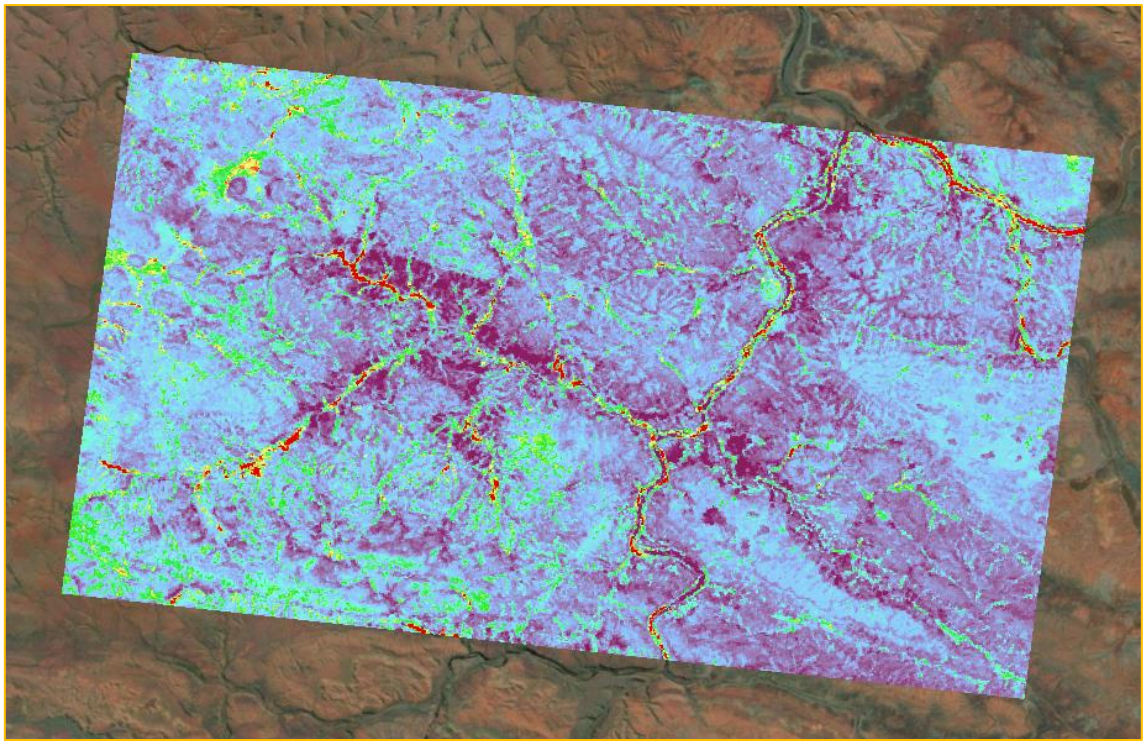
Ensure the **combine outputs** option is checked. This will combine all likelihood layers into one, showing an average of the likelihood output for the period 2013 – 2018. This can reduce seasonal variation and provide a more robust result.

Leave the remaining parameters as **default**.

When ready, click the **Run** button at the bottom of the panel. When finished, a new **likelihood.crf** layer will be added to the contents panel on the left as well as your current map automatically.

4.2.6 Explore the combined likelihood layer

The combined likelihood.crf layer should appear similar to below:



Spend some time investigating the output. Where do the highest likelihood values for the period 2013 – 2018 tend to occur?

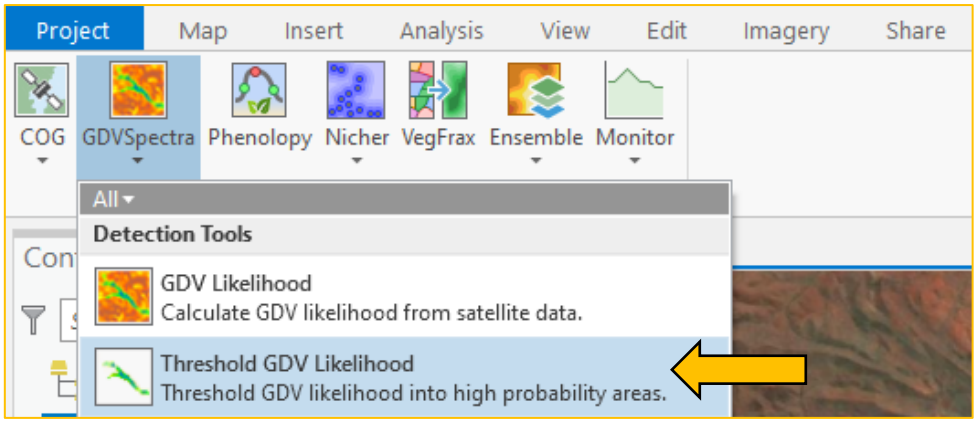
4.3 Threshold GDV likelihood layer

4.3.1 Considerations

- The input must be a NetCDF file of non-combined GDV Likelihood data generated prior using the GDVSpectra Likelihood tool.
- When using an occurrence shapefile, ensure the data is projected in the GDA94 Albers Australia (EPSG:3577) projection system.
- Presence (1s) and absence (0s) must be included in the input occurrence shapefile.

4.3.2 Initialise the tool

Click the GDVSpectra Threshold tool from the Tenement Tools top menu tab, like so:

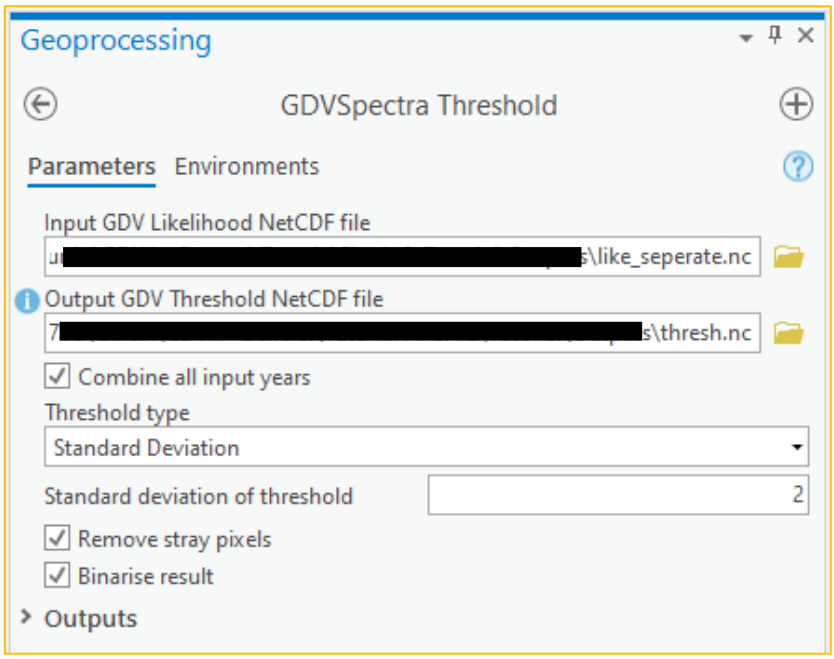


The GDVSpectra Threshold geoprocessing panel will appear on the right side of the screen.

4.3.3 Threshold GDV likelihood layers (non-combined)

The threshold tool can be used to isolate areas of highest likelihood from the GDV likelihood data generated via the GDVSpectra Likelihood tool using a thresholding technique. This technique can utilise either existing field presence and absence GDV species locations or standard deviation. We will apply the latter.

Set the parameters as follows:



Select the previous generated non-combined likelihood NetCDF **like_seperate.nc** as the input GDV Likelihood NetCDF file and save the output anywhere on your computer as **thresh.nc**.

Ensure the **combine all input years** is checked. This will threshold all likelihood data in combination. If you don't want to combine data, you may select one or more specific years instead.

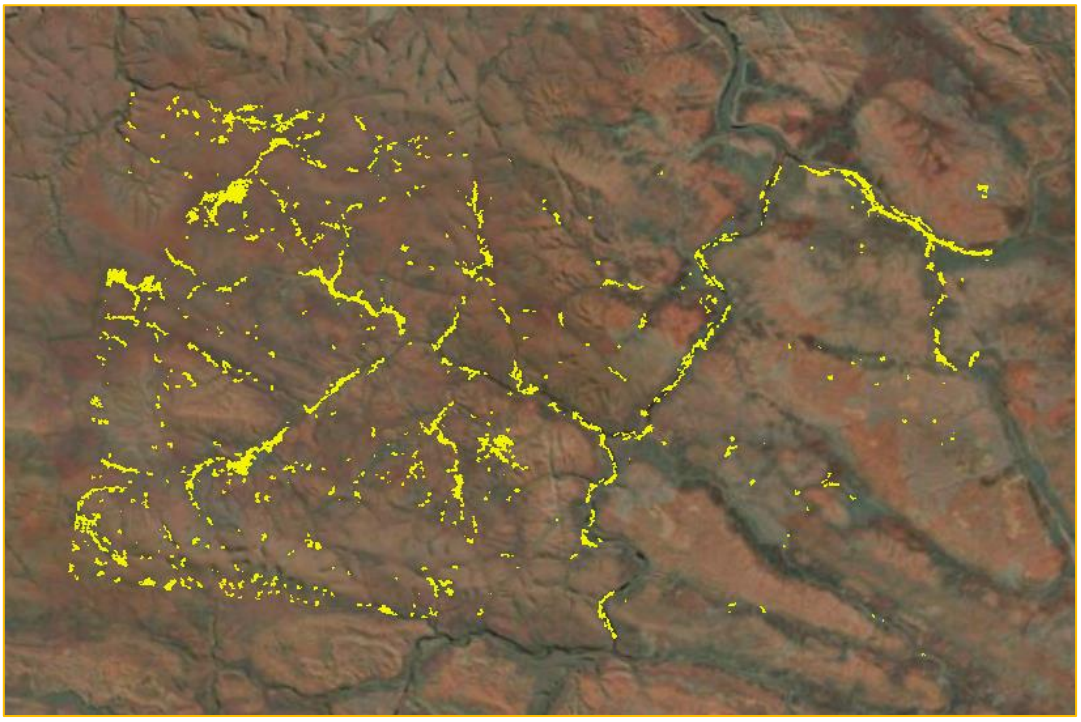
Set the **standard deviaton** value to 2. Increasing this value will return less but more sigificantly high likelihood areas.

Leave the remaining parameters as **default**.

When ready, click the **Run** button at the bottom of the panel. When finished, a new **likelihood_threshold.crf** layer will be added to the contents panel on the left as well as your current map automatically.

4.3.4 Explore the thresholded GDV likelihood

The thresholded GDV likelihood layer should appear similar to below:



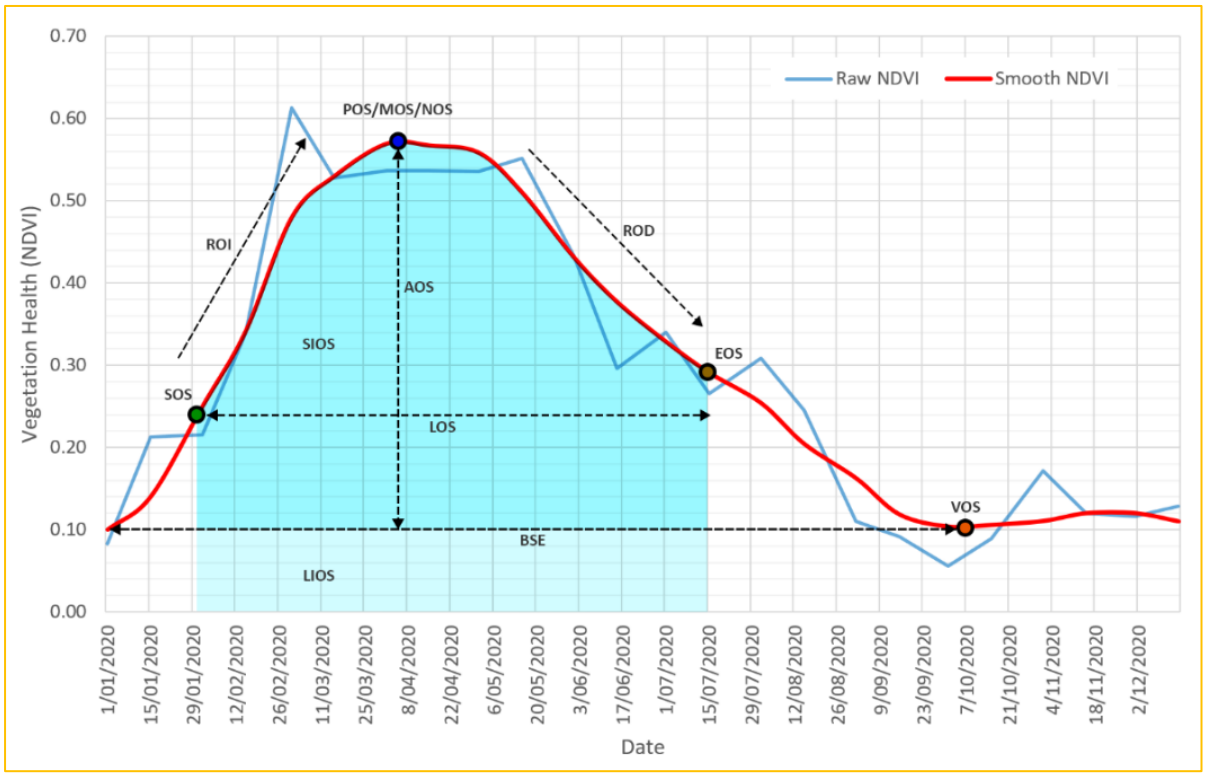
Spend some time investigating the output. How did the thresholding method go? If it over- or under-estimated high likelihood areas, try re-running the tool with a higher or lower standard deviation.

There are two more tools remaining within the GDVSpectra toolset: **trend analysis** and **Change Vector Analysis (CVA)**. These are used for assessing vegetation (and other) changes within a study area over time. As we will focus on change detection later when we look at the NRT Monitoring tools, we will leave these tools for then.

5 Phenolopy

5.1 Introduction

The Phenolopy (phenology + python) tool is used to combine a year or more worth of satellite data with the specific goal of extracting vegetation phenological characteristics. Over a dozen phenological metrics (i.e., phenometrics) representing various characteristics can be derived:



Metrics of any type of vegetation, from agricultural crops to native vegetation, can be extracted from satellite imagery using this tool. The methodology is based on the popular TIMESAT 3.3 software (http://web.nateko.lu.se/timesat/timesat.asp; Eklundh and Jönsson, 2012). See Appendix 16.1 for a deeper explanation of each metric.

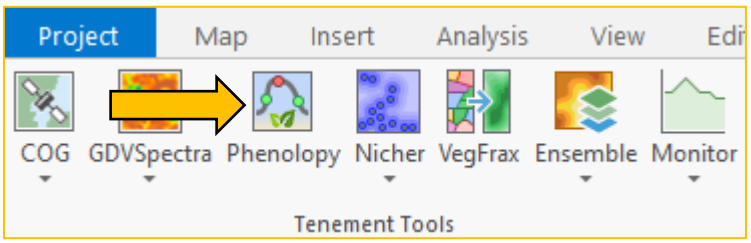
5.2 Use the Phenolopy tool

5.2.1 Considerations

- The input must be a NetCDF file of Landsat or Sentinel data obtained specifically from the COG Fetch (ODC) or COG Fetch (Legacy) tools.
- Phenolopy is only able to derive metrics for one year of data at a time. If multiple years are selected in the tool, the median of these years is used.

5.2.2 Initialise the tool

Click the Phenolopy tool from the Tenement Tools top menu tab, like so:



The Phenolopy geoprocessing panel will appear on the right side of the screen.

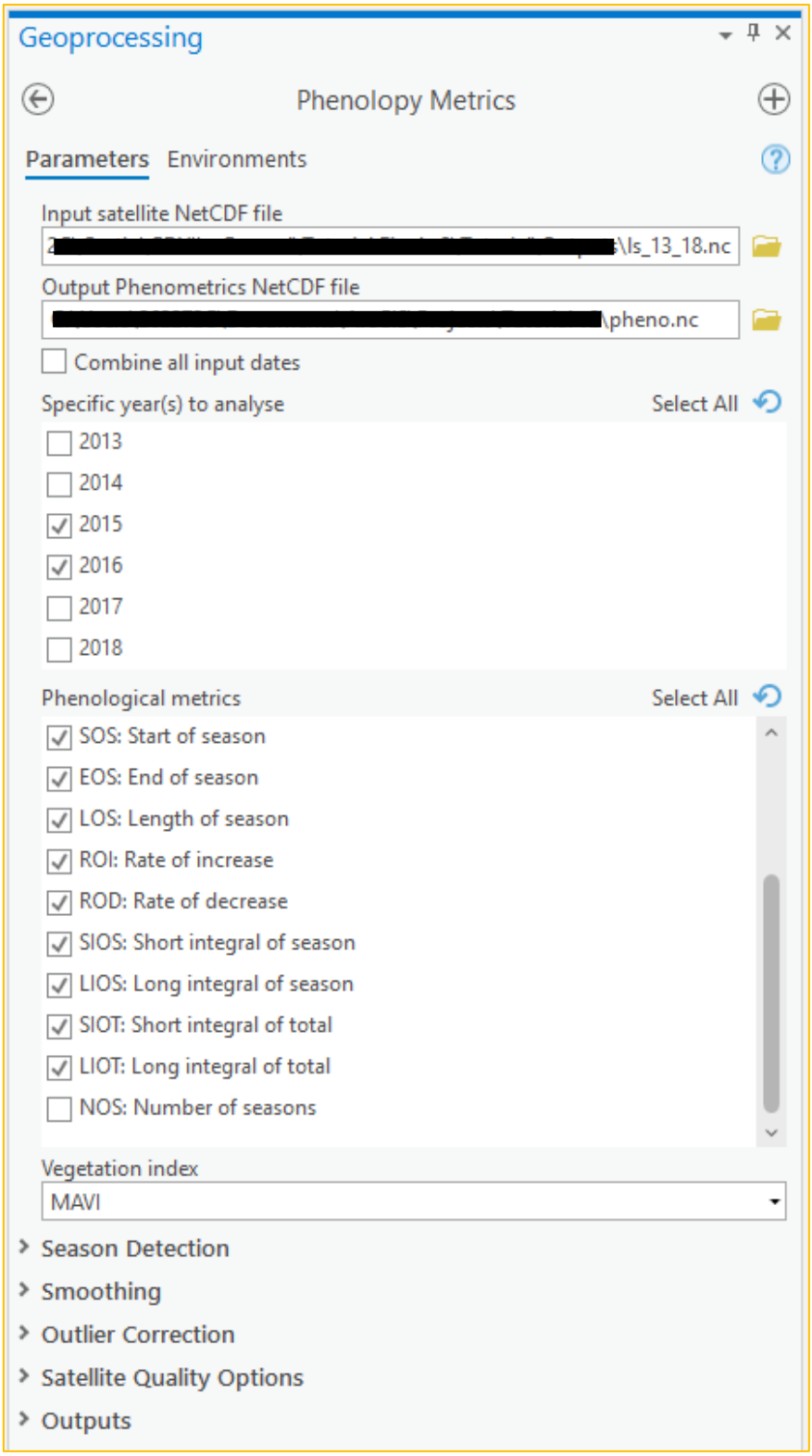
5.2.3 Generate phenological metrics

The Phenolopy tool provides numerous options for generating phenological metrics. In most cases, the default parameters can be considered optimal (these are also the defaults in TIMESAT).

However, the output is highly influenced by the quality of the satellite data. Based on a quick investgation of the input Landsat data, most years appear to contain sigifniciant bushfire events. Have a look for youself using the COG Explore tool.

This makes it somewhat difficult to choose an optimal year. The years least effected by bushfire appear to be 2015 and 2016, so let's use those years in the Phenolopy tool.

Set the parameters as follows:



Select the previous downloaded **ls_13_18.nc** as the input NetCDF file and save the output anywhere on your computer as **pheno.nc**.

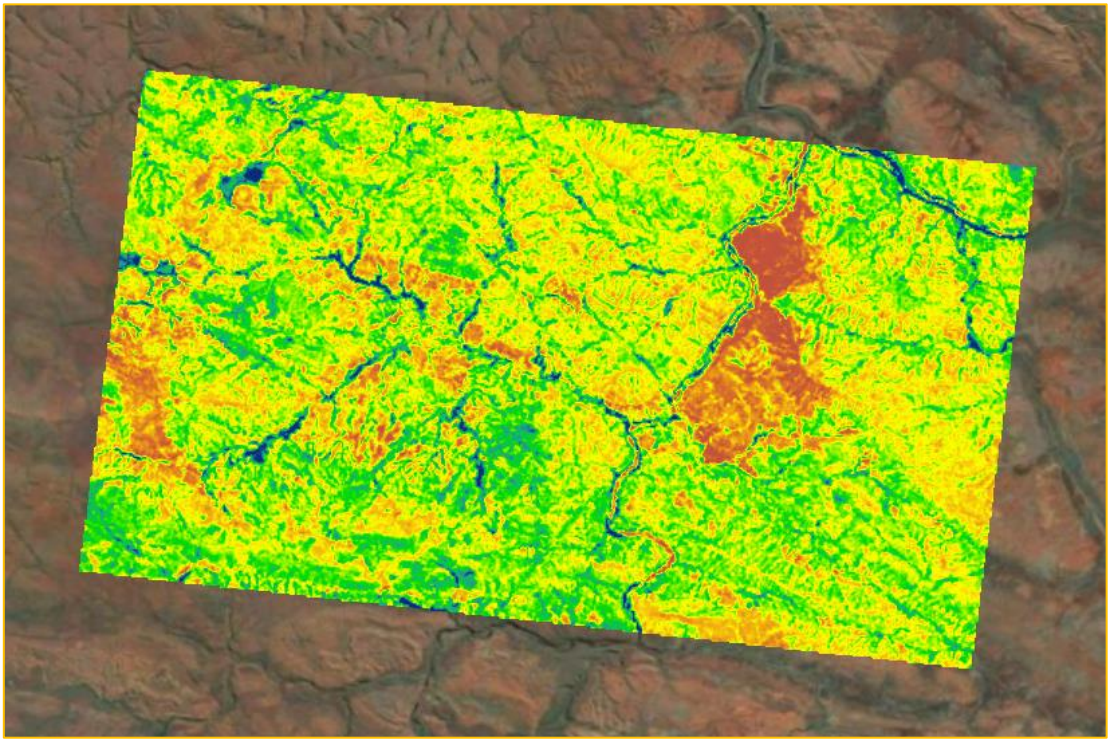
Unchek the **combine all input dates** option and select years **2015** and **2016** to use the median of these two years, which were our most bushfire-free years.

Leaving the rest as **default** is recommended, but go right ahead and experiment. For example, try adding more years. Try a different seasonal detection method. Also, try using Sentinel data for 2017 instead.

When ready, click the **Run** button at the bottom of the panel. When finished, numerous metrics will be added to a 'metrics' group in the contents panel on the left as well as your current map automatically.

5.2.4 Explore the various metrics

Turn all layers off expect the **liot_values** layer. LIOT (long integral of total) is considered a surrogate for total vegetation productivity per year (Eklundh and Jönsson, 2012). Here, the dark blue areas are the highest producing areas for the years 2016 and 2017. These appear to be related to drainage lines and riparian zones, and in our tests were strongly correlated with GDV.



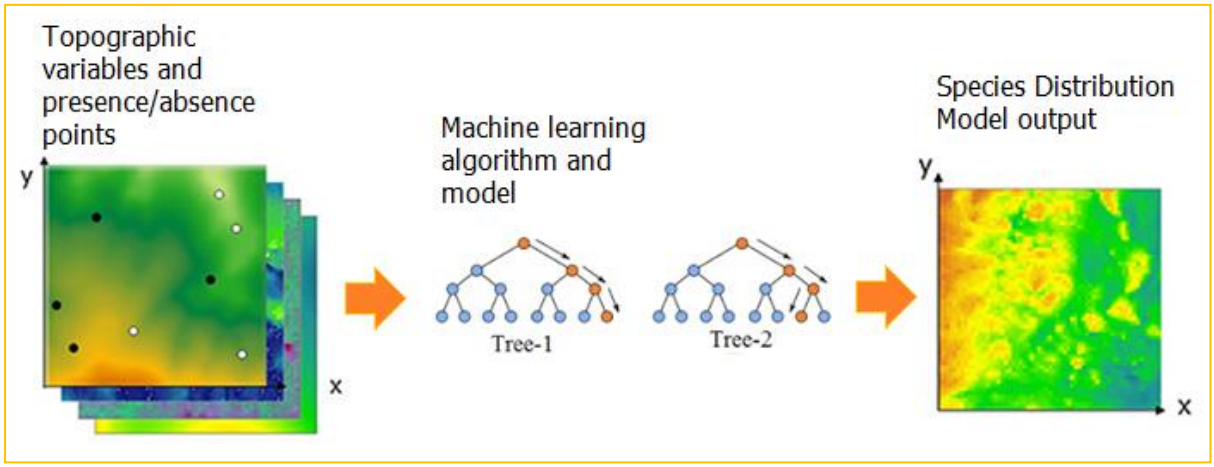
Check out some of the other metrics. From these metrics we can determine approximately what day of year certain vegetation started and ended its growing season (SOS, EOS), when its peak growth occurred (POS), and how rapidly it grew (ROI, ROS), just to name a few.

Metrics with the suffix **values** show values representing MAVI, whereas the suffix **times** values represent the day of year.

6 Nicher

6.1 Introduction

The Nicher tool is used to generate a species distribution model (SDM), also known as an ecological niche model (ENM). These models combine known locations of a target species (i.e., occurrences) and environmental data layers representing potential habitat with modern machine learning algorithms to correlate and predict potential distribution across a wider landscape.



Generally speaking, SDMs begins with user-provided occurrence points, which represent a single species (e.g., flora or fauna observations). Absence locations are then automatically generated by the model (i.e., pseudo-absences). These samples are then coupled with elevation-derived topographic variables and/or climate rasters to construct a statistical model using machine learning algorithms.

Once the algorithms have learned where species occur and under what environmental conditions, the probability of that species occurring elsewhere is estimated as some function of the environmental conditions of that place (Franklin, 2010).

6.2 Use the Nicher tool

6.2.1 Considerations

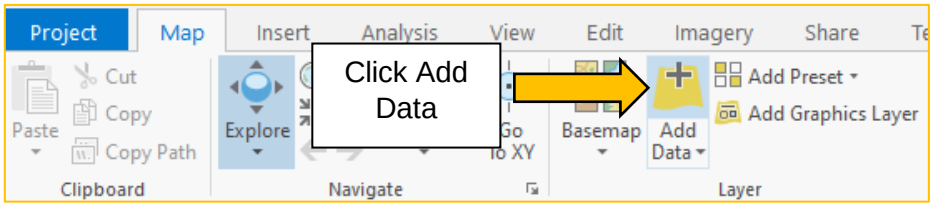
- The input data (occurrence points and raster layers) must be projected in the GDA94 Australian Albers (EPSG: 3577) projection.
- Input rasters must be GeoTiffs (.tif).
- All input data must spatially overlap each other or the tool will abort.
- The occurrence point data assumes all records are indicative of species presence.
- Infrastructure can be masked out of the SDM using the Nicher Masker tool.

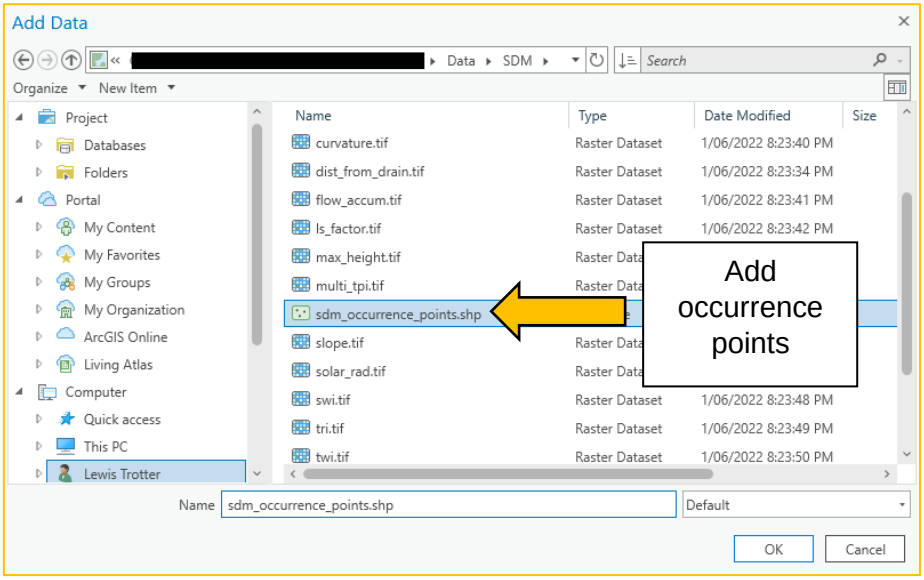
6.2.2 Add and explore provided data

Occurrence points and eighteen topographic variables have been included in the tutorial data.

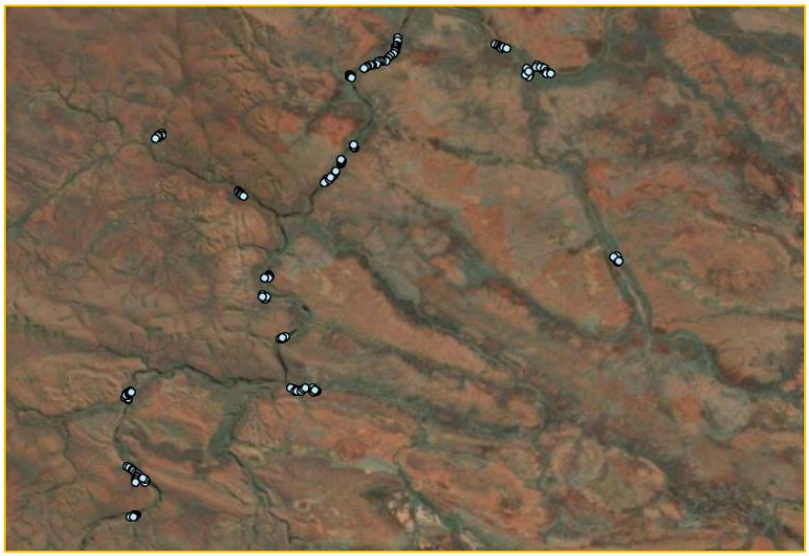
The occurrence points contains over 170 **simulated** locations of riparian zone *Eucalyptus* trees that were derived from aerial imagery.

Let's add the occurrence points to the map. Click Add Data button on the top menu and add the **sdm_occurrence_points.shp** file found within the SDM folder from the tutorial data:





The occurrence should appear on your map, like so:



The eighteen topographic variables represent various topographic phenomena within the study area and are all generated from a single 10m resolution digital elevation model (DEM).

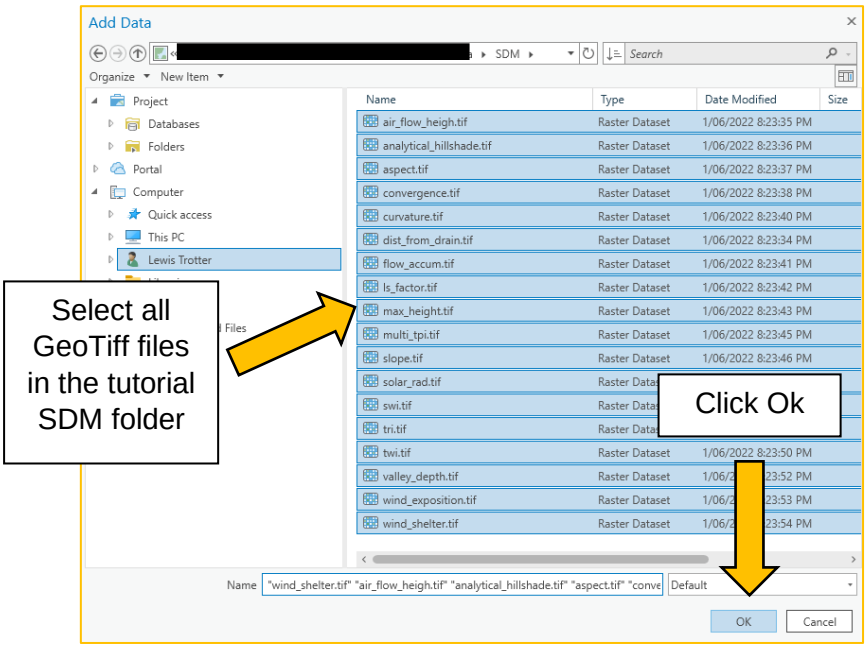
We recommend the free SAGA GIS software (http://www.saga-gis.org) for generating topographic variables from elevation data.

For your reference, the eighteen variables provided are summarised below:

Variable	Filename
Slope	slope.tif
Aspect	aspect.tif
Convergence	convergence.tif
Curvature	curvature.tif
Multi-scale Topographic Position Index (TPI)	multi_tpi.tif
Topographic Ruggedness Index (TRI)	tri.tif
Topographic Wetness Index (TWI)	twi.tif
SAGA Wetness Index	swi.tif
Flow Accumulation	flow_accum.tif
Landscape Factor	ls_factor.tif
Valley Depth	valley_depth.tif
Wind Exposition	wind_exposition.tif
Analytical Hillshade	analytical_hillshade.tif
Solar Radiation	solar_rad.tif
Distance to Drainage	drain_dist.tif
Air Flow Height	air_flow_height.tif
Wind Shelter	wind_shelter.tif
Max Height	max_height.tif

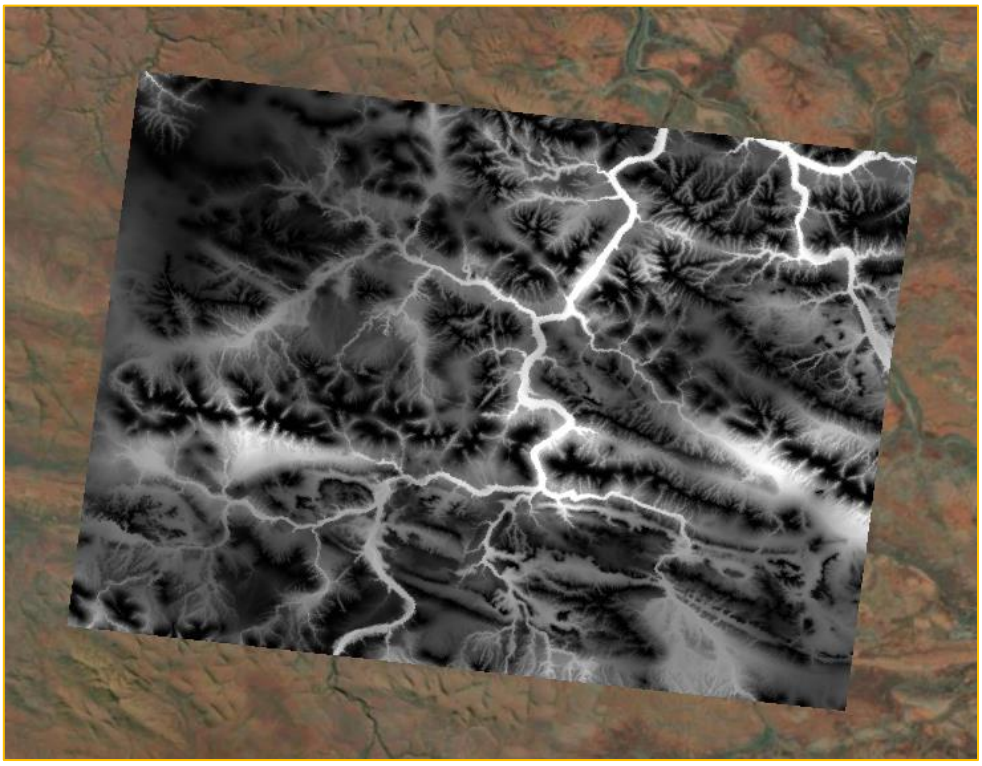
Many of these variables are commonly used in SDMs, though that doesn't mean all may be relevant here (e.g., Ball et al., 2020; Keppel et al. 2015; Yates et al. 2019).

Let's add the topographic variables to the current map and have a quick look. Click the Add Data button as above and **add all eighteen topographic variables**, like so:



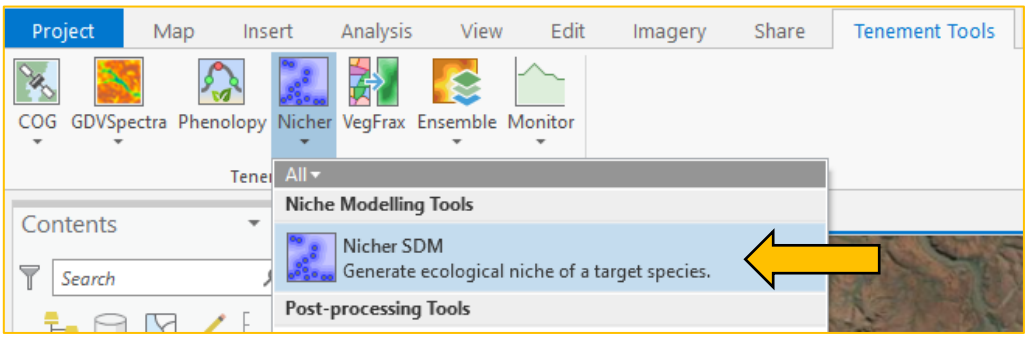
Spend some time exploring the occurrence points and topographic layers on your map. Try changing the symbology to these layers.

This is how valley depth appears on my map with default symbology:



6.2.3 Initialise the tool

Click the Nicher tool from the Tenement Tool top menu tab, like so:

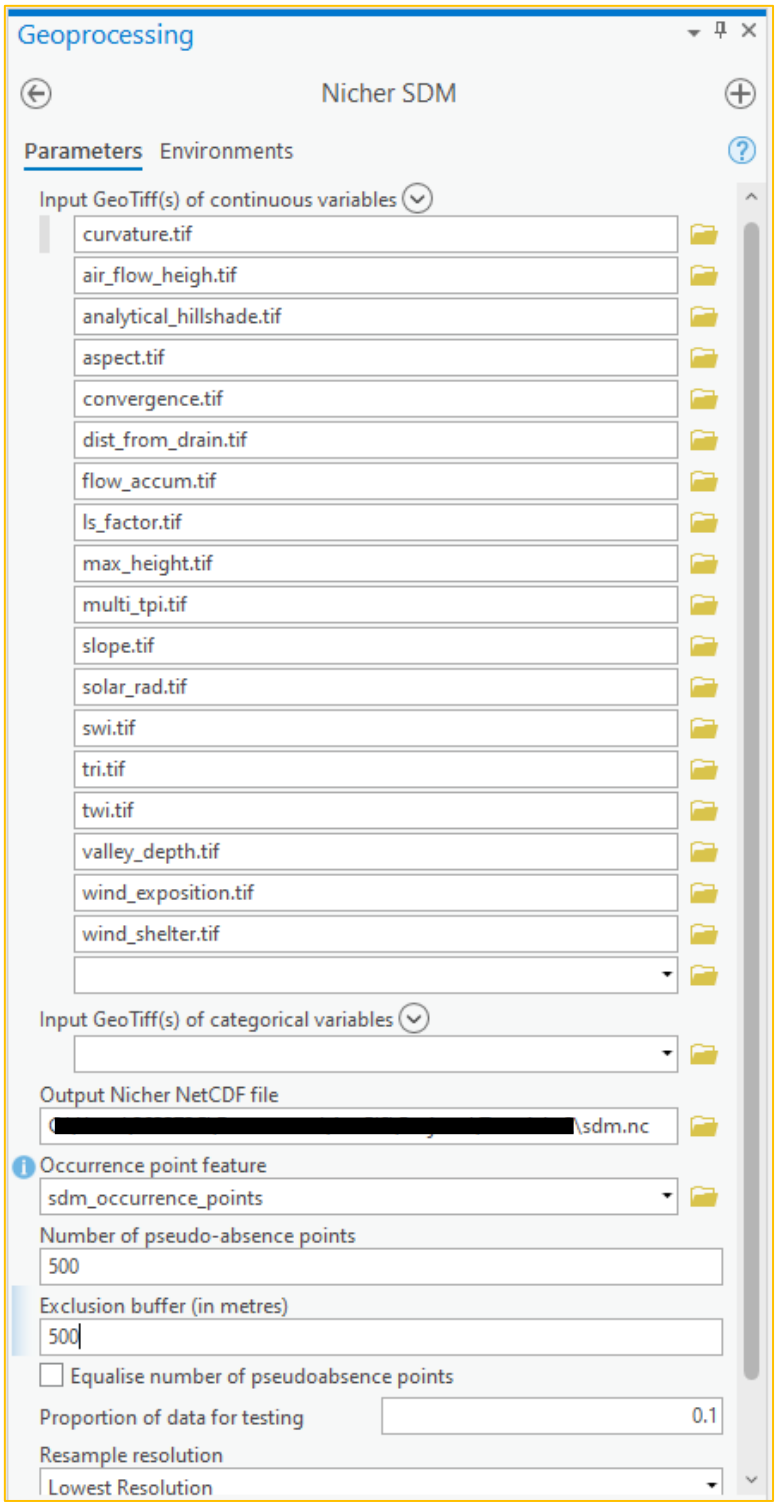


The Nicher geoprocessing panel will appear on the right side of the screen.

6.2.4 Generate species distribution model

The Nicher tool provides numerous options for generating SDMs. In most cases, the default parameters can be considered optimal, though experimentation is recommended.

Set the parameters are follows:



Add all topographic variables to the **continuous variables** control. We don't have any categorical layers (e.g., soil mapping), so leave the **categorical variables** control blank.

Save the output Nicher NetCDF file anywhere on your computer and call it **sdm.nc**.

Set the occurrence point feature to the **sdm_occurrence_points.shp** added earlier.

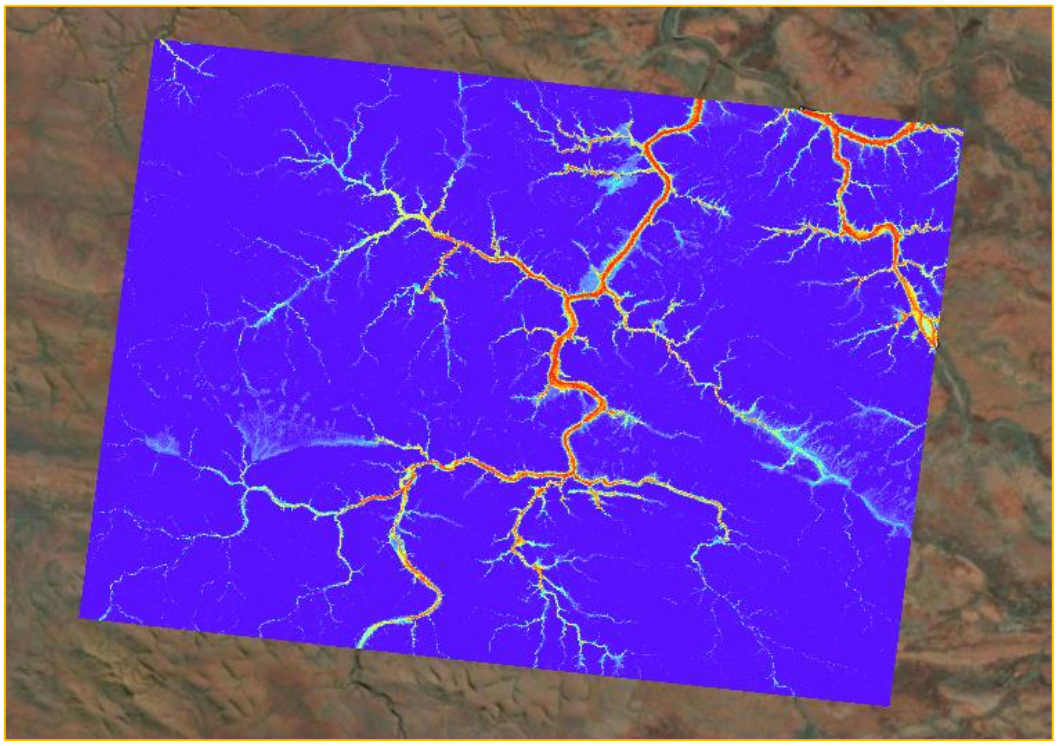
Change the number of generated **pseudo-absence points** to 500. While there is no general rule of thumb on selecting pseudo-absence count, typically the number is larger than occurrences by a significant amount (Barbet-Massin et al., 2012).

Leaving the **rest as default** is recommended, but go right ahead and experiment.

When ready, click the **Run** button at the bottom of the panel. This process shouldn't take long.

6.2.5 Explore the Nicher SDM

A single SDM model output will appear on your map when the tool is finished, like so:



The **'hotter' red colours** are values closer to 1 and represent areas in the landscape that are indicative of the **most suitable habitat** for our target species. In contrast, **'colder' blue colours** are areas that are **highly unlikely** to be suitable habitat.

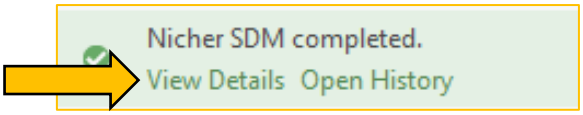
You can see how the SDM has highlighted the drainage lines and riparian areas where the vast majority of our simulated occurrence points were placed.

Take some time to explore the output or modify the model parameters and inputs.

6.2.6 Assess model performance

Understanding how the SDM **statistically performed** is an important step in the SDM process.

Nicher hides this information in the ArcGIS Pro result output window. To explore model performance information, **click the View Details link** that appears in the tool panel when the tool has finished processing:



A window will appear. Click the **Messages** tab and scroll down to the Variance Inflation Factor:

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Variance Inflation Factor Scores (VIFs)
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
General rule of thumb:
1 = No multicollinearity.
1 - 5 = Moderate multicollinearity.
> 5 = High multicollinearity.
> 10 = This predictor should be removed from the model.

                Variable  VIF Score
17          wind_shelter      1.232
3                 aspect      1.247
7              ls_factor      1.258
5        dist_from_drain      1.265
11             solar_rad      1.280
0              curvature      1.481
12                   swi      1.682
8             max_height      1.693
15          valley_depth      1.797
6             flow_accum      2.064
9              multi_tpi      2.198
4            convergence      2.217
14                   twi      2.445
1         air_flow_heigh      2.710
16       wind_exposition      3.848
10                 slope      4.117
2   analytical_hillshade      4.776
13                   tri      5.249
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

VIF measures the severity of multi-colinearity amonst our input variables. Typrically, you want to remove any of your highly correlated variables.

Based on the **general rule of thumb information**, do you have any highly correlated variables? You may want to remove any variable with a VIF > 5 and re-run the tool.

Another helpful measure is the Mean Variable Importance Score:

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Mean Variable Importance Scores
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
General rule of thumb:
The importance of variables in
the SDM model are listed below.
The higher the value, the more
important the variable.

           Variable Name  Mean Importance Score
16       wind_exposition               0.272596
1         air_flow_heigh               0.199927
15          valley_depth               0.189063
5        dist_from_drain               0.095047
8             max_height               0.071672
4            convergence               0.027678
9              multi_tpi               0.023161
14                   twi               0.016382
6             flow_accum               0.016142
11             solar_rad               0.015090
7              ls_factor               0.013746
13                   tri               0.011100
10                 slope               0.010337
12                   swi               0.009970
2   analytical_hillshade               0.008920
0              curvature               0.007667
3                 aspect               0.007639
17          wind_shelter               0.003866
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

MVI (or variable importance) indicates how important each topographic variable is within the SDM model prediction model. Generally, **the lower the value, the less important it is in predicting the target species**. This measure can also help you remove useless variables to improve the accuracy or robustness of your SDM.

To understand how the SDM model actually performed, two measures are provided.

First, training accuracy is provided:

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Model Training Accuracy Summary
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Mean Accuracy: 96.22. Mean Std. Deviation: 4.49.
Minimum Accuracy: 93.44. Maximum Accuracy: 99.18.
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

This measure provides some insight into how well the model was trained using the occurrence point training set automatically by the tool.

Generally, the Reciever Operating Characteristic (**ROC**) Area Under the Curve (**AUC**) metric is considered a better measure for model performance (e.g., Narkhede 2018) as it compares predicted values against real, known outcomes (i.e., testing set values).

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Area Under the Curve (AUC) Metrics
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
General rule of thumb:
The ROC AUC metric is indicative
of how well the current model was
able to predict the target feature.
A value <= 0.5 is a useless model. A
value closer to 0.75 is considered moderate
and a value of 1.0 is considered perfect.
The PR (Precision-Recall) AUC is considered
more robust when more absence points
exist than presence points.

ROC AUC: 1.0. PR AUC: 1.0
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

The AUC (and the stricter Precision-Recall Score AUC), **both ranked perfectly** (1.0). This is a perfect model based on our current occurrence points and chosen topographic variables. Your model result should be similar, though not necessarily exact. In reality, it is **unlikely you will score an AUC this high** using real-world data (we are using simulated data).

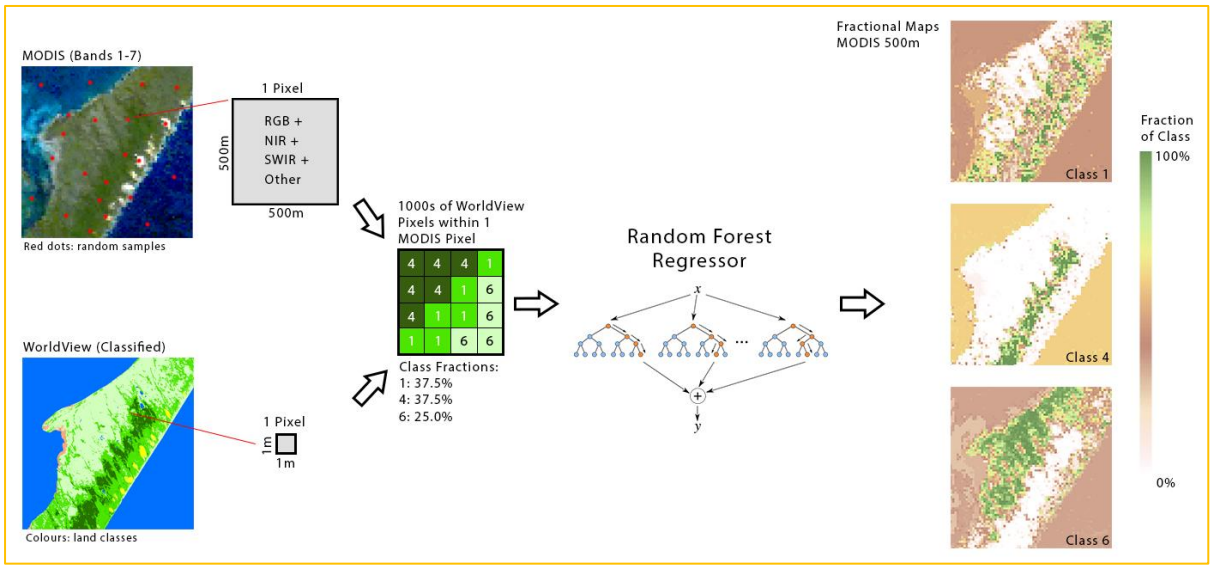
For comparison, tests using real occurrence points captured in the field resulted in AUCs ranging anywhere from 0.75 to 0.95, depending on the occurrence points and topographic variables used.

Finally, removing correlated and less important variables will also likely influence the AUC values. It is possible this model is overfitted.

7 VegFrax

7.1 Introduction

The VegFrax tool is used to extrapolate high-resolution classified map information (e.g., vegetation community classes) on to larger extents of lower-resolution satellite imagery (e.g., Landsat) using class fractions and random forest regression:



The method extracts pixel values from a high-resolution classified image that fall within the extents of larger randomly selected pixels of a lower-resolution satellite data. In other words, if a high-resolution raster has a pixel size of 1m and the moderate image has a pixel size of 10m, each random sample will have a window that captures 100 1m class pixels within.

Next, the fraction of each class per window is calculated. For example, 25 *Eucalyptus* woodland pixels would = 25% within a window. A model is built from these frequency values and used to predict the fractional values onto the lower-resolution pixels satellite band information.

The major benefit of this method is that small extents of high-resolution data can be extrapolated out onto much larger extents over a landscape.

7.2 Use the VegFrax tool

7.2.1 Considerations

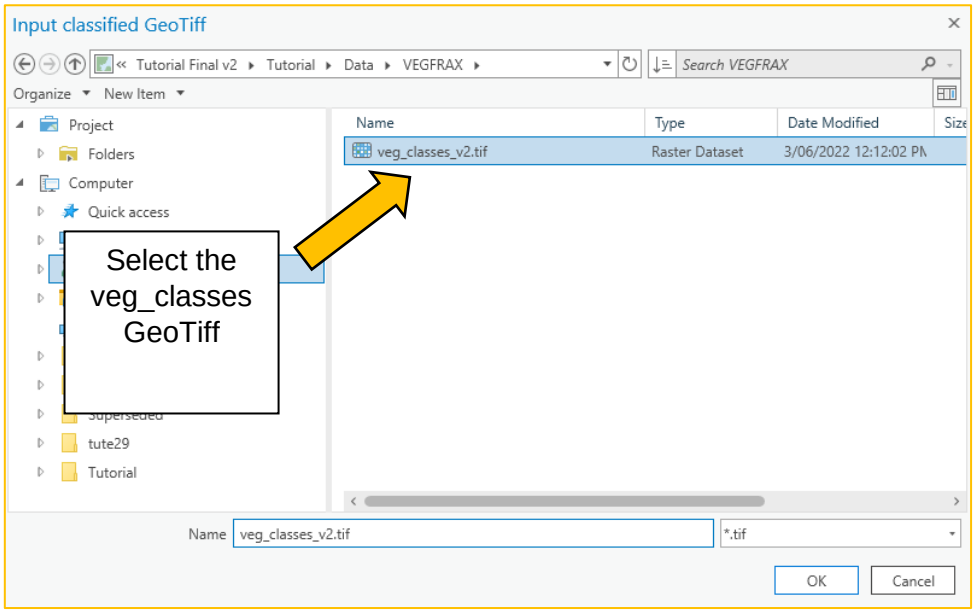
- The input data for this tool must be projected in the GDA94 Australia Albers (EPSG: 3577) projection.
- The input classified GeoTiff must be higher resolution than the input satellite NetCDF.
- The input data must spatially overlap.

- Avoid including classes with very few pixels as they may not be captured adequately when generating training samples.

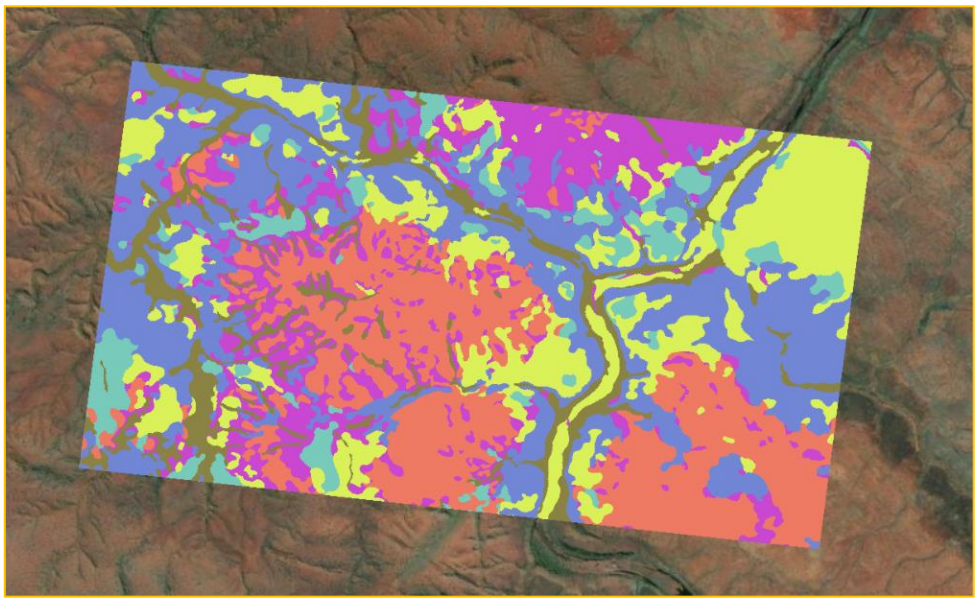
7.2.2 Add and explore the classified GeoTiff

A high-resolution, **simulated** classified vegetation GeoTiff layer has been provided in the tutorial data. This layer has six vegetation types (classes 0 – 5).

Let's add this data to the map. Once again, click the Add Data button and navigate to the VEGFRAX folder in the tutorial data:



The high-resolution vegetation class map will appear on your map. Here's how mine looks (your colours are likely different):



The six classes in this layer represent **estimated** broad vegetation community types. These communities were generated using an un-supervised clustering algorithm on a high-resolution (1m resolution) RGB satellite image.

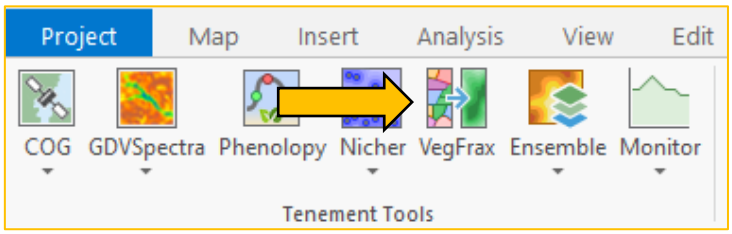
Based on an assessment of aerial imagery, the classes best resemble the following:

Class	Description
0	Duricrust plains and plateau remnants supporting hard spinifex grasslands.
1	Plateaus and ridges supporting soft spinifex grasslands.
2	Rugged plateaus, ridges and large hills supporting hard spinifex grasslands.
3	Small hills and ranges with soft spinifex grasslands.
4	Mesas, breakaways and stony plains with spinifex grasslands.
5	Riparian area and moderate to large drainage lines with *Eucalyptus* species.

The class we're most interested in is the *Eucalyptus* species in riparian and drainage areas (i.e., class 5).

7.2.3 Initialise the tool

Click the VegFrax tool from the Tenement Tool top menu tab, like so:

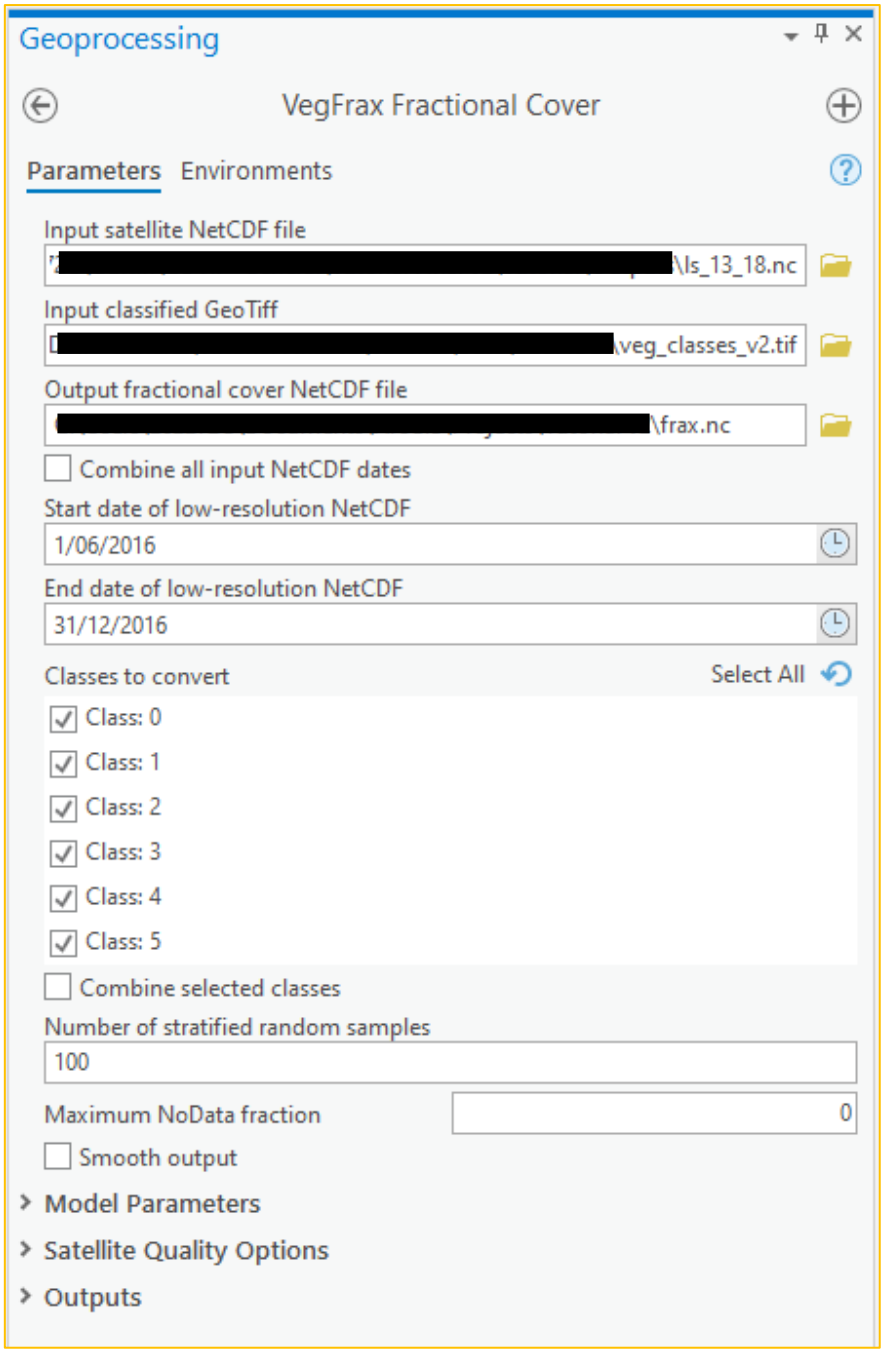


The VegFrax geoprocessing panel will appear on the right side of the screen.

7.2.4 Generate the class fractional covers

The VegFrax tool provides numerous options for generating fractional cover maps. In most cases, the default parameters can be considered optimal, though experimentation is recommended.

Set the parameters as follow:



Set the input satellite NetCDF to the **ls_13_18.nc** file we downloaded earlier and the input classified GeoTiff to the **veg_classes_v2.tif** from the tutorial data.

Save the output NetCDF file as **frax.nc** anywhere on your computer.

Uncheck the **combine all input NetCDF dates** and set the start and end dates to **01/06/2016** and **31/12/2016**, respectively. This will ensure only the drier half of 2016 is used, potentially maximalising persistent vegetation on the satellite imagery.

Ensure all classes **in classes to convert** are selected. This will ensure the tool produces a separate fractional map for each class. You could also just select the relevant classes that you need (e.g., only class 5).

Ensure the **combine selected classes** option is unchecked. If this is checked, then selected classes will be combined in the output. This is helpful for merging similar classes.

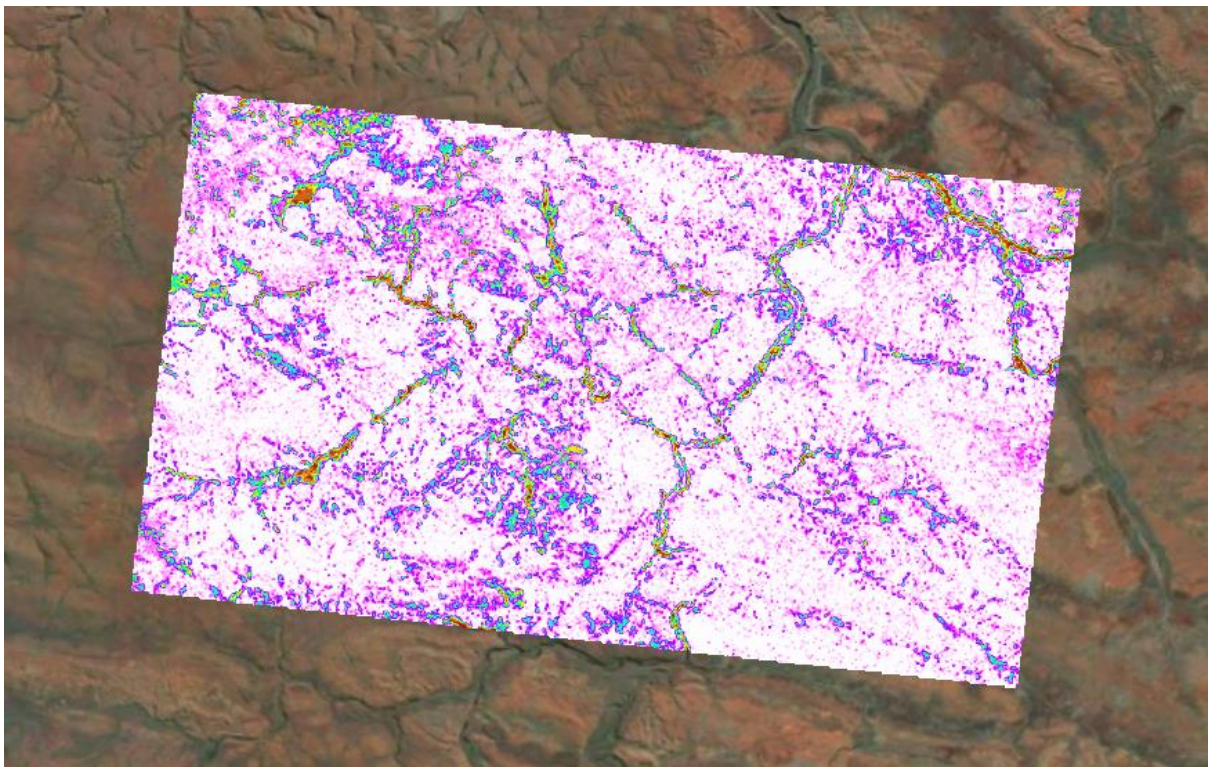
Leaving the **rest as default** is recommended, but feel free to experiment. Increasing the number of model iterations will result in a potentially more accurate output, but will take longer to process.

When ready, click the **Run** button at the bottom of the panel. This process shouldn't take long.

7.2.5 Explore the VegFrax fractional cover maps

Six separate fractional layers should appear on your map in a group layer called **fractions**.

Explore these fractional maps and see if they match with your high-resolution classified GeoTiff. Here's how **class 5** should appear:



The higher the value indicates that a higher fractional of that class was predicted at that location on the lower-resolution satellite data.

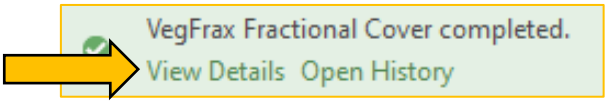
You can see the dark red/brown pixels appear to follow the riparian or drainage lines represented on our classified map.

In contrast, some classes may appear to 'do worse' than others (e.g., less clear, noisier). This is likely due to the regression model having difficulty separating spectrally similar classes. These are classes we'd want to ignore or merge.

7.2.6 Assess model performance

Like Nicher, understanding how the fractional model (i.e., random forest regression) performed when predicting each requested class.

VegFrax also hides this information in the ArcGIS Pro result output window. Once more, **click the View Details link** that appears in the tool panel when the tool has finished processing.



A window will appear. Click the **Messages** tab and some accuracy metrics should be displayed:

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Class Accuracy Metrics
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Class 0 MSE: 0.1 MAE: 0.14 RSME: 0.12.
Class 1 MSE: 0.22 MAE: 0.18 RSME: 0.2.
Class 2 MSE: 0.13 MAE: 0.13 RSME: 0.09.
Class 3 MSE: 0.2 MAE: 0.23 RSME: 0.17.
Class 4 MSE: 0.22 MAE: 0.22 RSME: 0.19.
Class 5 MSE: 0.15 MAE: 0.15 RSME: 0.15.
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

Mean Squared Error (MSE) is a widely used metric for determining general model performance when using random forest regression. This is a helpful metric as it presents the amount of error detected in the model in the units of the data (i.e., a value between 0 and 1 for fractional maps).

How did the classes perform? Most performed reasonably well. Class 5 saw **~15% error**, meaning our fractional values may be out by 0.15.

Your results may vary slightly, mostly due to the random selection of training and testing samples.

8 Ensemble

8.1 Introduction

The Ensemble tool is used to combine two or more spatial ‘evidence’ layers using Dempster-Shafer belief theory (Dempster, 1967; Shafer, 1976).

This approach not only allows our evidence to be combined but also allows for evidence uncertainty and conflicts to be determined and visualised.

Four outputs are produced by the Ensemble tool, including belief, disbelief, plausibility and belief-interval (i.e., uncertainty).

Prior to ensemble modelling, all input layers must be prepared using fuzzy membership functions. We will walk through the process below.

8.2 Prepare evidence layer inputs

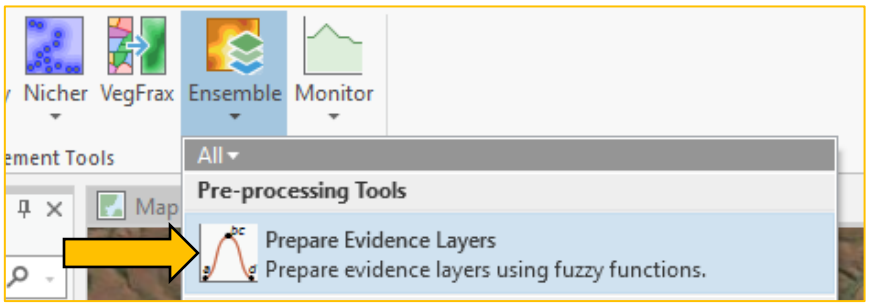
8.2.1 Considerations

- Inputs can be either a NetCDF or GeoTiff.
- Inputs must be projected in the GDA94 Australia Albers (EPSG:3577) projection system.

8.2.2 Initialise the Ensemble Sigmoider tool

Before we generate an ensemble model, all input values have to be rescaled to a new value range 0 – 1. We can do this using sophisticated fuzzy membership sigmoidal functions.

Click the Ensemble Sigmoider tool from the Tenement Tool top menu tab, like so:



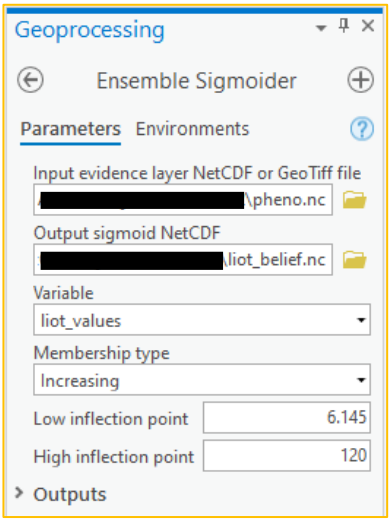
The Ensemble Sigmoider geoprocessing panel will appear on the right side of the screen.

8.2.3 Rescale evidence layers

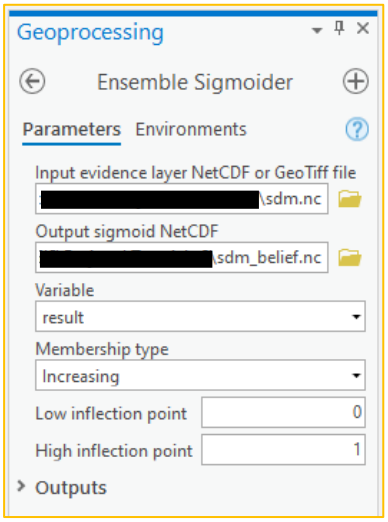
This tool can only take one input at a time, so we need to run it several times in order to prepare all our evidence layers.

Set the parameters for each evidence layer as follows:

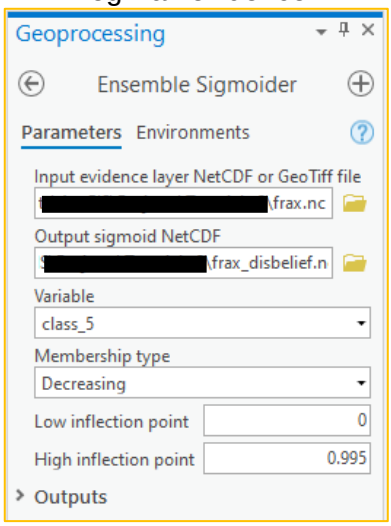
Phenometrics evidence



Nicher evidence



VegFrax evidence



Using the output NetCDFs from the tools you ran prior (Phenolopy, Nicher and VegFrax), convert each into fuzzy ensemble evidence using the settings shown above.

Here, we have assigned the Phenolopy **LIOT metric to an increasing sigmoidal**, which will rescale all values (from lowest to highest) to 0-1. Notice we have reduced the default high inflection point from default down to 120? This is done to **reduce the influence** of very high LIOT values during rescaling. **Do the same** for the **Nicher** output, except leave the low and high inflection as whatever value is set as default.

These two layers of evidence will represent **belief** evidence in the ensemble model, where higher values are indicative of areas in the landscape we highly believe our species to be.

The **VegFrax** evidence is slightly different. Here, we set a **decreasing** membership type. This rescales the values 1-0 instead, where the original fractional high values are rescaled towards 0 and the lower values are rescaled towards 1. This evidence will be used as disbelief in the ensemble model, and higher values will represent areas we highly believe our species **will not exist** (i.e., disbelief).

Explore the outputs of each tool run and get a feeling for what the fuzzy membership functions are doing to the data.

8.3 Generate ensemble model

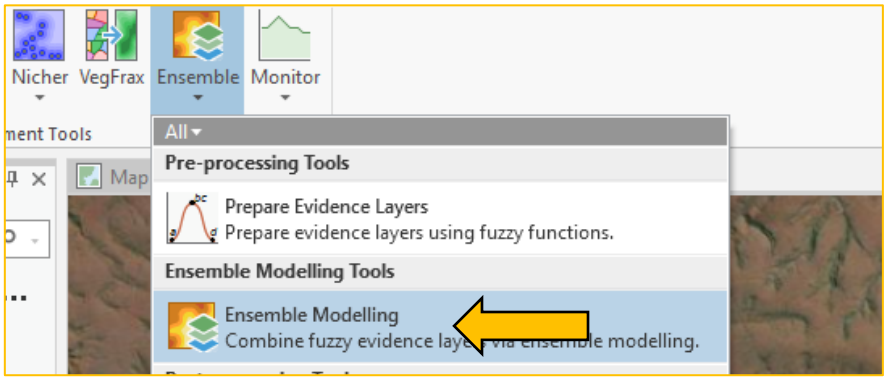
8.3.1 Considerations

- Inputs must have been run through the Ensemble Sigmoider tool prior.
- Inputs must be projected in the GDA94 Australia Albers (EPSG:3577) projection system.

8.3.2 Initialise the Ensemble Model tool

Now that the evidence layers have been prepared, we can finally use them in the ensemble model tool.

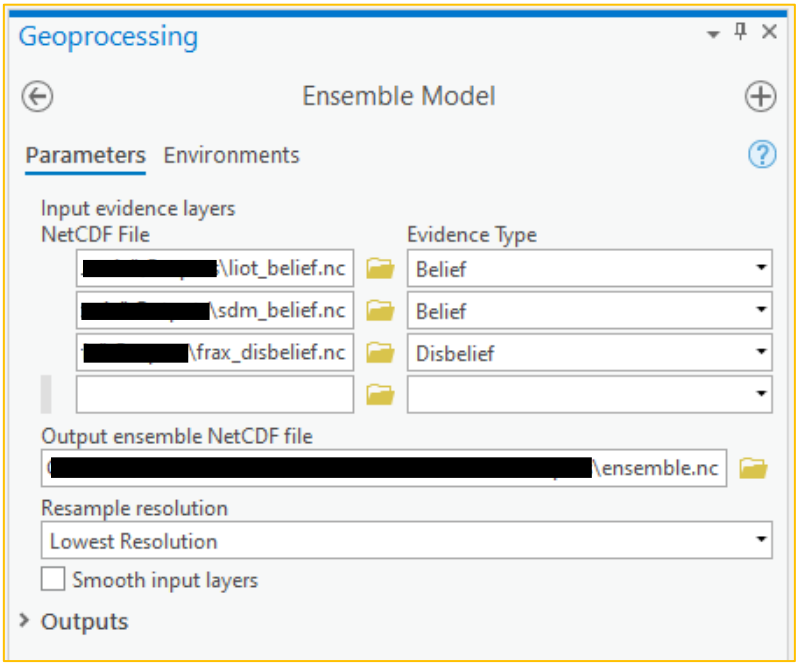
Click the Ensemble Model tool from the Tenement Tool top menu tab, like so:



The Ensemble Model geoprocessing panel will appear on the right side of the screen.

8.3.3 Generate ensemble model

The Ensemble Model tool is quite straightforward to use. Set the parameters are follows:



Ensure the **liot_belief.nc** and **sdm_belief.nc** files are set the evidence type to **belief**. Set the **frax_disbelief.nc** as **disbelief** evidence type. We need at least one of both.

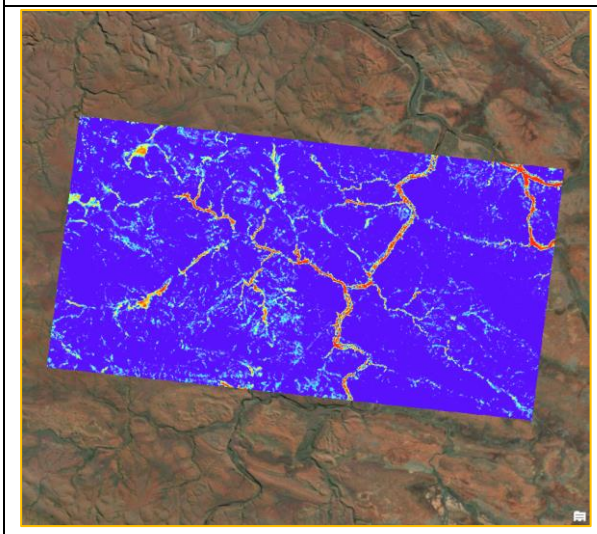
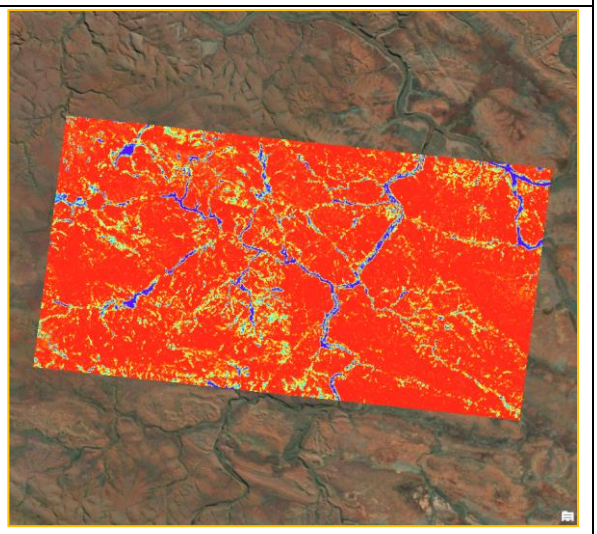
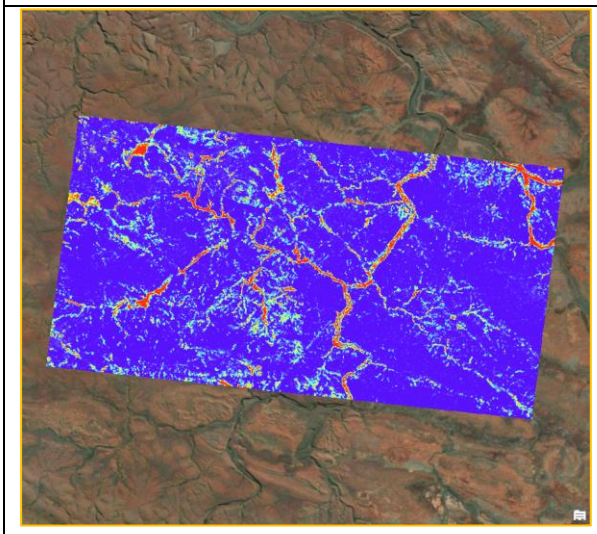
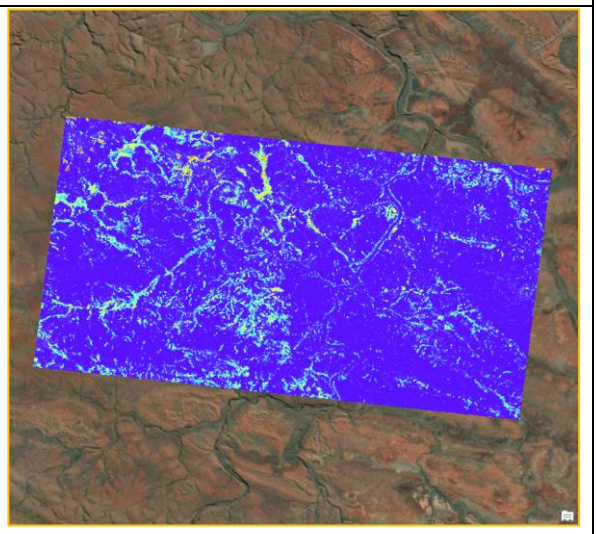
Save the output NetCDF file as **ensemble.nc** anywhere on your computer.

Leave all the remaining parameters as **default**.

When ready, click the Run button at the bottom of the panel. This process should be very quick.

8.3.4 Explore the ensemble outputs

Four separate ensemble model output layers should appear on your map in a group layer called **'ensemble'**. These layers are called **belief**, **disbelief**, **plausibility** and **interval**:

Belief	Disbelief
	
Plausibility	**Interval (i.e., uncertainty)**
	

The **belief** layer represents all combined belief evidence (i.e., LIOT and the SDM). This layer represents areas in the landscape where we **strongly believe our target species exists**. The higher the value is indicative of stronger belief.

The **disbelief** layer is similar as belief except **inverted**; this represents areas where we **strongly disbelieve our target species exists**. In other words, higher values are indicative of areas we do not believe the species exists.

The **plausibility** layer is interesting, it represents the degree to which we **cannot reject** the hypothesis that our target species exists. The higher the value, the more plausible the chance that our species exists.

Finally, the **interval** layer (or belief-interval) is the difference between belief and plausibility. It is often referred to as a **confidence** layer as it shows areas where evidence conflicted and uncertainty exists. The higher the value, the more uncertainty. This is a useful layer for determining if and where more evidence is required. Do you have any areas that might need to be investigated further?

9 NRT Monitoring

9.1 Introduction

The NRT Monitor toolset is designed to provide you with tools to create monitoring areas around any vegetation (GDV or not) in Australia and track on-going changes (e.g., growth, stability or decline) in near real-time from now and into the future.

This is achieved using a modern change detection algorithm known as EWMACD (Brooks et al., 2013) and Edyn (Brooks et al., 2017). Essentially, the algorithm generates a harmonic regression model from satellite-derived vegetation values beginning at a specified year and tracks how far vegetation at later dates has deviated from the initial harmonic model.

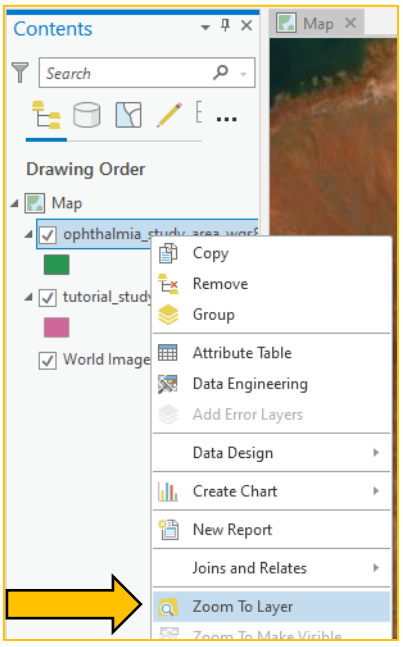
This approach has been implemented into the NRT Monitor toolset, along with the ability to visualise change, send alerts via email when specific changes occur and continuously monitor without the need for user input.

We will walk through all major functions of the NRT Monitor toolset (as well as some simpler change detection methods) in this section.

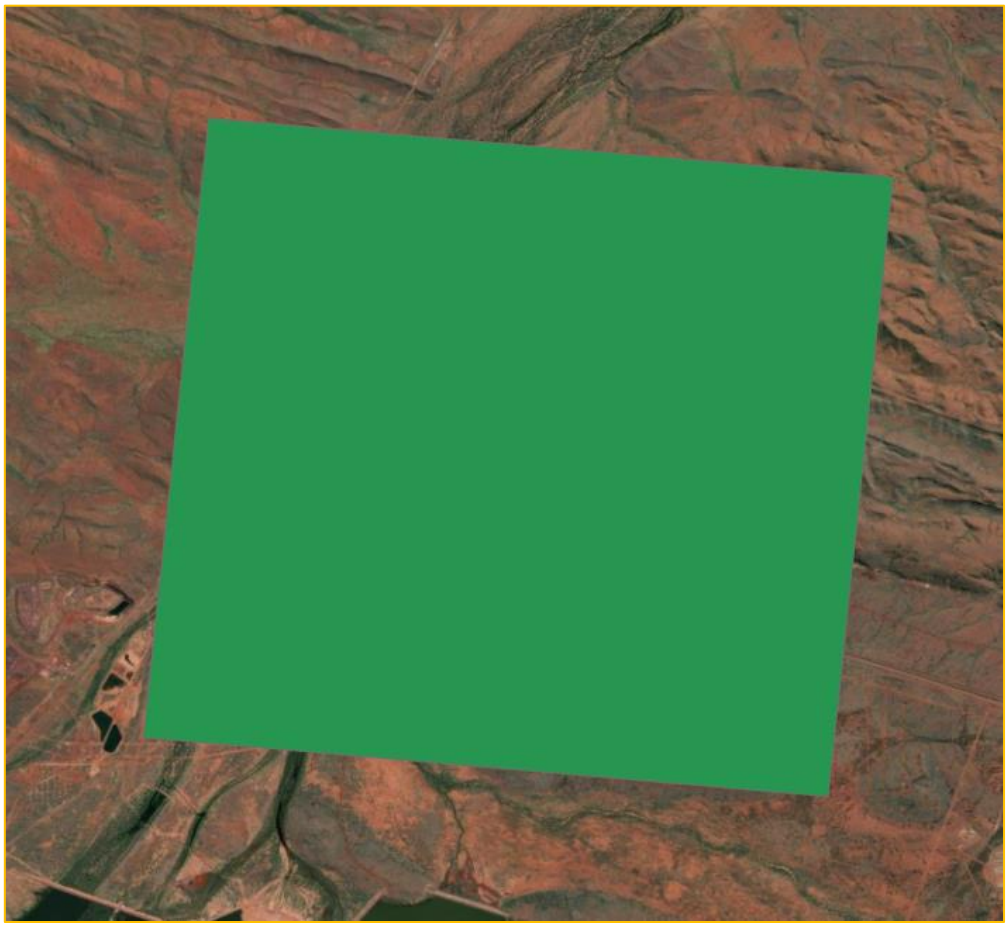
9.2 Switch to new study area

If you recall, we added two study area shapefiles at the beginning of the tutorial. Let's use the second area for our monitoring exercises.

Right click the layer called **ophthalmia_study_area_wgs84** in the contents pane and click Zoom to Layer:



The map view will zoom to the selected area and should look similar to this (likely with different colouring):



Spend some time changing the **symbology** of the study area feature (e.g., remove the fill colour and set the border to black). Likewise, have a look around the study area and the underlying aerial imagery to get a sense of the area.

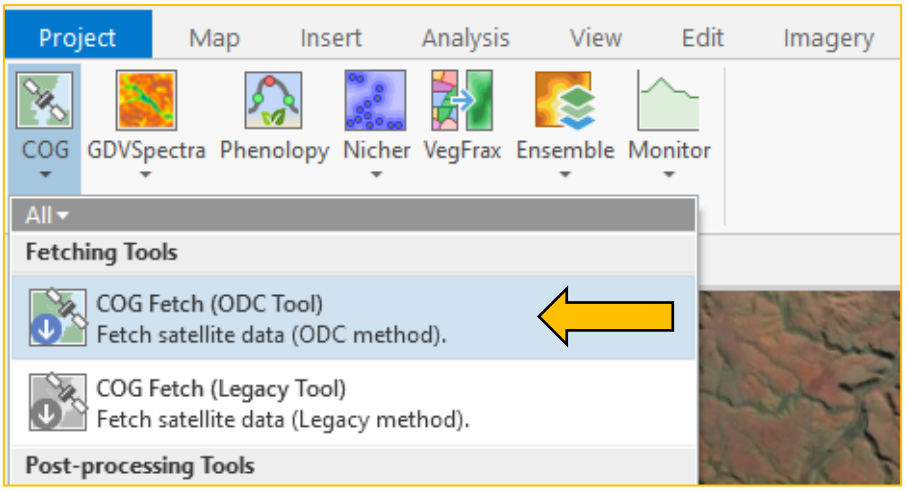
This area is just north of the Ophthalmia Dam near the town of Newman in the Pilbara, Western Australia. Notable impacts (e.g., drainage line obstruction) have been observed to the GDV north of the dam at various the dates over the last several decades. We will investigate some of these impacts.

9.3 Preliminary vegetation change assessment

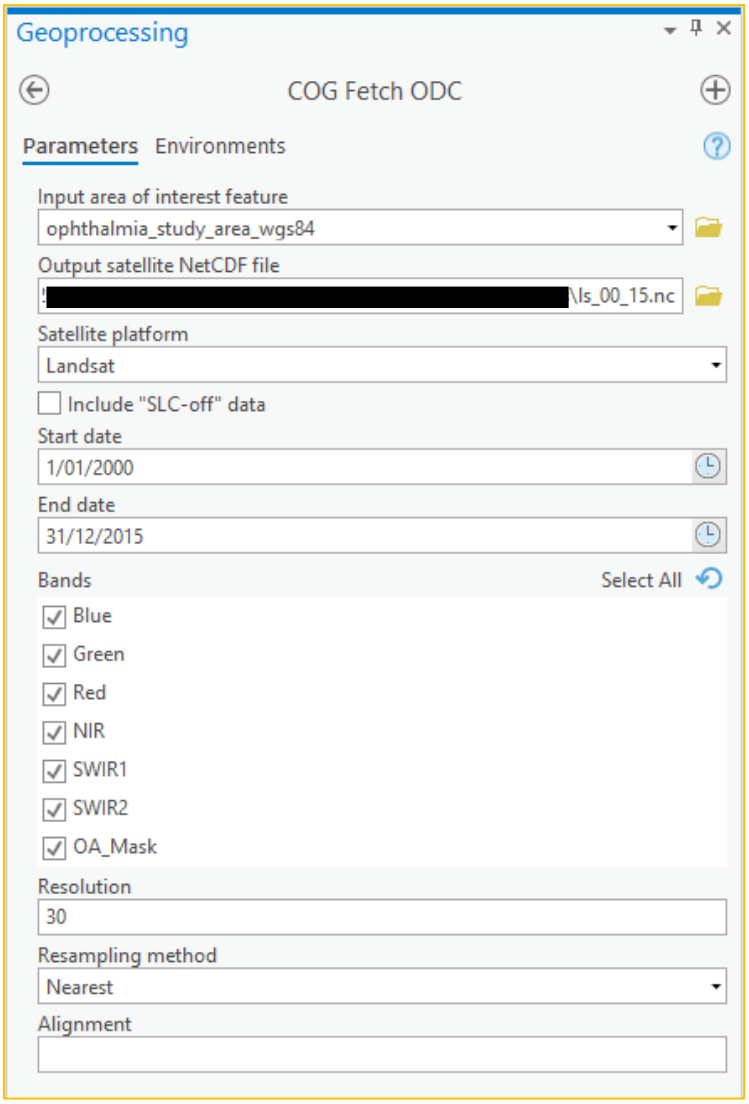
9.3.1 Initialise and run the COG Fetch tool

Similar to before, let's download some Landsat satellite data for our new study area. We will download all available satellite data from **2000 to 2015** and use it for a preliminary investigation of the study area's vegetation dynamics over time.

Click the COG Fetch (ODC) tool from the Tenement Tools top menu tab, like so:



The COG Fetch (ODC) geoprocessing panel will appear on the right side of the screen. In the geoprocessing, set the parameters as follows:



Select the **ophthalmia_study_area_wgs84** layer as the input area of interest this time. Save the output NetCDF file anywhere on your computer and call it **ls_00_15.nc**.

Set satellite platform as **Landsat** for the period **01/01/2000** to **31/12/2015**.

Leave the remaining parameters as **default**.

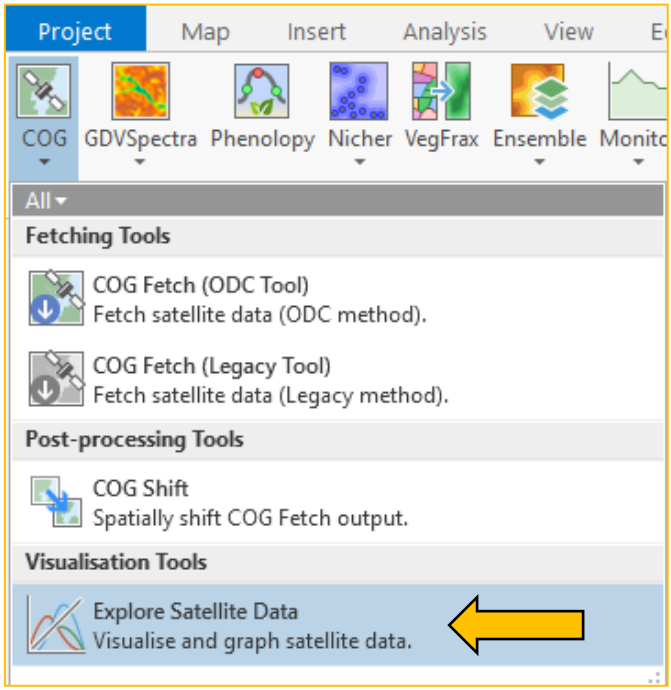
When ready, click the **Run** button at the bottom of the panel. The download should take approximately 5 - 10 minutes on a decent network.

When this process is complete, you should see a single **.nc** file approxmately 400mb in size at the output location set above.

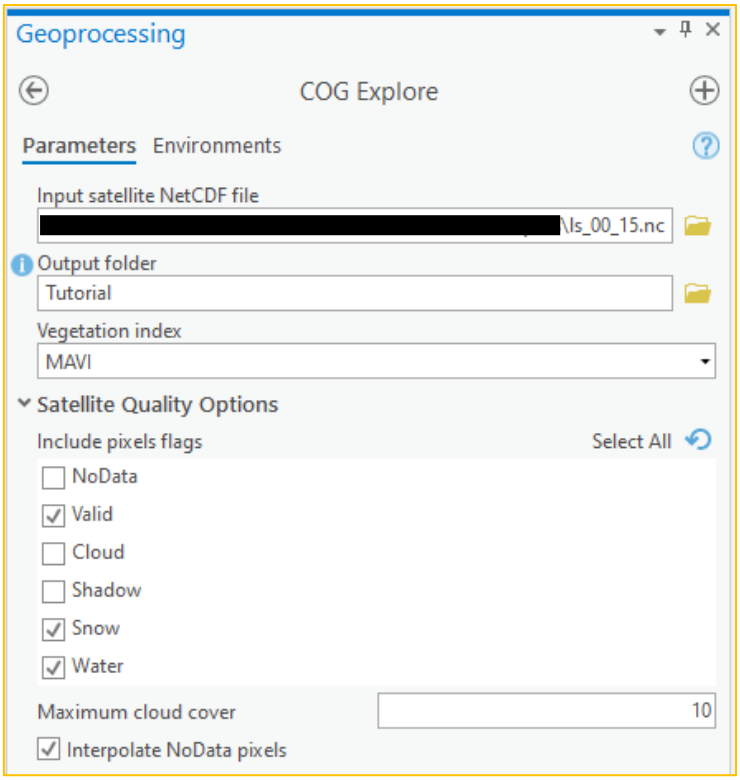
9.3.2 Initialise and run the COG Explore tool

As before, visualise the raw satellite data as the vegetation index known as **MAVI**.

Click the COG Explore tool from the Tenement Tools top menu tab, like so:



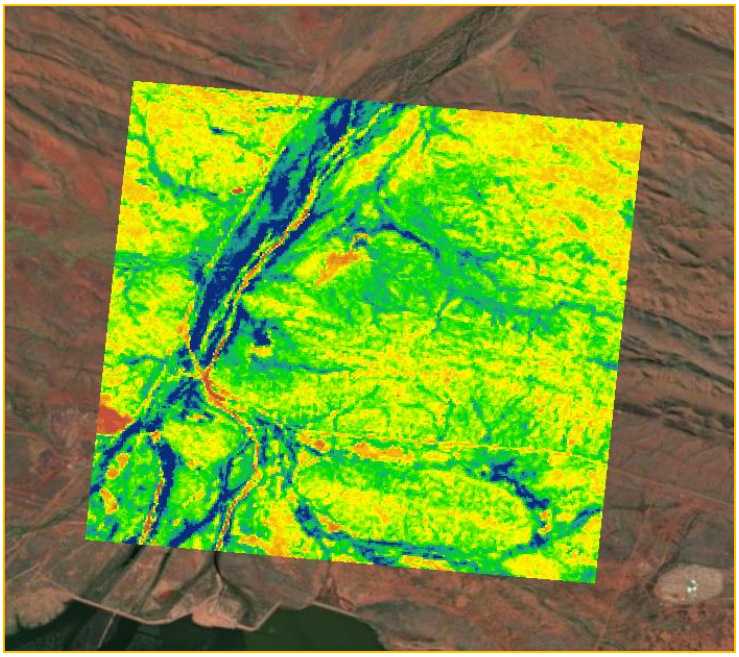
The COG Explore geoprocessing panel will appear on the right side of the screen. In the geoprocessing panel, set the parameters as follows:



Set the input NetCDF as the one we just downloaded, **ls_10_15.nc**. Set the output folder to any appropriate **folder** on your computer for the outputs of the tool (a **.crf** file).

Finally, set the vegetation index to **MAVI** (though feel free to use any). Leave the remaining parameters as **default**.

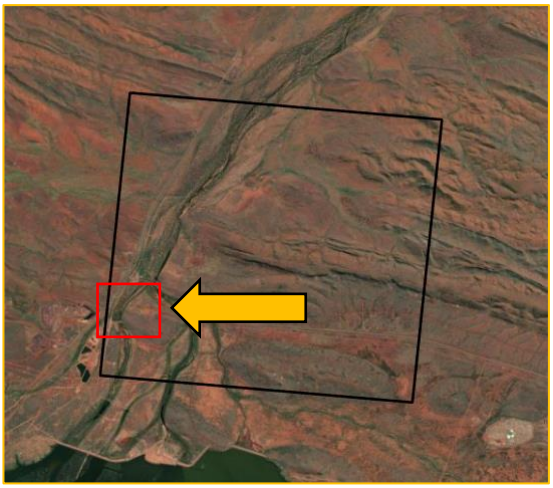
When ready, click the **Run** button at the bottom of the panel. When finished, the satellite data represented as MAVI will automatically appear on your map. It should look similar to below:



9.3.3 Zoom to impact area

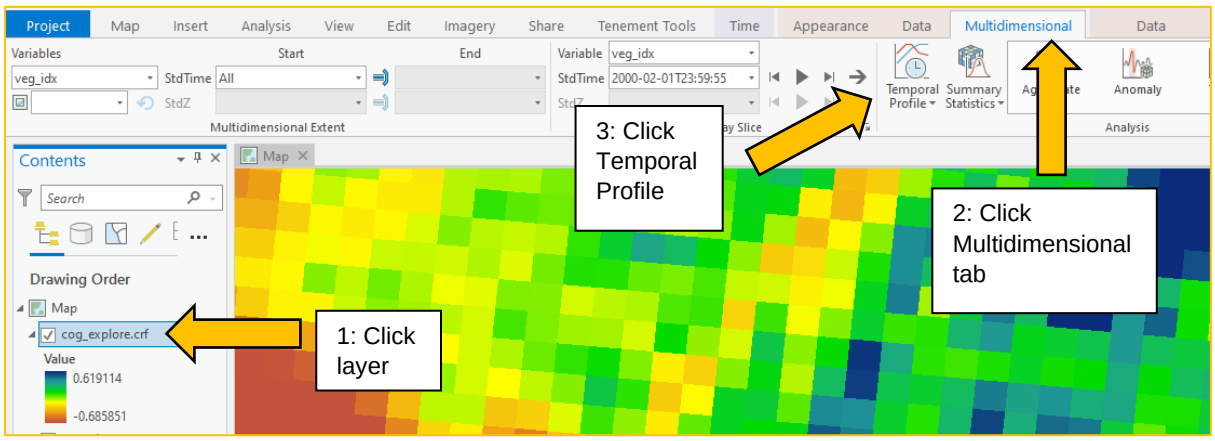
We will explore some vegetation containing *Eucalyptus victrix* that was impacted between **2007 and 2008** as a result of drainage line obstruction due to installation of dam infrastructure.

Zoom in towards the general area shown below:



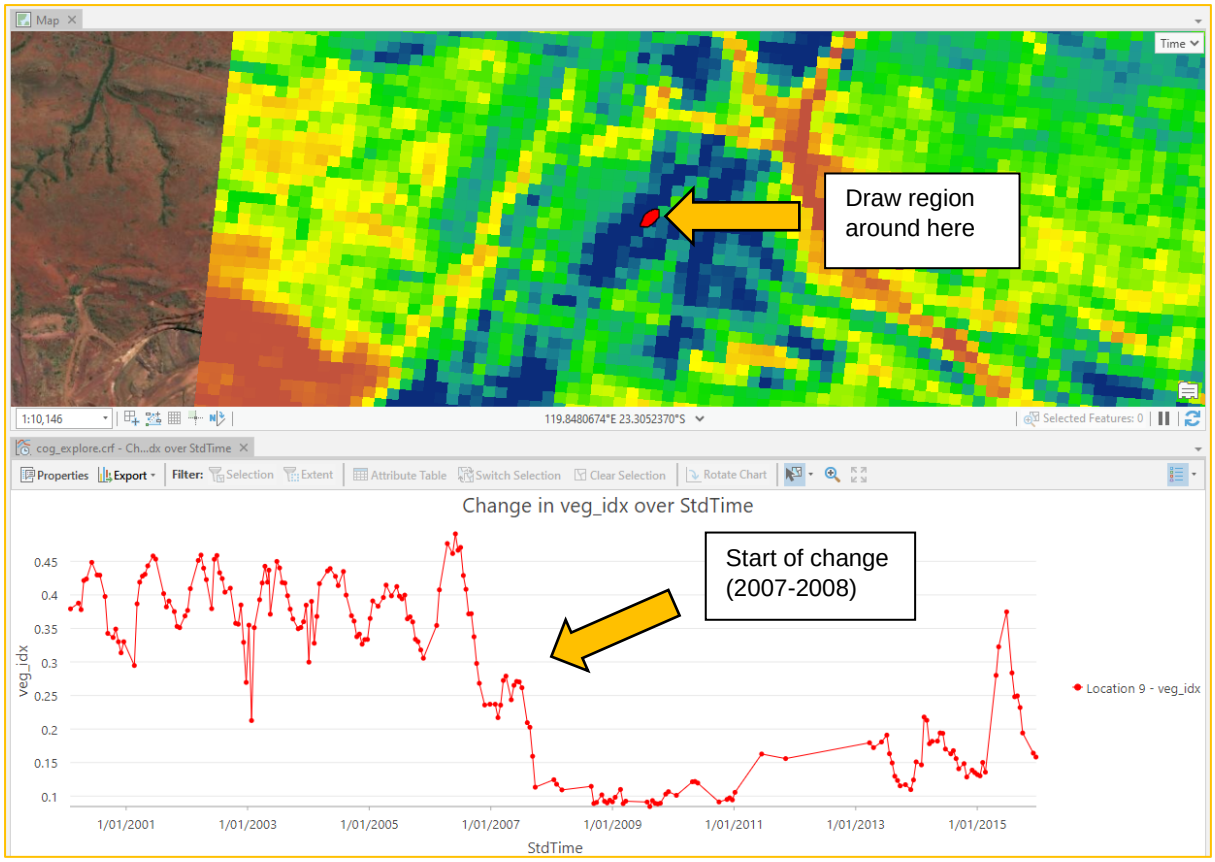
9.3.4 Graph the vegetation history

Let's graph the vegetation's history from 2000 to 2015. Click the **cog_explore.crf** layer in the left side contents panel (1), then the **Multidimensional** tab (2), and finally click the **Temporal Profile** button (3):



A **new graphing panel** should appear at the bottom half of your screen, along with some controls in a panel on the right.

As before, **draw a region of interest** around the historical impact area, like so:



We can see that around **2007 and 2008** the impact to the vegetation is observable from the temporal profiles.

Spend some time exploring other vegetation before moving on. Were there any other notable changes to vegetation in the area? Any significant vegetation growth areas?

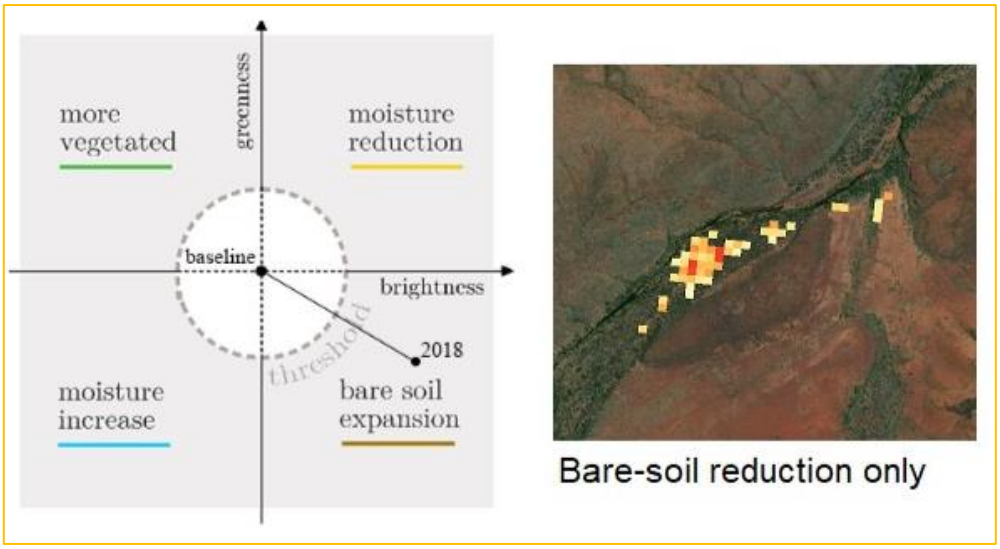
9.4 Detect change via Change Vector Analysis (CVA)

9.4.1 Introduction

The CVA tool allows us to compare a range of images across multiple years captured at similar dates (e.g., dry season) against a specified baseline image. Every image in the range is compared to the baseline.

The CVA algorithm produces an angle and magnitude map for every image. The angle map can be classified into one of four change types (e.g., moisture and vegetation increase or decrease), depending on angle value (see above figure). The magnitude indicates how intense the change was.

Our implementation produces four change outputs per year, based on the below figure:



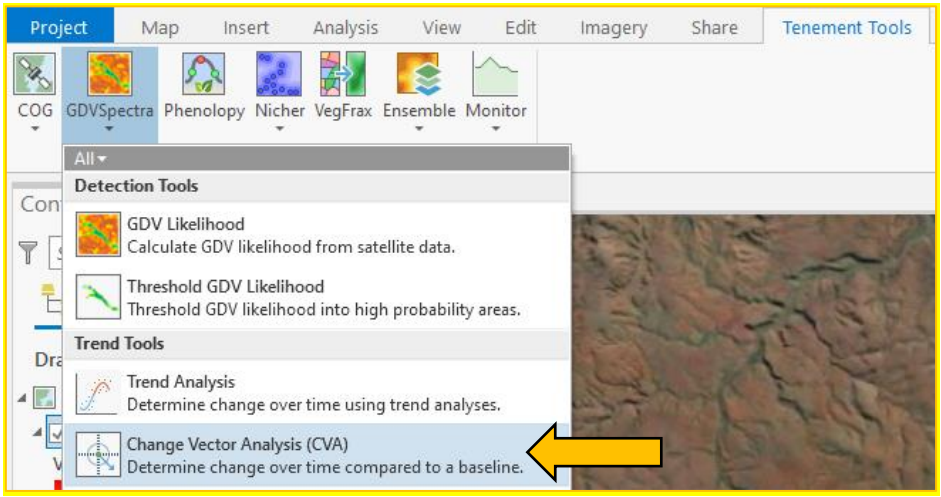
We will now jump back to the GDVSpectra CVA tool to do a preliminary change assessment. The Change Vector Analysis (CVA) tool is useful for quick analysis of change in a study area.

9.4.2 Considerations

- The input must be a NetCDF file of raw satellite data downloaded prior using the COG Fetch (ODC) or COG Fetch (Legacy) tools.
- The optional mask input NetCDF file must be the threshold NetCDF output from the GDVSpectra Threshold tool.
- At least two years of data is required within the raw satellite NetCDF file.

9.4.3 Initialise the tool

Click the GDVSpectra CVA tool from the Tenement Tools top menu tab, like so:

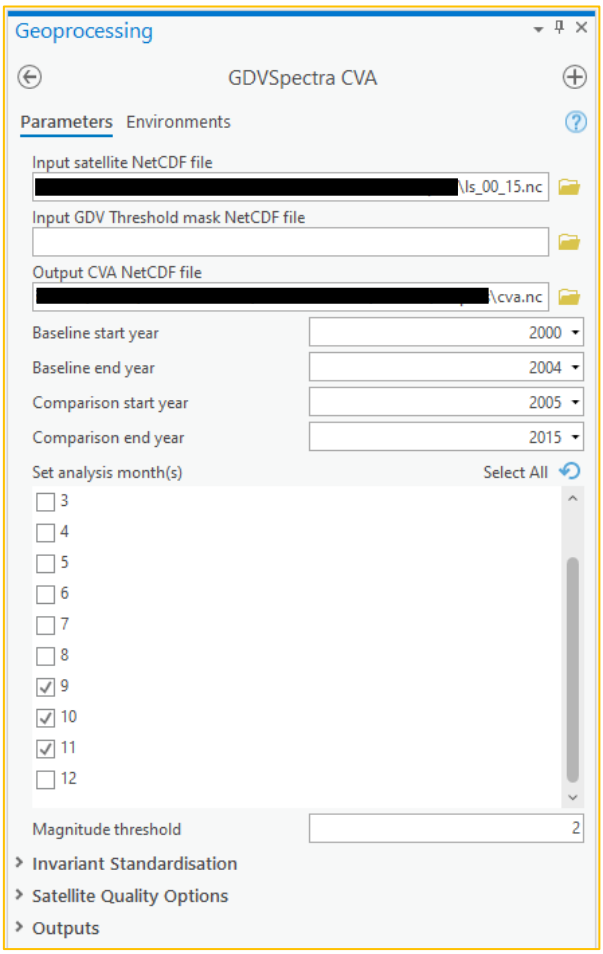


The GDVSpectra CVA geoprocessing panel will appear on the right side of the screen.

9.4.4 Perform Change Vector Analysis

Let's perform a CVA using the previously downloaded Landsat data for the study area.

Set the parameters as follows:



Set the input satellite NetCDF to the Landsat data NetCDF **ls_00_15.nc** we have just downloaded. Set the output NetCDF file anywhere on your computer and call it **cva.nc**.

We now need to set the baseline year(s) and the comparison year(s). Set the **baseline range to start at 2000 and end at 2004**. This will create a baseline of combined imagery that exists for the period 2000 to 2004, reducing seasonal fluctuation.

Set the **comparison start and end year from 2005 to 2015**. Each year here will be compared to the baseline aggregate.

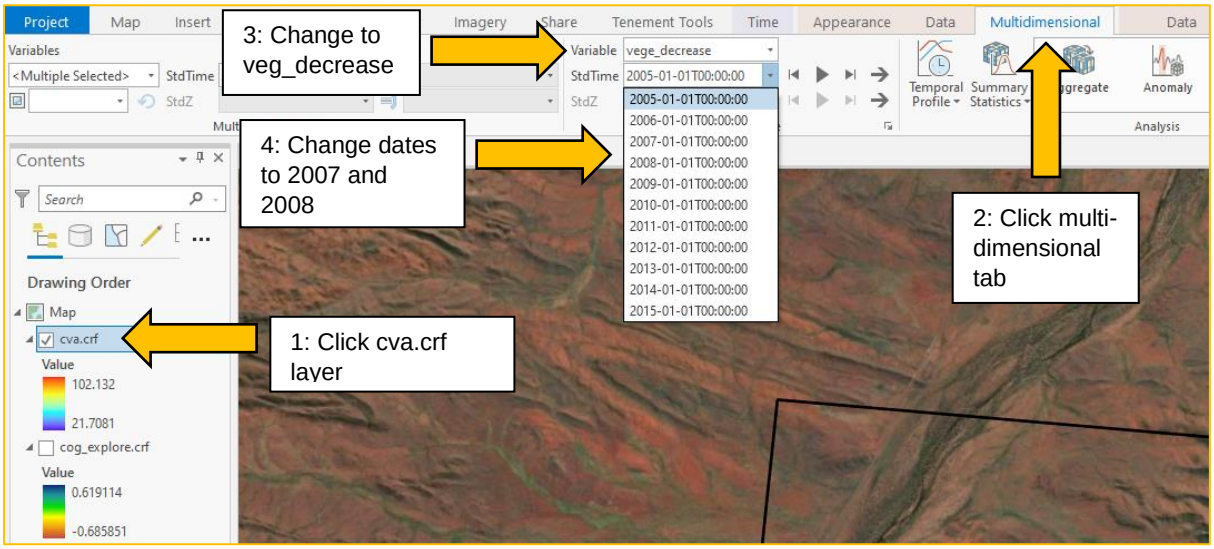
Leave the remaining parameters as **default**. This will use the **dry period months (months 9, 10, 11)** for comparisons.

When ready, click the **Run** button at the bottom of the panel. The download should take approximately 5 - 10 minutes on a decent network.

9.4.5 Visualise vegetation decline change

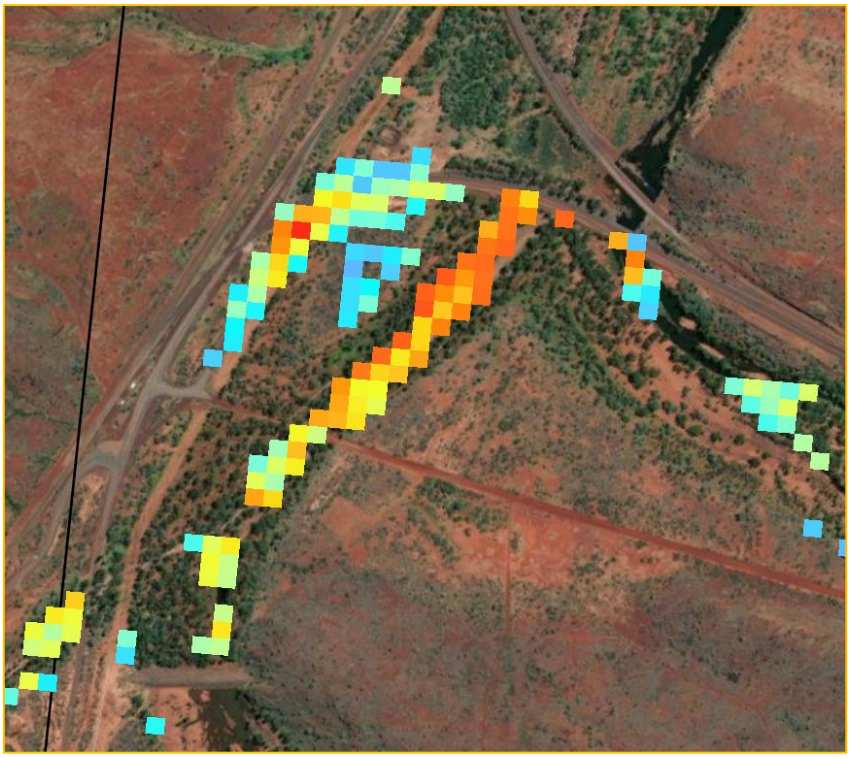
The output initially shows areas of moisture decline for the first comparison year (2005). Let's see how vegetation decline looks instead.

Click on the **cva.crf** layer in the contents panel (1) to enable to **Multidimensional tab** (2). Change the Variable to **veg_decrease** (3). Browse different images at different dates using the **StdTime drop-down** control (4):



Take your time scrolling through some of the different dates to get a better idea of how the vegetation changes throughout the years. How does the change detection look when in 2007 or 2008?

It should show some high magnitude change in 2008, like so:



Here, **'hotter' pixels** are areas that saw significant vegetation decrease in **2008** when compared to the baseline of vegetation data based on an aggregate of 2000 to 2004. This can be helpful in determining areas to start monitoring.

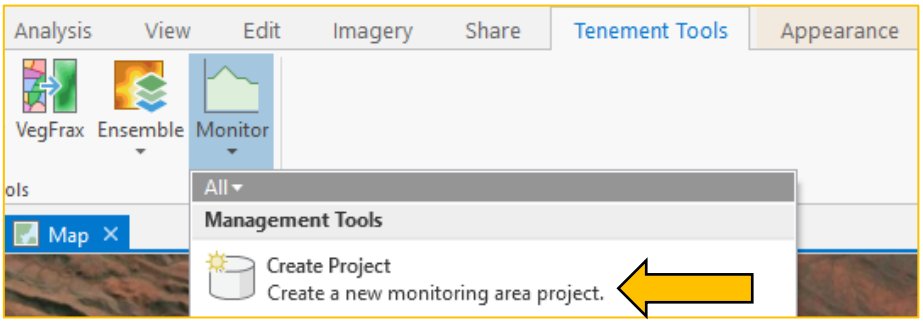
9.5 Create a new monitoring project

9.5.1 Considerations

- Monitoring areas cannot be created without a properly created project via the NRT Create Projects tool.

9.5.2 Initialise the tool

Click the Create Project tool from the Tenement Tools top menu, like so:

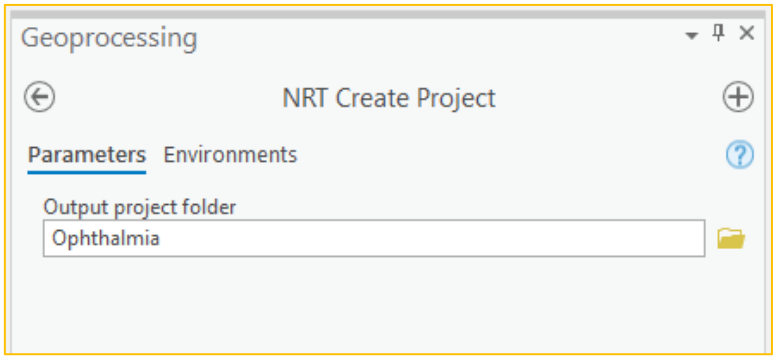


The NRT Create Projects geoprocessing panel will appear on the right side of the screen.

9.5.3 Create a new project

A project must be setup before monitoring areas can be created. This project contains a specific folder structure and geodatabase required by all tools within the NRT Monitoring toolset. Without a project, all NRT Monitoring tools will fail to run.

Set the parameters as follows:



Simply **create a new folder called Ophthalmia** somewhere on your computer and select it in the tool using the little **yellow folder icon** above (tip: you can create new folders from within the prompt that appears).

When a folder has been selected, click the **Run** button at the bottom of the panel. When finished, an empty layer called **monitoring_areas** will be added to the contents panel on the left, automatically.

This process designates the selected folder to be the project folder for our current monitoring work. Have a look inside the folder on your computer. You should see a geodatabase called **monitoring_areas.gdb** within.

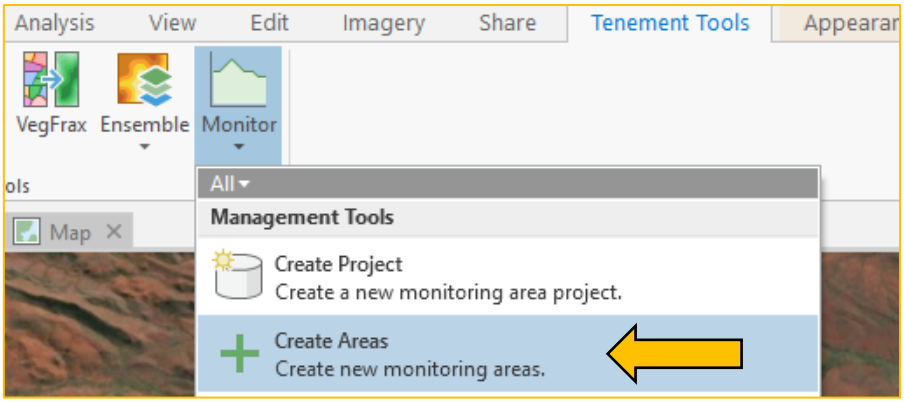
9.6 Create new monitoring areas

9.6.1 Considerations

- A monitoring area must not be smaller than 30m^2 (i.e., a single Landsat pixel).
- Only monitoring area layers created using the Create Project tool are accepted.

9.6.2 Initialise the tool

Click the Create Areas tool from the Tenement Tools top menu, like so:

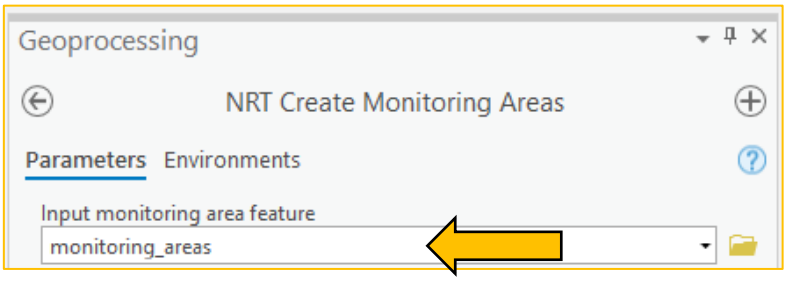


The NRT Create Areas geoprocessing panel will appear on the right side of the screen.

9.6.3 Create monitoring area

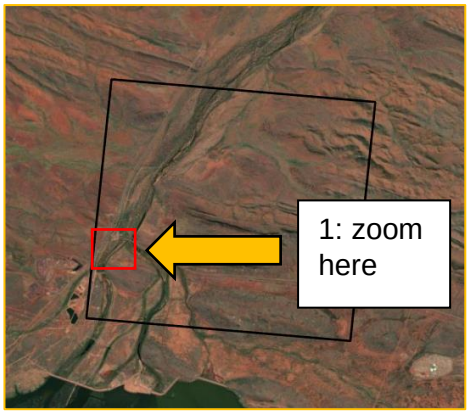
The Create Monitoring Area tool is used to draw monitoring area boundaries and assign change detection parameters.

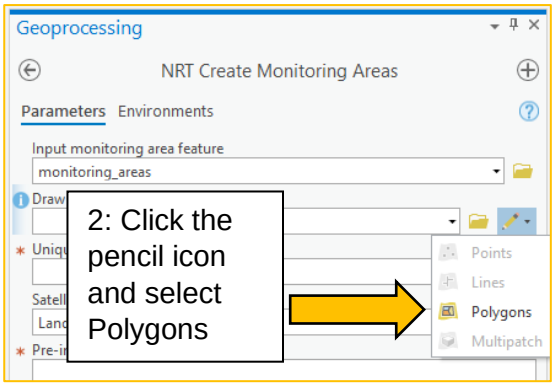
In the **NRT Create Areas** tool, first select the **monitoring_areas** layer as the input monitoring areas feature, like so:

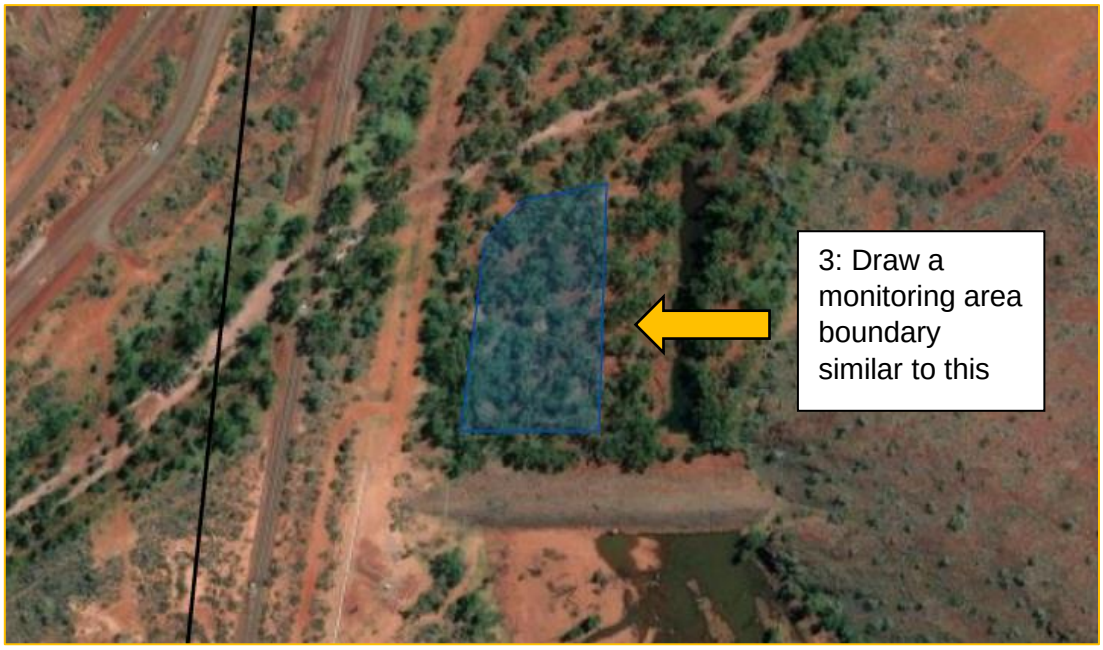


Once selected, some **new parameters** should appear in the geoprocessing panel.

Now, **zoom to the area** (1) we explored using CVA earlier (if not already there). Click the pencil icon and click **Polygons** (2). A new temporary layer will be added to the table of contents. Finally, draw **a new monitoring** area boundary (3):

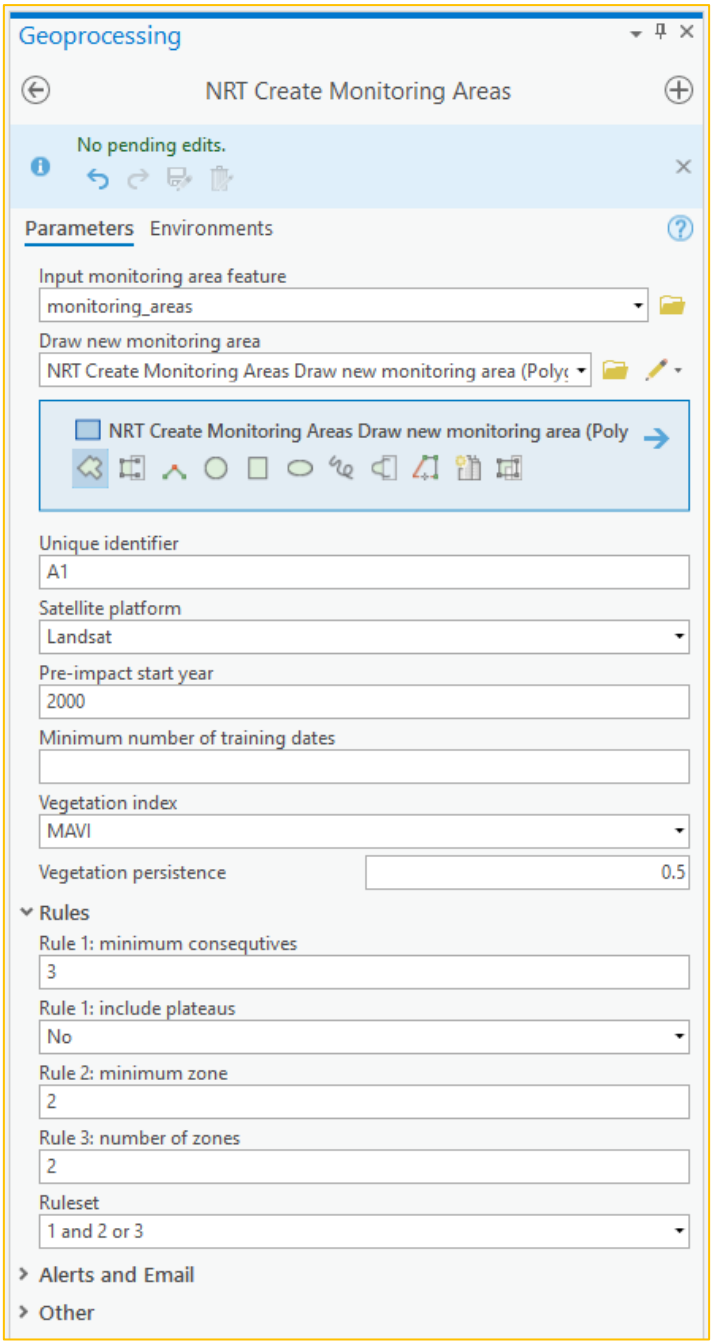






Once the monitoring area has been drawn, the next step is to set various parameters that control the change detection and alert systems used by NRT Monitoring tools. We will set the **general parameters** first and explain others, below.

Ensure your geoprocessing panel looks like this:



Set the **unique identifier to A1**. This is used to help us identify the monitoring area easier. Every monitoring area must have a unique identifier, and anything can be used.

Set the **satellite platform to Landsat**. Significant impact occurred at this monitoring area in 2007, so we need to use Landsat here because Sentinel is only available from 2015 onwards. For newer areas, you'd likely want to use Sentinel.

The parameter **pre-impact start year** is very important. Set it to the year **2000**. This is the year that will be used as the starting point to train the change detection algorithm. Typically, a stable, healthy period of vegetation of **one or more years** is suitable. Setting this year to a period of impacted, unstable vegetation will result in inaccurate monitoring.

Leave **the minimum number of training dates to empty**. This will force to choose an optimal number of satellite images to use for training length. Override this if you would like to force this **minimum training period** to be longer or shorter. As a general rule of thumb, **15 is one year of Landsat, and 40 is one year of Sentinel**.

We will leave the remaining parameters as **default**, though a summary of other important parameters is provided here:

Parameter	Description	Example
Vegetation persistence	Controls the 'sensitivity' of how quick the change algorithm will register deviation. A lower value will register deviations sooner but will capture more insignificant deviations (i.e., noise). A higher value will result in the inverse.	N/A
Rule 1	Controls the number of consecutive inclines (growth) or declines (impact) dates required before an alert is registered. Include plateaus will include 'no change' dates following inclines or declines (see appendix).	Appendix 16.2
Rule 2	Set the minimum zone a deviation must occur within before it will be registered as an alert. This is used to exclude change of low significance.	Appendix 16.3
Rule 3	Controls the number of zones needed to register a large "spike" deviation as an alert. A lower value will register relatively small "leaps" between dates, whereas a higher value will only register very significant "spikes".	Appendix 16.4
Ruleset	Rules can be used in isolation or combination to create unique alert triggers. See appendix for an explanation. Note: the blue appendix rows are most useful.	Appendix 16.5
Alert based on method	The change detection algorithm used provides two modes: static and dynamic. The static method will compare all new vegetation data to the training period model forever. The dynamic method will recalibrate itself and create new training periods when the vegetation has deviated far enough away from the original period.	Appendix 16.6
Direction of change to trigger alert	Controls the direction and type of change that will register a deviation. The best explanation is provided visually within the appendix.	Appendix 16.7

When ready, click the **Run** button at the bottom of the panel. When finished, the tool will create a new monitoring area in the project database and visualise it, like so:

You can create as many monitoring areas as you need for any area in Australia. You can also modify and/or delete your monitoring areas at any time using the NRT Modify Area and NRT Delete Area tools.

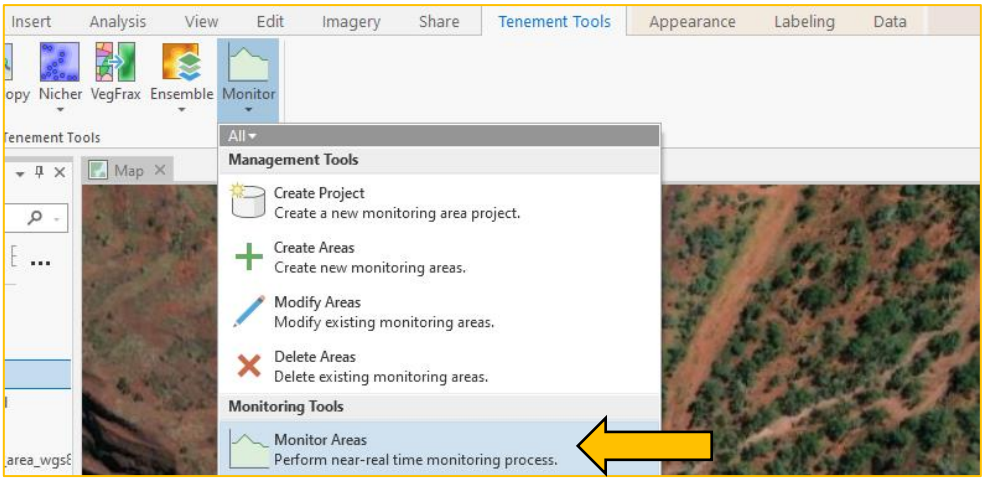
9.7 Begin monitoring areas

9.7.1 Considerations:

- The tool requires a monitoring_areas feature created prior by the NRT Create Project tool and must contain at least one monitoring area.

9.7.2 Initialise the tool

Click the NRT Monitor Areas tool from the Tenement Tools top menu, like so:

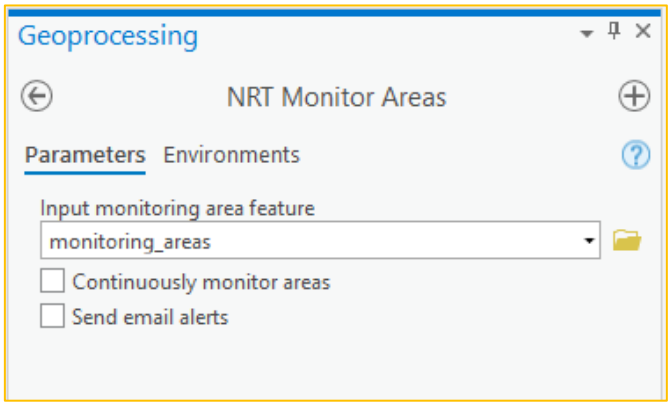


The NRT Monitor Areas geoprocessing panel will appear on the right side of the screen.

9.7.3 Begin monitoring areas

The NRT Monitor Areas tool can run in two different modes: 'on-demand' and 'continuous'. In this tutorial, we will only demonstrate the former. Likewise, the tool can send emails automatically when alerts are triggered – this feature is also not demonstrated here.

In the NRT Monitor Areas tool panel, set the parameters like so:



Set the input monitoring area feature to the **monitoring_areas** feature containing our single monitoring area we just created. The tool will get the latest satellite data and update change model for every area in this feature each time we run the tool.

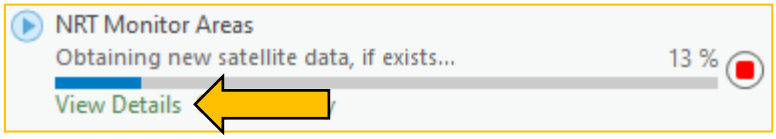
Leave the **continuously monitor areas** parameter **unchecked**. This will force the tool to only run through each monitoring area once, and will then exit. This is 'on-demand' mode.

Also leave the **send email alerts** parameter **unchecked**. Use this to automatically send emails whenever a monitoring area has triggered an alert.

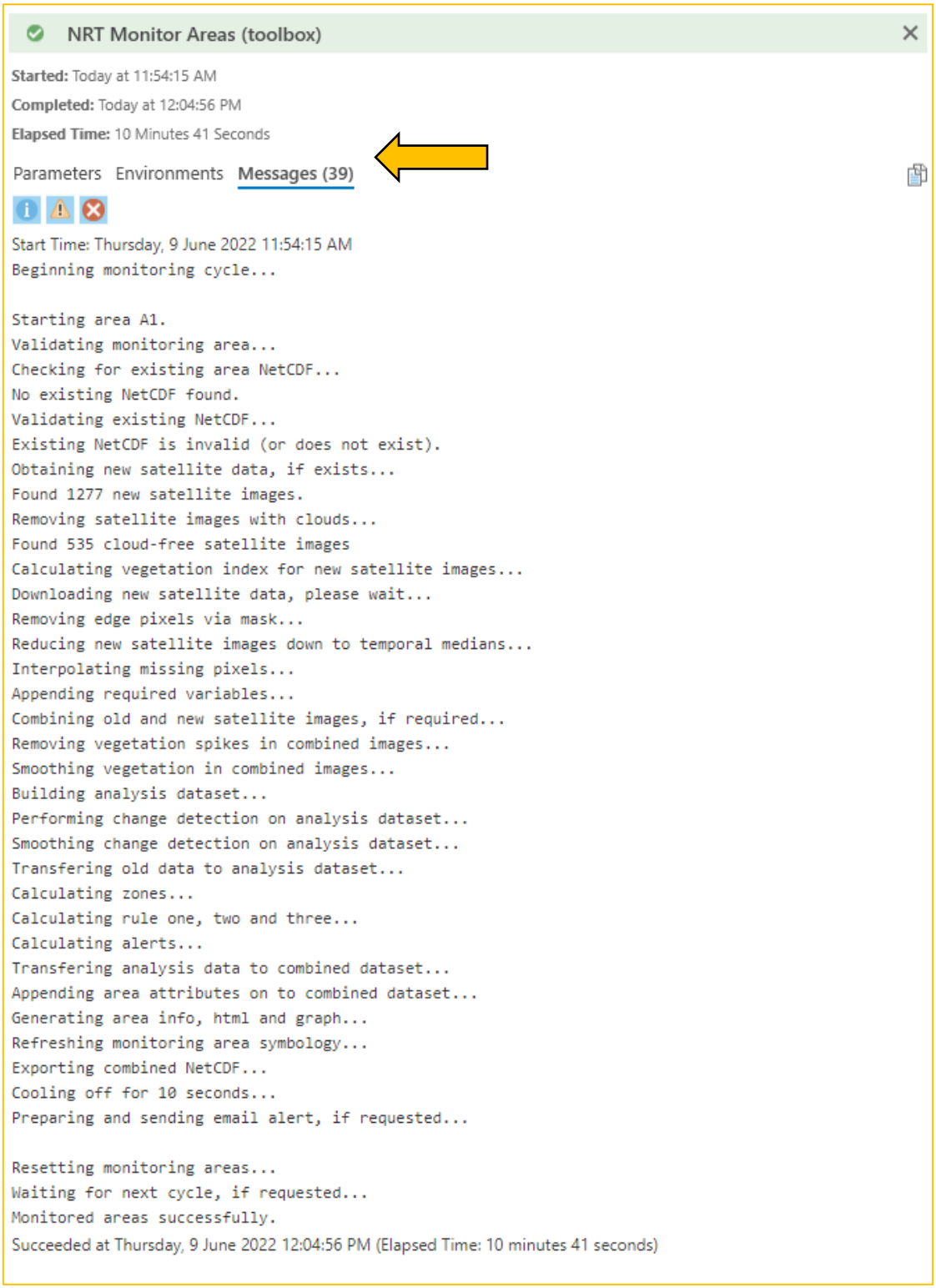
When ready, click the **Run** button at the bottom of the panel. The process needs to download satellite data, so it may take a few minutes to complete.

9.7.4 Observe the monitoring tool process

The NRT Monitor Area tool process can be observed while it's working by clicking on the **View Details link** as the tool processes:



A prompt will appear. Click on the **Messages tab** to see a real-time status update showing what the tool is doing for each monitoring area. This is particularly helpful when running in **continuous** mode:



Once the tool has finished running, the existing symbology of the monitoring area(s) will be updated to match the latest zone and change direction for the latest satellite imagery.

Here's how the area appears as of **June 9th, 2022**:

The **fill colour** (purple) represents the **current zone** the area is detected to be in, and the **label** provides the **zone number and direction** (minus for negative zone).

The **border colour** (black) indicates the **current direction of change** the monitoring area is experiencing. Currently, the **black border indicates 'no direction'**.

If the border were **red**, the area would be in **decline** and if **blue**, the area would be in **incline (growth)**.

This visual information is supposed to provide a **quick idea of how a monitoring area is performing** just at a glance. We can assess change much deeper using the NRT Graph Area tool.

9.8 Graph monitoring area

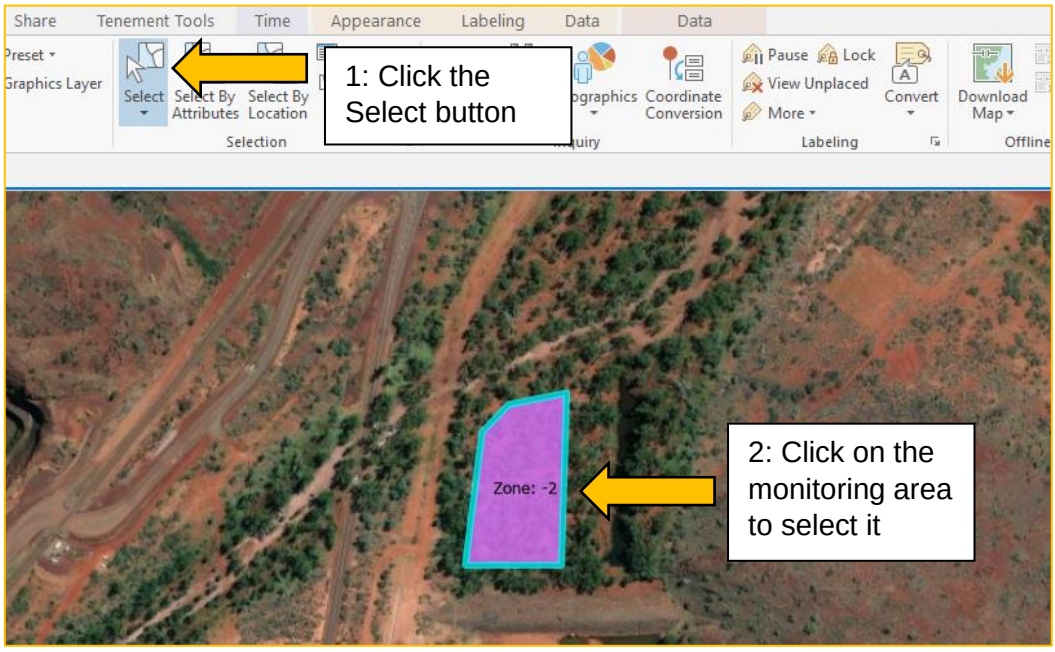
9.8.1 Considerations

- The graphing tool can only be used on monitoring areas created using the NRT Create Monitoring Areas tool.
- A monitoring must have been monitored at least once in order for the graph tool to work.

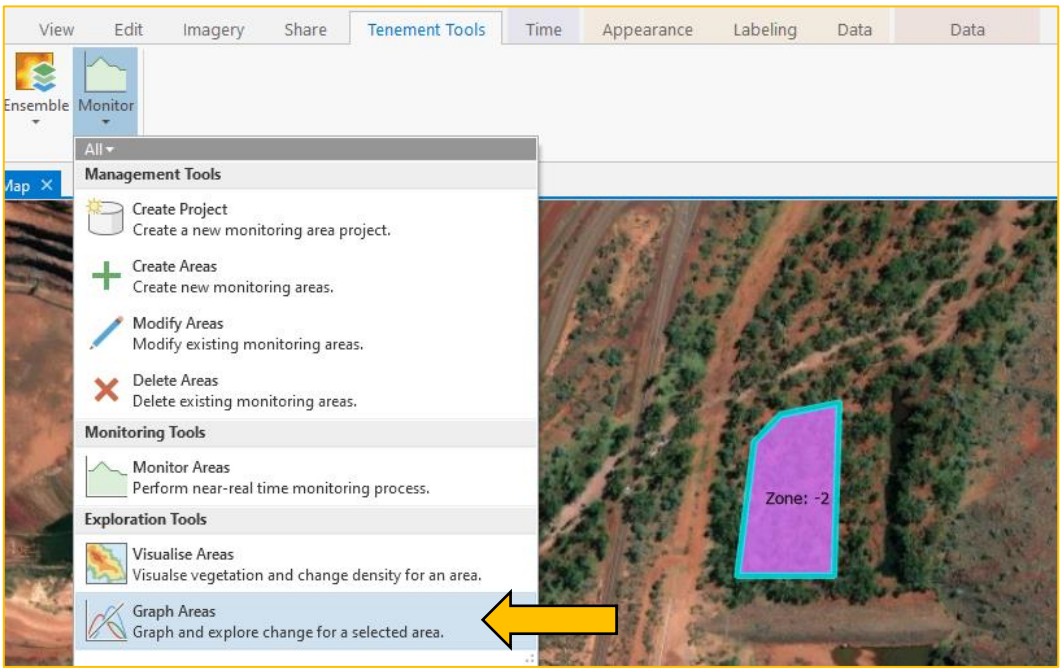
9.8.2 Select the monitoring area

The NRT Graph Monitor Area tool works a little differently to the others. It can only be run on a single monitoring area at a time, and that monitoring area must be currently selected on the map.

First, enable the **Select** tool (1) found in the **Map tab** on the top menu, and then click your monitoring area to **select it** (2). The **border will change colour** when selected:

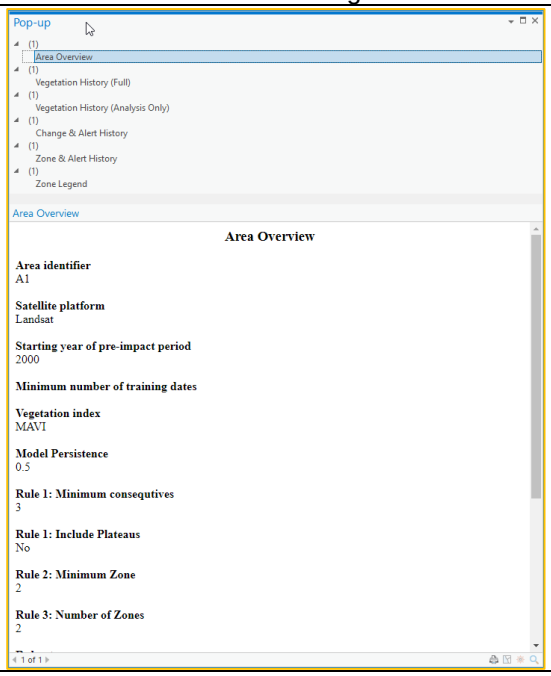
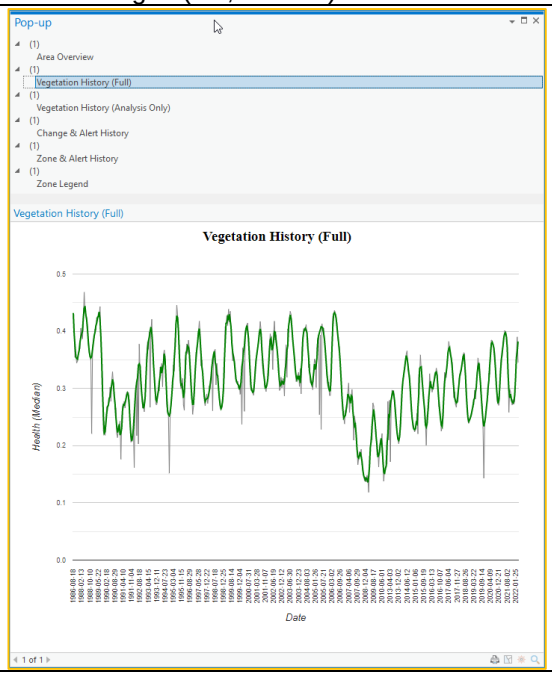


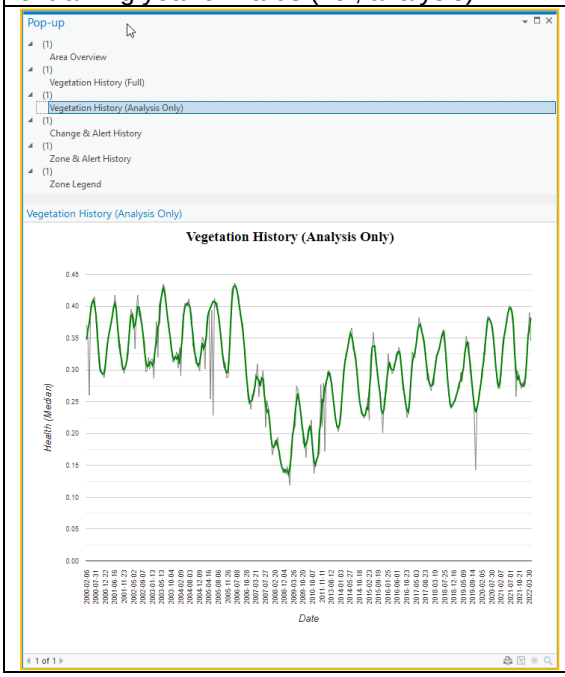
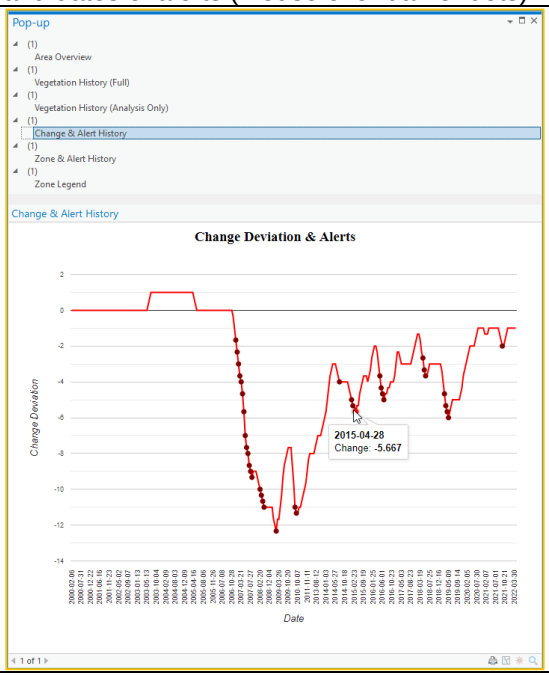
With the monitoring area selected, go back to the **Tenement Tools tab on the top menu** and run the **NRT Graph Areas** tool:

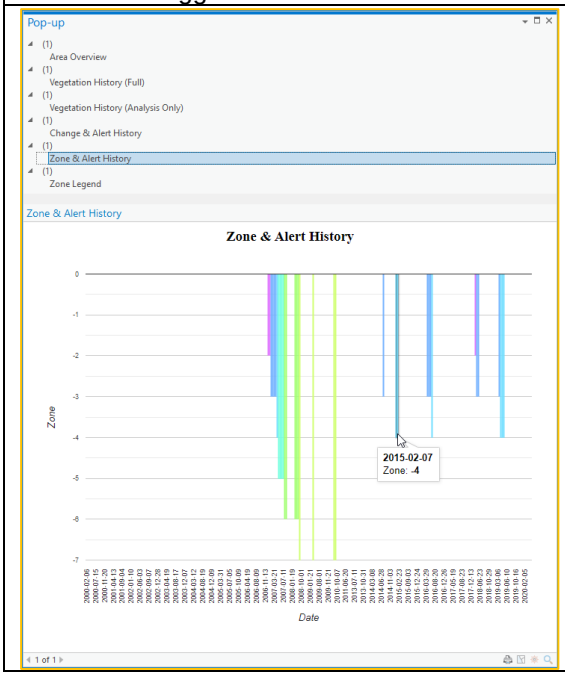
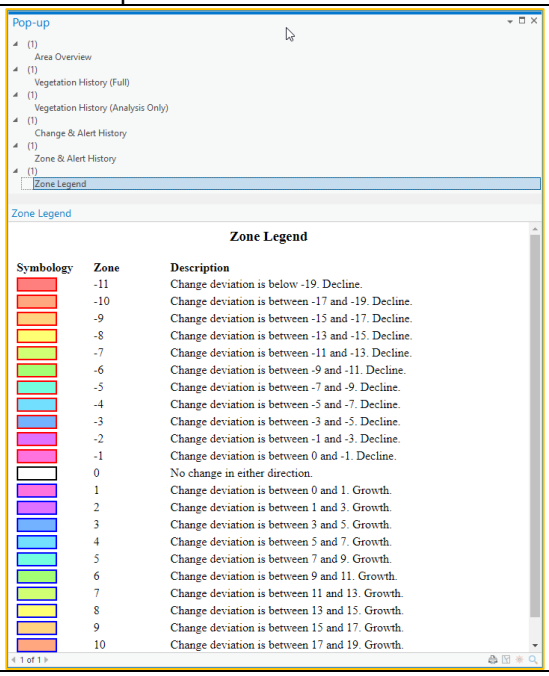


9.8.3 Explore monitor area graphs

A **pop-up window** will appear on your screen showing information regarding the selected monitoring area. Click on the headings to change the information. A summary of each is provided below:

Area Overview	Vegetation History (Full)
Shows parameters set during monitoring area creation for selected monitoring area.	Shows vegetation index values for all available satellite images (i.e., all-time).
	

Vegetation History (Analysis Only)	Change and Alert History
Shows vegetation index values used from start of training year onwards (i.e., analysis)	Shows change deviation calculated by algorithm and dates of alerts (mouse over darker dots).
	

Zone and Alert History	Zone Legend
Shows alert dates and zone colours of when alerts were triggered.	Basic legend outlining what zones and zone colours represent.
	

Of these, perhaps the most important are the **vegetation and change deviation history graphs**. The vegetation graphs show you the raw vegetation index values used to generate the change deviation information.

Based on our selected start of pre-impact training year, the **change model used at least one year of data from 2000 onwards to calibrate the change model**. Getting this starting year correct is vital for an optimal change model. We chose a year that looks like its relatively stable and healthy.

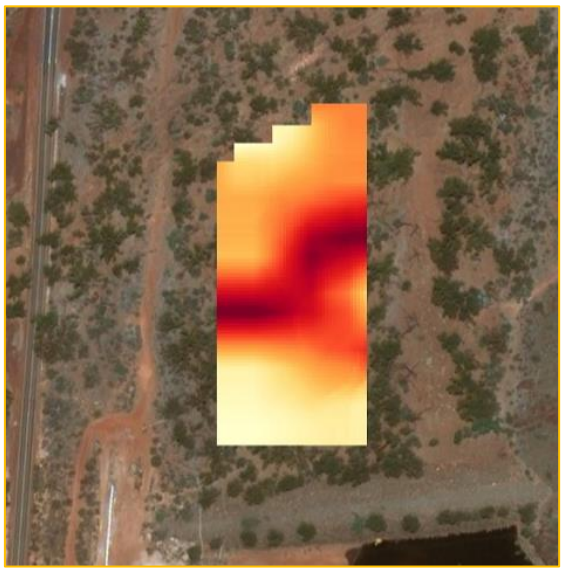
At some point around the **end of 2016**, we can see vegetation begins to change to a declining trajectory. The change deviation graph noticed this and picked up this change relatively early, hence the emergence of negative deviations.

On-going and gradual regrowth over the following several years also appears on the change deviation graph, as **deviation values begin to move back up towards 0 (i.e., no change)**. This indicates the vegetation is beginning to grow back to match the pre-impact period, at least based on the raw vegetation data.

9.9 Visualise monitoring area change

9.9.1 Introduction

The NRT Visualise Areas tool generates a spatial representation of change within a monitoring area. This is useful for indicating where most change has (and hasn't) occurred. This is achieved using a kernel density method:



The various density layers output by this tool are summarised below:

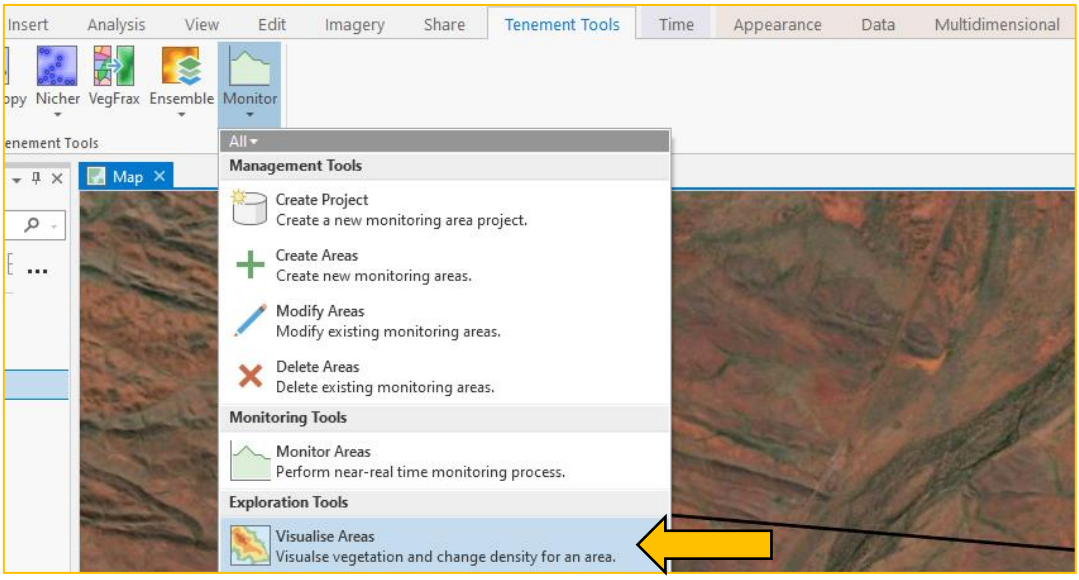
vege_avg_all_time.tif	Mean of vegetation index from start of training period up until today.
vege_std_all_time.tif	Standard deviation of vegetation index from start of training period to today.
vege_latest_time.tif	Vegetation index of most recent satellite image.
change_max_all_time_incline.tif	Maximum vegetation growth (incline) value detected by the change algorithm from start of training period to today.
change_max_all_time_decline.tif	Maximum decline value detected by the change algorithm from start of training period to today.
change_avg_all_time_incline.tif	Mean vegetation growth (incline) value detected by the change algorithm from start of training period to today.
change_avg_all_time_decline.tif	Mean vegetation decline value detected by the change algorithm from start of training period to today.
change_latest_time_incline.tif	Latest vegetation growth (incline) value detected by the change algorithm.
change_latest_time_decline.tif	Latest vegetation decline value detected by the change algorithm.
alerts_cnt_all_time_incline.tif	Count of all vegetation growth (incline) alerts detected by the change algorithm.
alerts_cnt_all_time_decline.tif	Count of all vegetation decline alerts detected by the change algorithm.

9.9.2 Considerations

- Only monitoring area layers created using the Create Project tool are accepted.

9.9.3 Initialise the tool

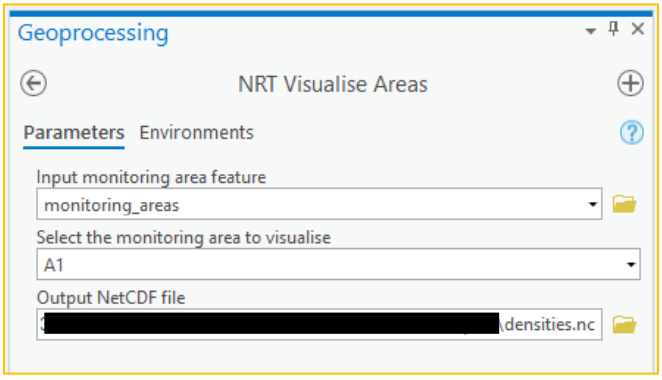
Click the NRT Visualise Areas tool from the Tenement Tools top menu tab, like so:



The NRT Visualise Areas geoprocessing panel will appear on the right side of the screen.

9.9.4 Visualise all-time change

We'll now generate density layers for various vegetation and change dynamics over time for our monitoring area. Set the parameters as follows:



Set the input monitoring area feature to the **monitoring_areas** layer we created earlier. Select the monitoring area to visualise as **A1**, the area we drew before.

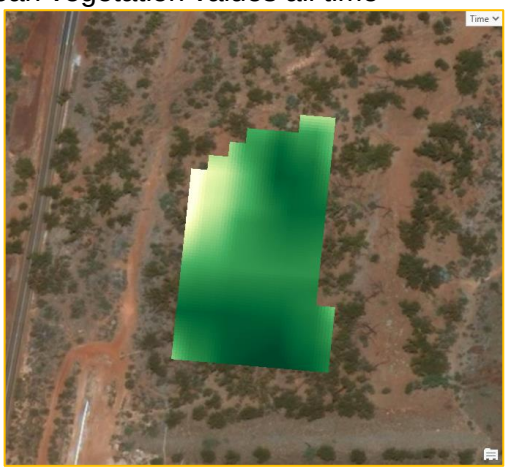
Finally, set the output NetCDF to **densities.nc** and save it anywhere on your computer.

When ready, click the **Run** button at the bottom of the panel. The process needs to download satellite data, so it may take a few minutes to complete.

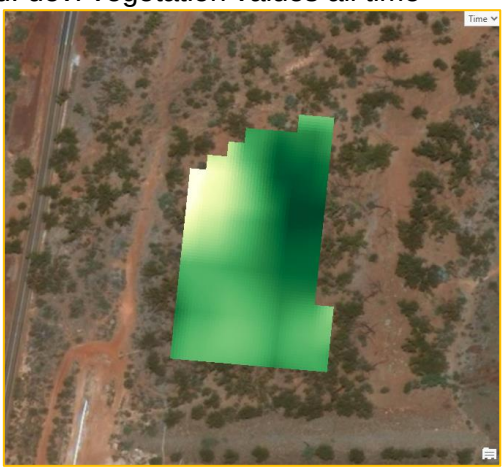
9.9.5 Visualise vegetation and change densities

When the tool is finished processing, around a dozen layers will be added to a new group layer called **nrt_visual** on the left contents panel. The following summarises these outputs:

vege_avg_all_time.tif
Mean vegetation values all time

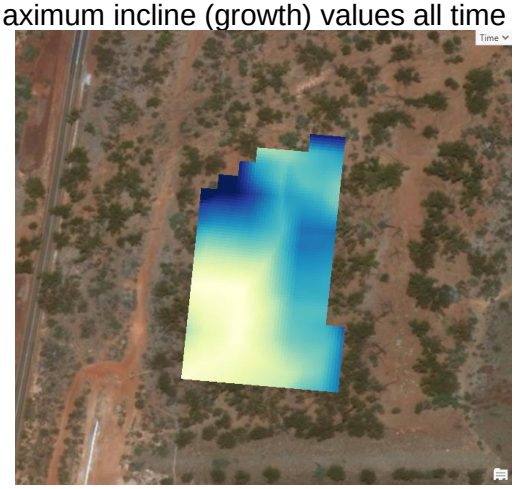


vege_std_all_time.tif
Std. dev. vegetation values all time

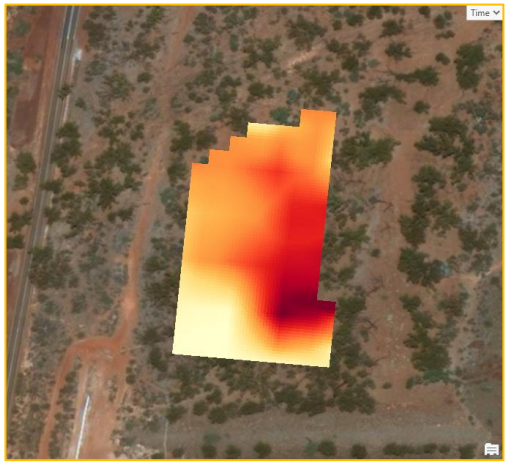


vege_latest_time.tif
Latest vegetation values

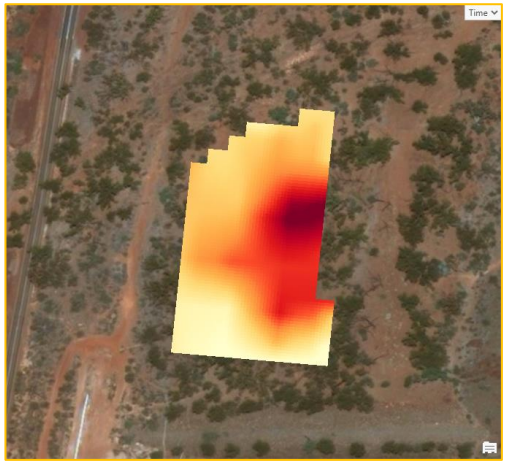
change_max_all_time_incline.tif
Maximum incline (growth) values all time



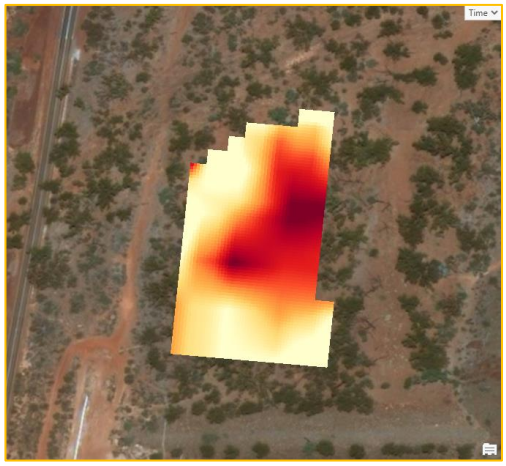
change_max_all_time_decline.tif
Maximum decline values all time



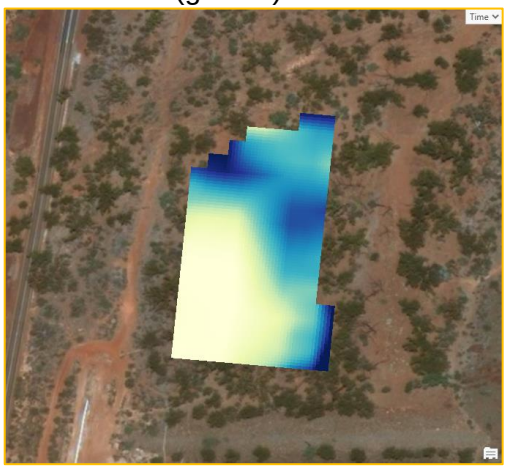
change_avg_all_time_decline.tif
Mean decline values all time



change_latest_time_decline.tif
Latest decline values

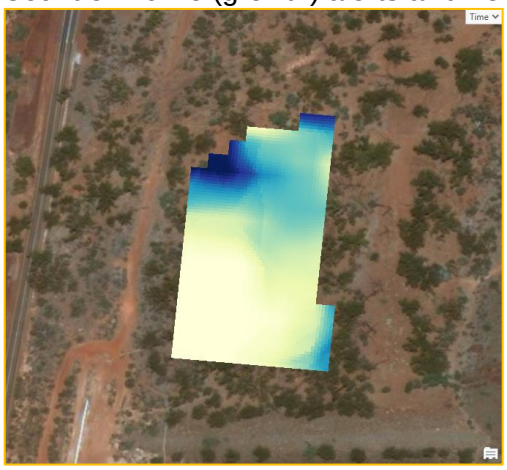


change_avg_all_time_incline.tif
Mean incline (growth) values all time

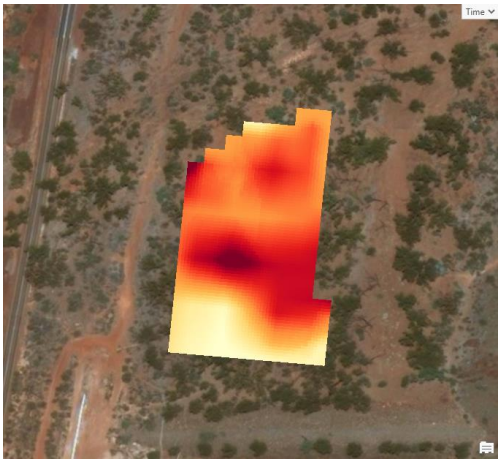


change_latest_time_incline.tif
Latest incline (growth) values

alerts_cnt_all_time_incline.tif
Count of incline (growth) alerts all time



alerts_cnt_all_time_decline.tif
Count of decline alerts all time

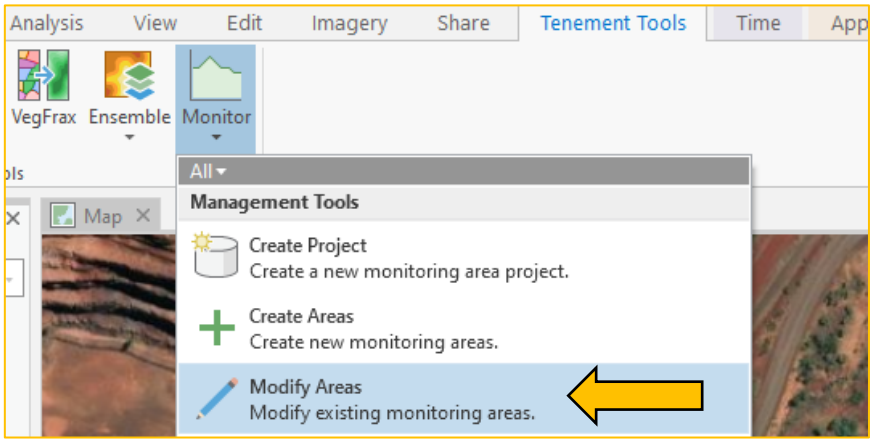


These densities show areas within your monitoring area where the highest or most frequent changes occurred. The **“hotter” (decline)** or **“colder” (incline)** the area, the more intense or frequent the change occurred. Are there any “hotspots” in your current monitoring area?

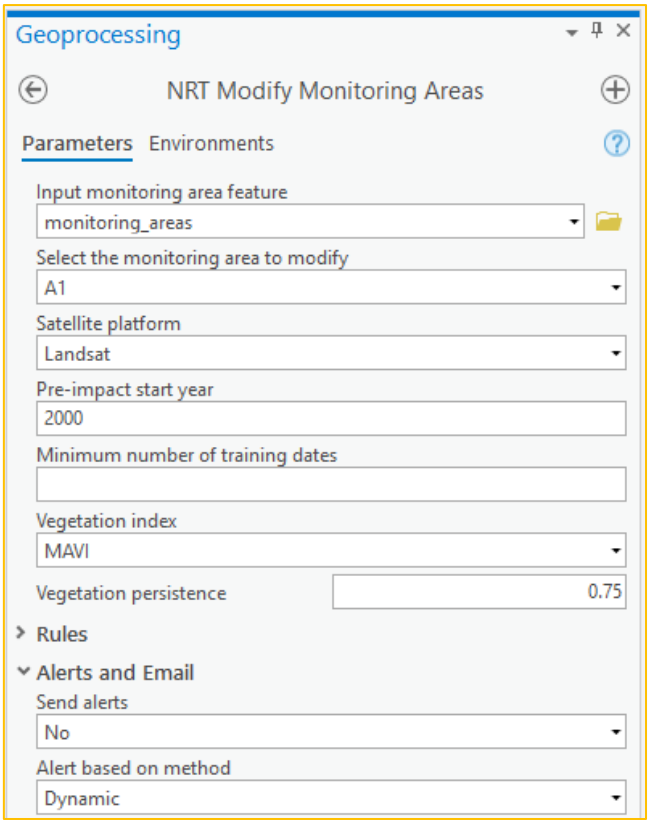
9.10 Extra: monitor area with dynamic method

As an extra exercise, we will monitor our previous area using the dynamic method (we used static earlier) and see how it differs.

First, we will modify our existing monitoring area and change the change detection method. We can modify existing monitoring areas easily. Initialise the **NRT Modify Area** tool, like so:



The NRT Modify Area geoprocessing panel will appear on the right. We need to modify the **change method and vegetation persistence value** for area **A1**, like so:



Set the input monitoring area feature to the **monitoring_areas** feature. Select the **A1** monitoring area to modify it.

Change the **vegetation persistence** value to **0.75**. Generally, the dynamic method **produces better results with a higher persistence than static** (e.g., an additional 0.25).

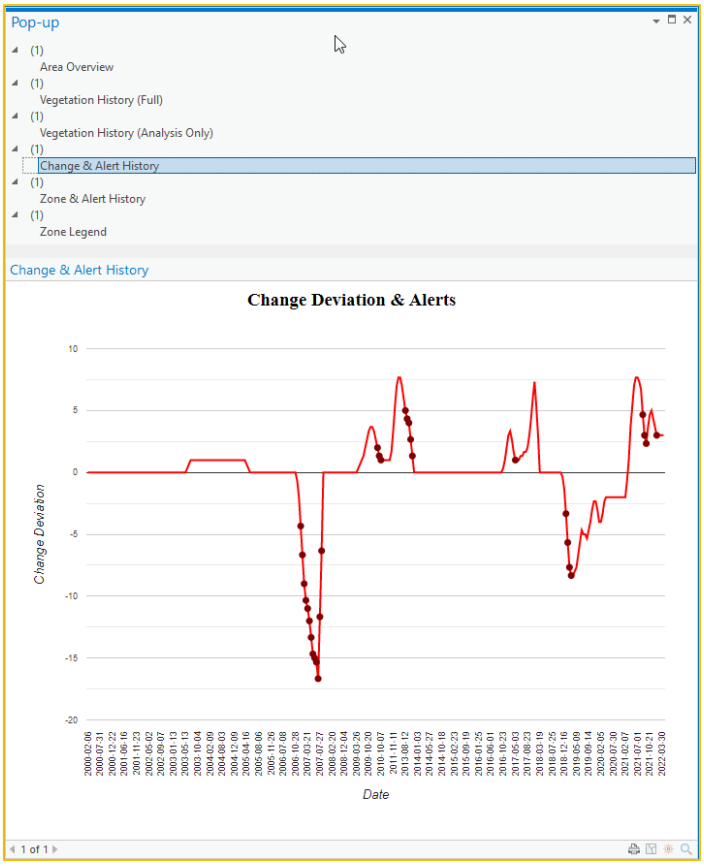
Finally, change the **alert based on method** to **Dynamic**.

When ready, click the **Run** button at the bottom of the panel. The process needs to re-download satellite data from before, so it may take a few minutes to complete (again).

9.10.1 Explore the dynamic change graph

As before, select the monitoring area and run the **NRT Graph Areas** tool:

The dynamic change history graph **should look quite different** to the static version. Here's mine:



The **dynamic** method produces different looking change graphs to the static method, mostly due to the way it **recalibrates** whenever vegetation has deviated significantly from the original preliminary impact training period.

The **static** method always uses the same initial training period, and therefore **never recalibrates**.

When the dynamic method recalibrates, the change deviation graph will **quickly return back to zero**. This indicates a new **change model has been generated** using a new period of vegetation that better represents the **current vegetation health regime**.

This recalibration will occur again only if vegetation begins to deviate away from this new regime.

See Appendix 16.6 for further information.

10 References

Ball, J.W., Robinson, T.P., Wardell-Johnson, G.W., Bovill, J., Byrne, M. and Nevill, P.G., 2020. Fine-scale species distribution modelling and genotyping by sequencing to examine hybridisation between two narrow endemic plant species. Scientific reports, 10(1), pp.1-12.

Barbet-Massin, M., Jiguet, F., Albert, C.H. and Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many?. Methods in ecology and evolution, 3(2), pp.327-338.

Brooks, E.B., Wynne, R.H., Thomas, V.A., Blinn, C.E. and Coulston, J.W., 2013. On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data. IEEE Transactions on Geoscience and Remote Sensing, 52(6), pp.3316-3332.

Brooks, E.B., Yang, Z., Thomas, V.A. and Wynne, R.H., 2017. Edyn: Dynamic signaling of changes to forests using exponentially weighted moving average charts. Forests, 8(9), p.304.

Dempster, A.P., 2008. Upper and lower probabilities induced by a multivalued mapping. In Classic works of the Dempster-Shafer theory of belief functions (pp. 57-72). Springer, Berlin, Heidelberg.

Eklundh, L. and Jönsson, P., 2012. TIMESAT 3.1 software manual. Lund University, Sweden, pp.1-82. Franklin, J., 2010. Mapping species distributions: spatial inference and prediction. Cambridge University Press.

Keppel, G., Robinson, T.P., Wardell-Johnson, G.W., Yates, C.J., Van Niel, K.P., Byrne, M. and Schut, A.G., 2017. A low-altitude mountain range as an important refugium for two narrow endemics in the Southwest Australian Floristic Region biodiversity hotspot. Annals of Botany, 119(2), pp.289-300.

Narkhede, S., 2018. Understanding auc-roc curve. Towards Data Science, 26, pp.220-227.

Phillips, S.J. and Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, 31(2), pp.161-175.

Shafer, G., 1976. A mathematical theory of evidence. Princeton university press.

Yates, C.J., Robinson, T., Wardell-Johnson, G.W., Keppel, G., Hopper, S.D., Schut, A.G. and Byrne, M., 2019. High species diversity and turnover in granite inselberg floras highlight the need for a conservation strategy protecting many outcrops. Ecology and Evolution, 9(13), pp.7660-7675.

Zhu, G., Ju, W., Chen, J.M. and Liu, Y., 2014. A Novel Moisture Adjusted Vegetation Index (MAVI) to reduce background reflectance and topographical effects on LAI retrieval. PloS one, 9(7), p.e102560.

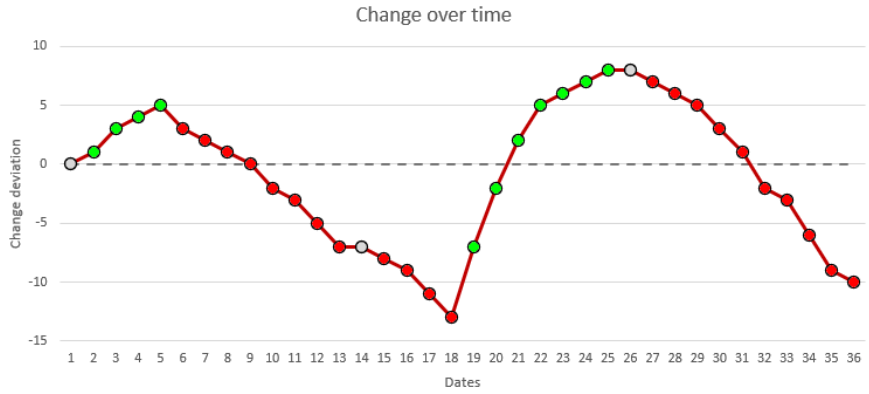
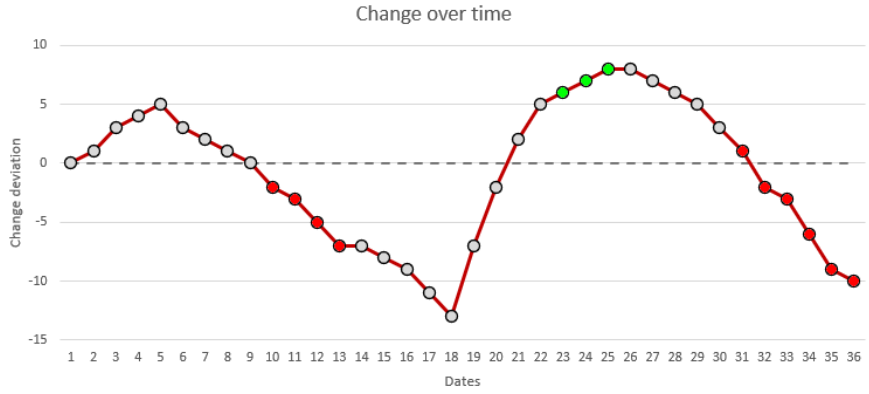
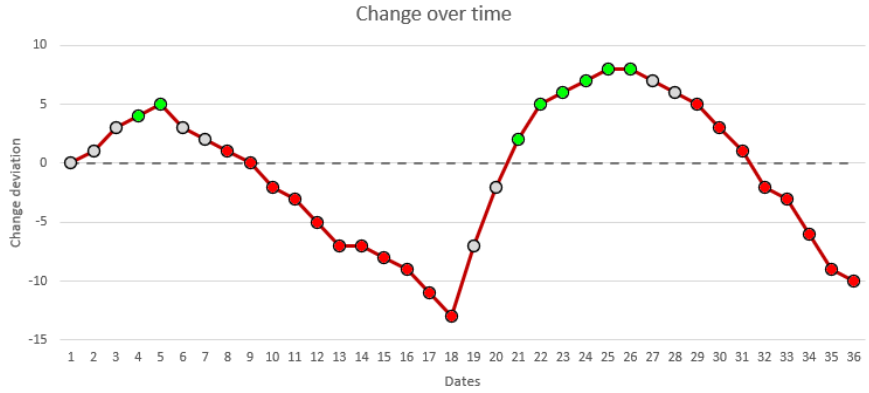
11 Appendix

11.1 Summary of phenometrics

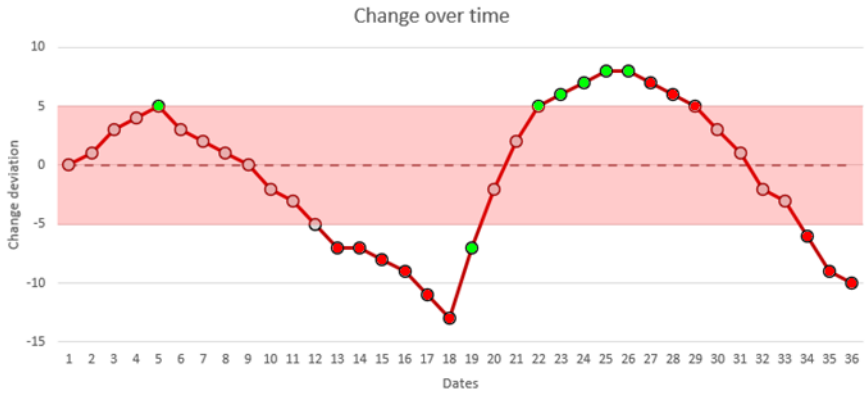
Code	Name	Description	Method
POS	Peak of Season	Highest vegetation value and time of season.	Maximum value in a time series.
MOS	Middle of Season	Mean vegetation value and time of values in top 80% of season.	Mean value and time where the left and right slope edges have increased and decreased to the 80% level of the season, respectively. Typically only used in calculation of other metrics.
VOS	Valley of Season	Lowest vegetation value and time of season.	Minimum value in a time series.
BSE	Base	Mean of the lowest vegetation values in season.	Mean value of the lowest vegetation values to the left and right of Peak of Season. Typically only used in calculation of other metrics.
SOS	Start of Season	Vegetation value and time at the start of season.	Six methods available: 1) seasonal amplitude; 2) absolute amplitude; 3) Relative amplitude; 4) LOESS STL Trend line; 5) First value of positive slope; and 6) Median value of positive slope.
EOS	End of season	Vegetation value and time at the end of season.	Six methods available: 1) seasonal amplitude; 2) absolute amplitude; 3) Relative amplitude; 4) LOESS STL Trend line; 5) First value of negative slope; and 6) Median value of negative slope.
LOS	Length of Season	Length of time (number of days) between the start and end of season.	The day of year at SOS minus EOS. Typically only used in calculation of other metrics.
ROI	Rate of Increase	The rate of vegetation "green up" at the beginning of season.	Calculated as the ratio of the difference between the left 20% and 80% levels and the corresponding time difference.
ROD	Rate of Decrease	The rate of vegetation "green down" at the end of season.	Calculated as the ratio of the difference between the right 20% and 80% levels and the corresponding time difference.
AOS	Amplitude of Season	The amplitude of vegetation values for season.	The difference between the maximum value and the VOS/BSE value. Typically only used in calculation of other metrics.
SIOS	Short Integral of Season	Represents the seasonally active vegetation and provides a larger value for herbaceous vegetation cover and smaller value for evergreen vegetation cover.	Calculated using the trapezoidal rule on the total vegetation values between season start and end minus the VOS/BSE level value.

LIOS	Long Integral of Season	Represents the total productivity of vegetation when in season.	Calculated using the trapezoidal rule between the total vegetation values between season start and end.
SIOT	Short Integral of Total	Represents total vegetation productivity throughout the season, and provides a larger value for herbaceous vegetation cover and smaller value for evergreen vegetation cover.	Calculated using the trapezoidal rule on the total vegetation values minus the VOS/BSE level value.
LIOT	Long Integral of Total	Represents the total productivity of vegetation throughout the season.	Calculated using the trapezoidal rule between the total vegetation values between season start and end.
NOS	Number of Seasons	Total number of seasons (i.e. prominent graph peaks) in time series.	Peaks detected using scipy find_peaks and any peaks are over 3 months apart.

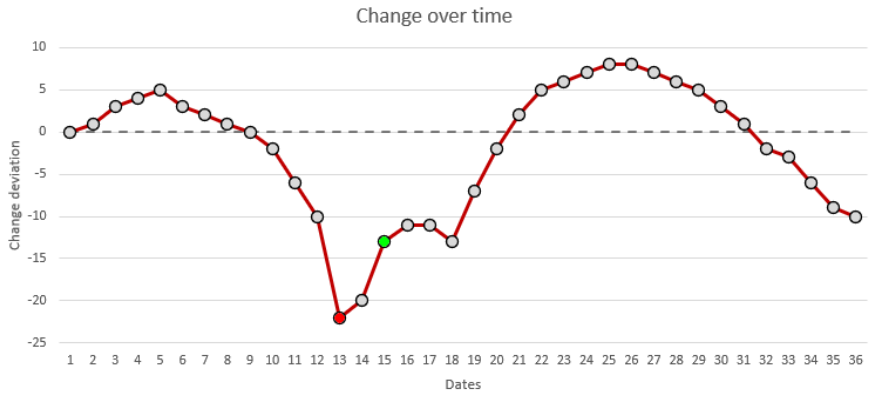
11.2 Rule one (number of consecutive and plateaus)

Rule 1 consecutives set to 3 (default). Include plateaus set to No. After 3 inclines (green) or declines (red), date is flagged. No plateaus included (see dates 13 to 14).	
Rule 1 consecutives set to 1. Include plateaus set to No. After 1 incline (green) or decline (red), date is flagged. No plateaus included (see dates 13 to 14).	
Rule 1 consecutives set to 5. Include plateaus set to No. After 5 inclines (green) or declines (red), date is flagged. No plateaus included (see dates 13 to 14).	
Rule 1 consecutives set to 3 (default). Include plateaus set to Yes. After 3 inclines (green) or declines (red), date is flagged. Plateaus included (see dates 13 to 14).	

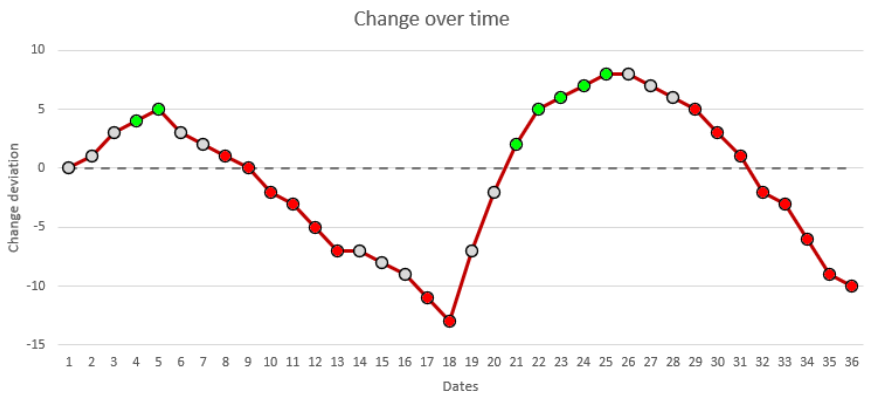
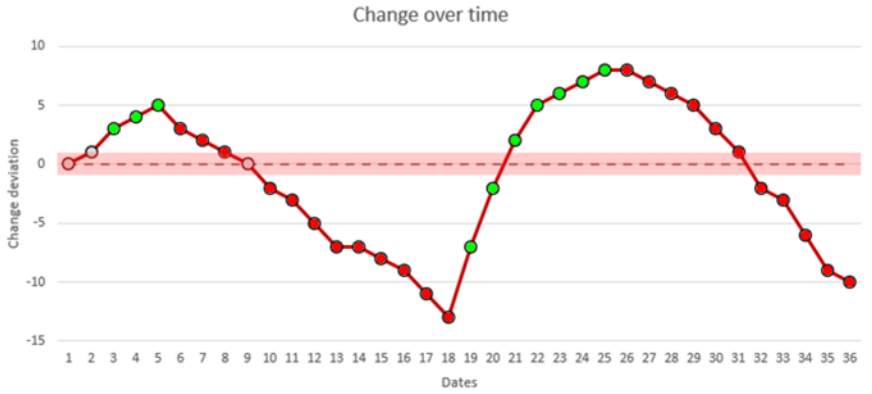
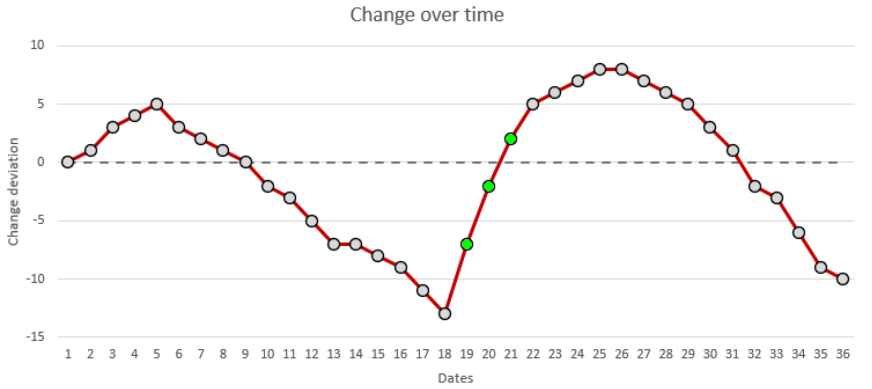
11.3 Rule two (minimum allowed zone)

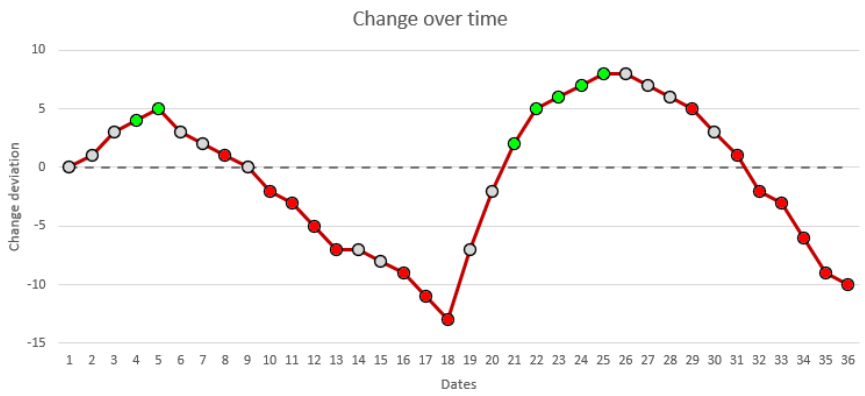
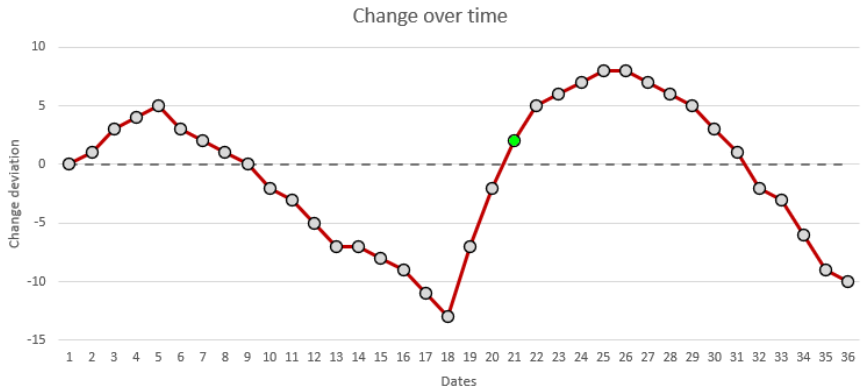
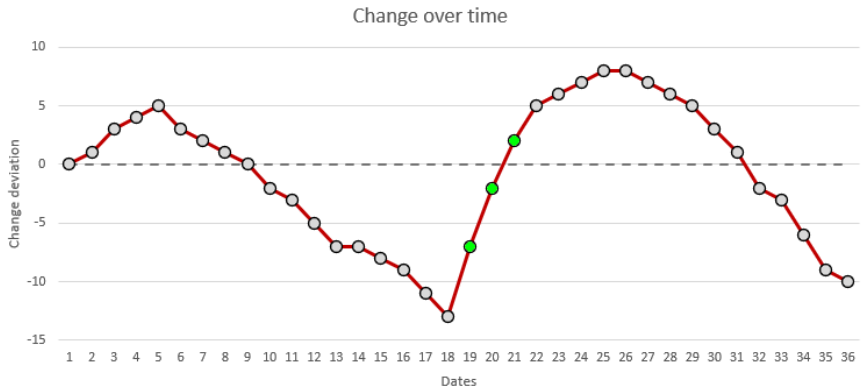
Rule 2 minimum zone set to 2 (default). Any values within the red exclusion zone will not be flagged. Incline (green) and decline (red) shown. Zone value of 2 excludes any change deviation between 1 and -1.	
Rule 2 minimum zone set to 4. Any values within the red exclusion zone will not be flagged. Incline (green) and decline (red) shown. Zone value of 5 excludes any change deviation between 5 and -5.	

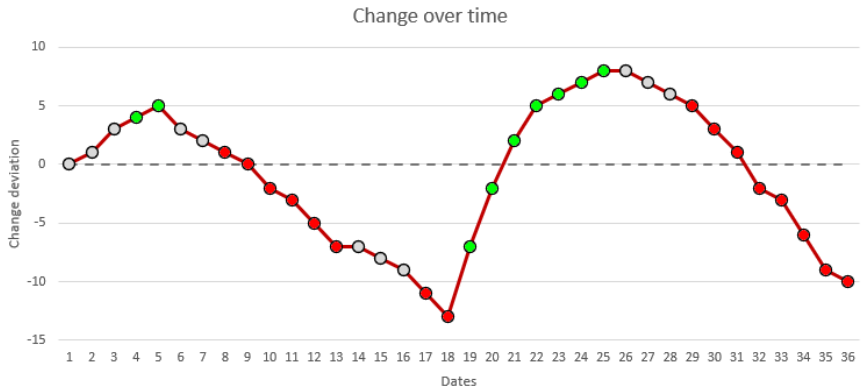
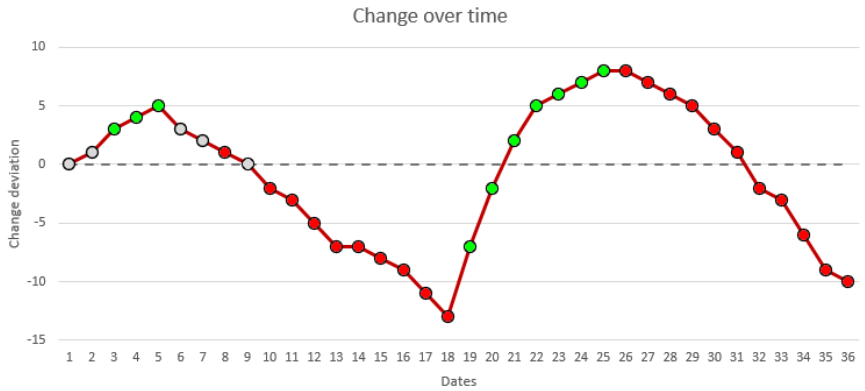
11.4 Rule three (number of zones “jumped”)

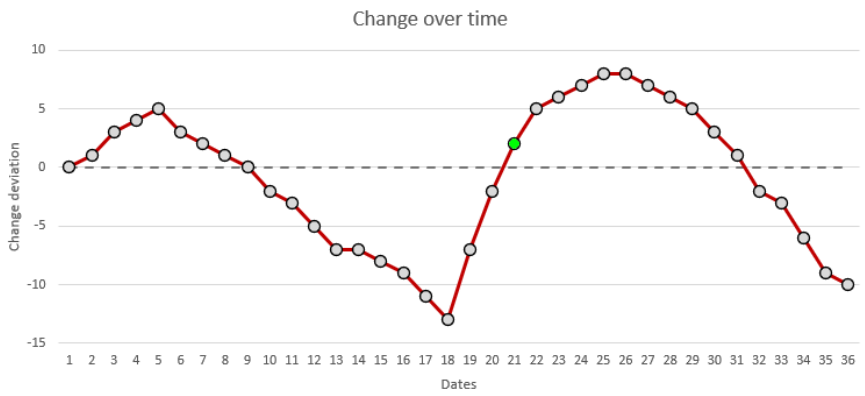
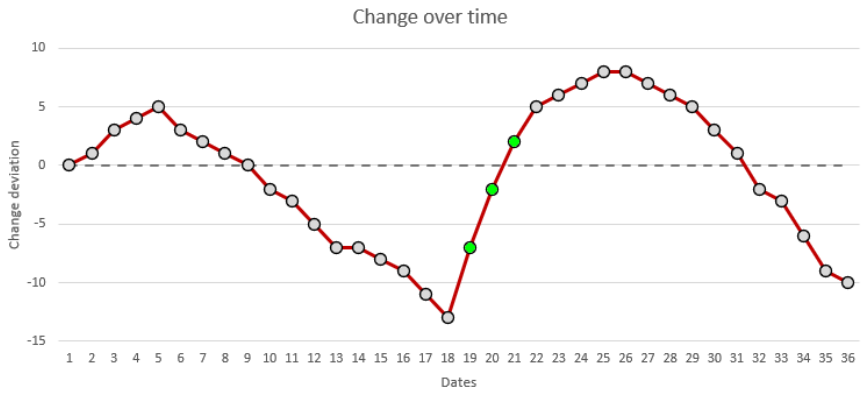
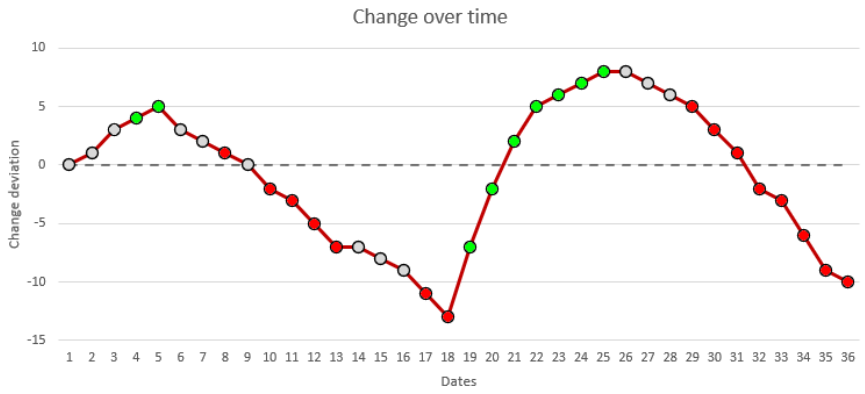
Rule 3 number of zones “jumped” set to 2 (default). When a jump of 2 or more zones detected, alert is flagged. Incline (green) and decline (red) shown. Zone leap of 2 requires 4 or more deviations for flag.	
Rule 3 number of zones “jumped” set to 4. When a jump of 3 or more zones detected, alert is flagged. Incline (green) and decline (red) shown. Zone leap of 4 requires 6 or more deviations for flag.	

11.5 Rulsets

Ruleset set to **1 only**. Only rule 1 (consecutive inclines/declines) flagged as alerts. Default rule values (3 consecutives and include plateaus set to No).	
Ruleset set to **2 only**. Only rule 2 (deviations outside of minimum zone) flagged as alerts. Default rule value (minimum zone set to 2).	
Ruleset set to **3 only**. Only rule 3 (number of zones jumped) flagged as alerts. Default rule value (number of zones jumped set to 2).	

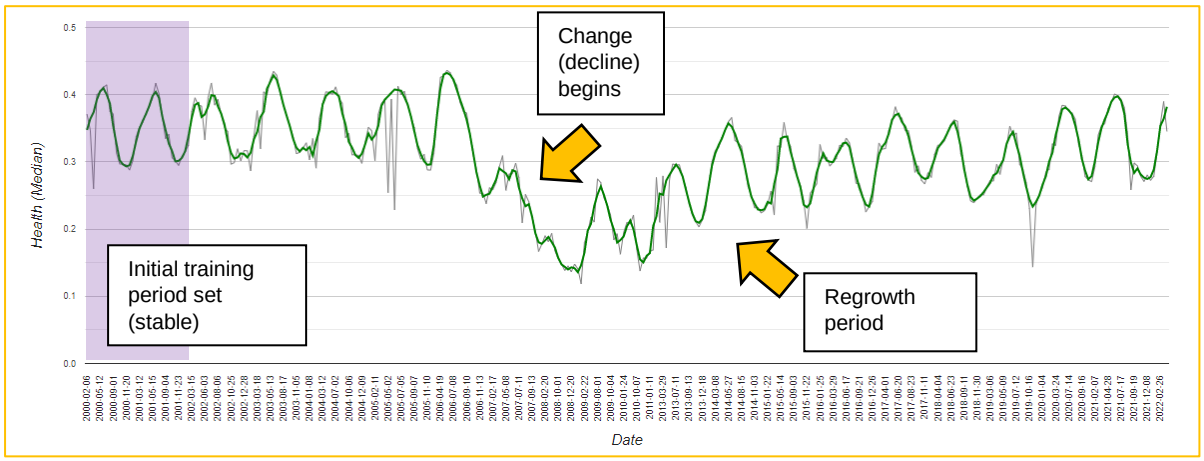
Ruleset set to **1 and 2**. Where **BOTH** rule 1 (consecutives) **AND** rule 2 (deviations outside minimum zone) occur flagged as alerts. Default rule values used (see above).	
Ruleset set to **1 and 3**. Where **BOTH** rule 1 (consecutives) **AND** rule 3 (number of zones jumped) occur flagged as alerts. Default rule values used (see above).	
Ruleset set to **2 and 3**. Where **BOTH** rule 2 (deviations outside minimum zone) **AND** rule 3 (number of zones jumped) occur flagged as alerts. Default rule values used (see above).	

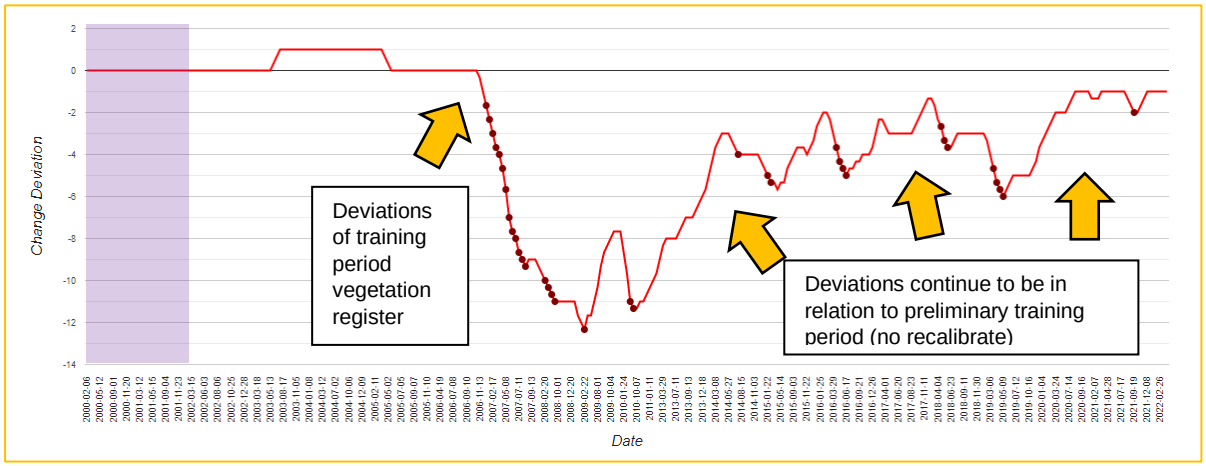
Ruleset set to **1 or 2**. Where **EITHER** rule 1 (consecutives) **OR** rule 2 (deviations outside minimum zone) occur flagged as alerts. Default rule values used (see above).	
Ruleset set to **1 or 3**. Where **EITHER** rule 1 (consecutives) **OR** rule 3 (number of zones jumped) occur flagged as alerts. Default rule values used (see above).	
Ruleset set to **2 or 3**. Where **EITHER** rule 2 (deviations outside minimum zone) **OR** rule 3 (number of zones jumped) occur flagged as alerts. Default rule values used (see above).	

Ruleset set to **1 and 2 and 3**. Where **ALL** rule 1 (consecutives) **AND** rule 2 (outside minimum zone) **AND** rule 3 (zones jumped) occur flagged as alerts. Default rule values used (see above).	
Ruleset set to **1 or 2 and 3**. Where rule 1 (consecutives) **OR** rule 2 (outside minimum zone) AND rule 3 (zones jumped) occur flagged as alerts. Default rule values used (see above).	
Ruleset set to **1 and 2 or 3**. Where **BOTH** rule 1 (consecutives) **AND** rule 2 (outside minimum zone) **OR** rule 3 (zones jumped) occur flagged as alerts. Default rule values used (see above).	
Ruleset set to **1 or 2 or 3**. Where **BOTH** rule 1 (consecutives) **AND** rule 2 (outside minimum zone) **OR** rule 3 (zones jumped) occur flagged as alerts. Default rule values used (see above).	

11.6 Alert method

Vegetation (2000 – 2022) and static change model output

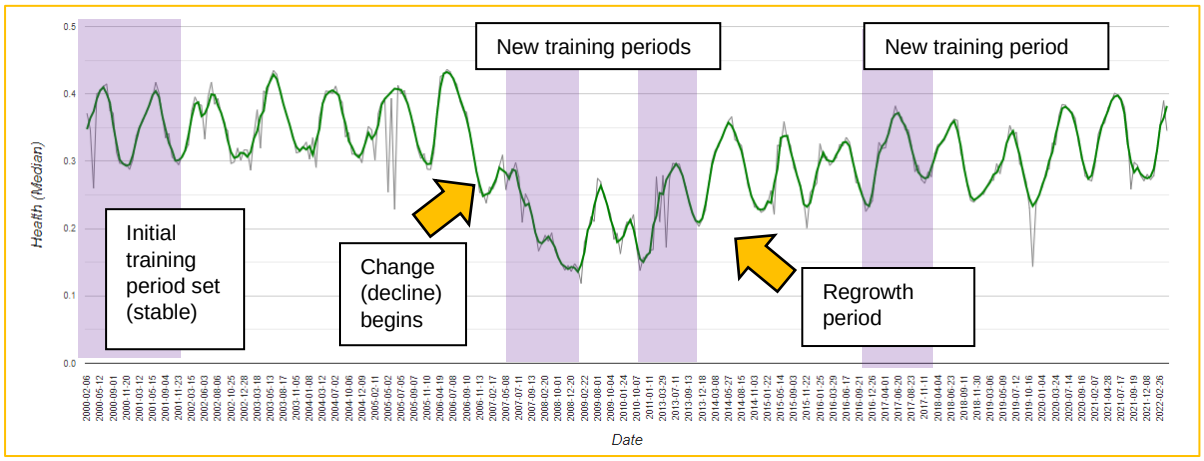




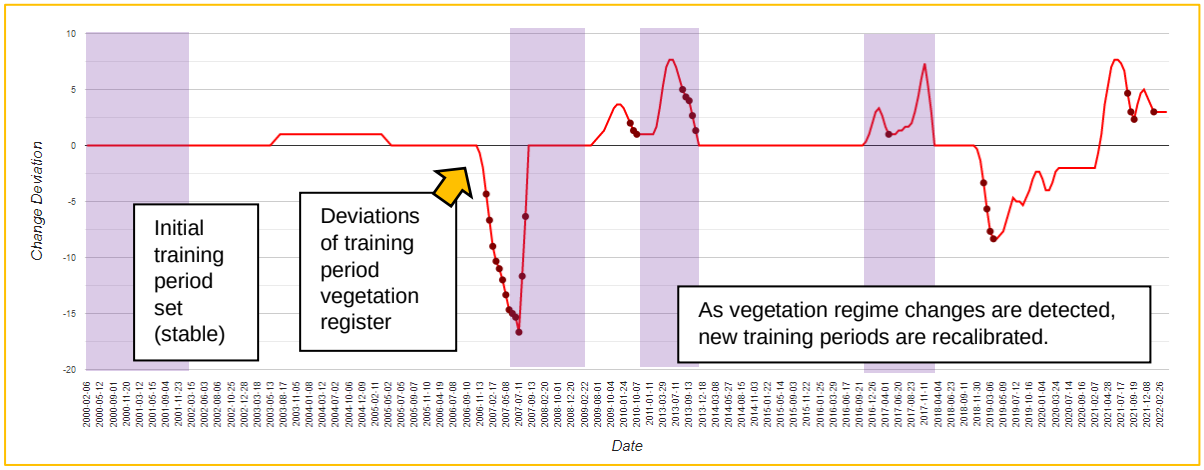
The static method requires a single training period (set by the starting year). Here this is approximated and represented as the purple bar. As new vegetation index values are added to the monitor area database, these values are continuously compared to a harmonic regression model created by algorithm during the training period forever. The model will never recalibrate or build a new training period, even if the vegetation is radically different now.

As a general rule of thumb, this method is optimal when detection of severe impacts in highly stable and green vegetation is of focus.

Vegetation (2000 – 2022) and dynamic change model output



Dynamic change model output



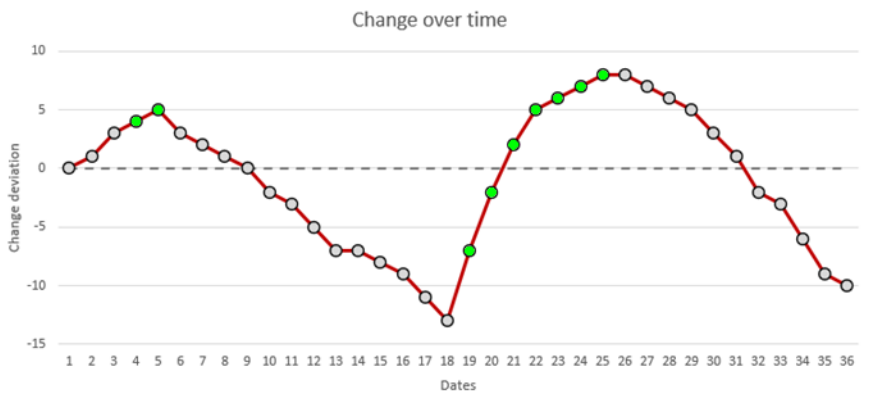
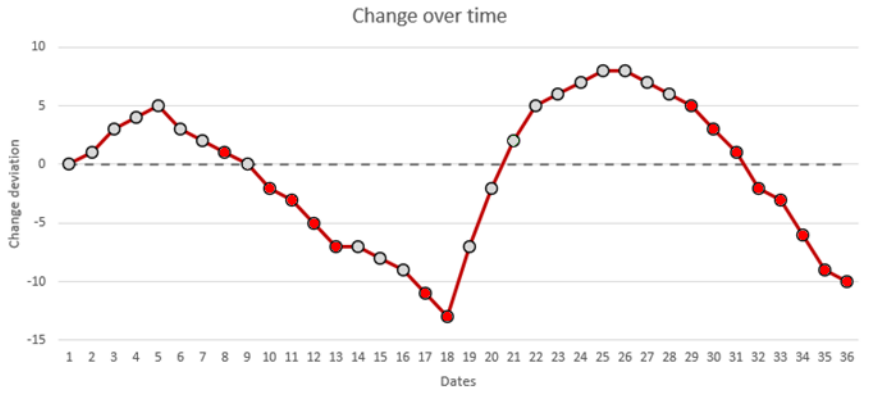
The dynamic method requires an initial training period (set by the starting year). Here this is approximated and represented as the purple bar. As new vegetation index values are added to the monitor area database, these values are continuously compared to a harmonic regression model created by algorithm during the training period.

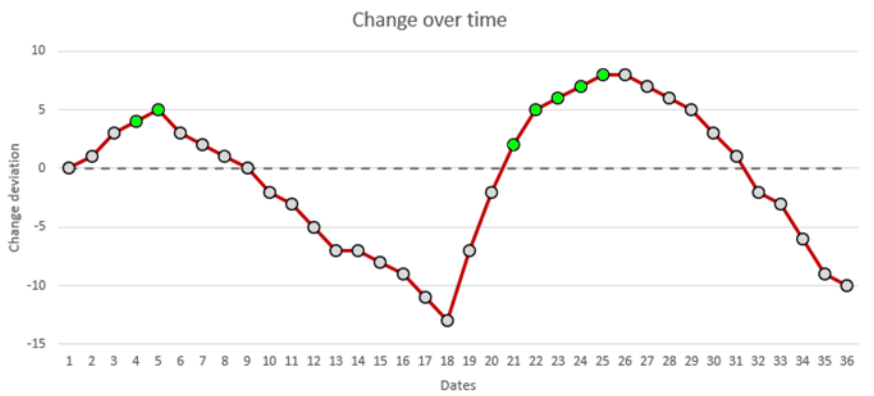
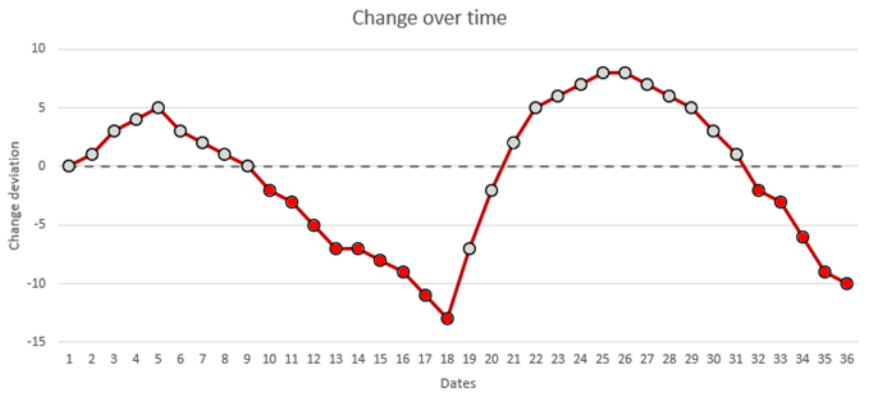
However, unlike the static method, if vegetation index values begin to deviate too far from the initial training period, the model will recalibrate based on what it has calculated to be a new vegetation regime. This is best demonstrated looking at the second purple bar from the left: after a large impact, the model recalibrated and the change deviations reset to zero.

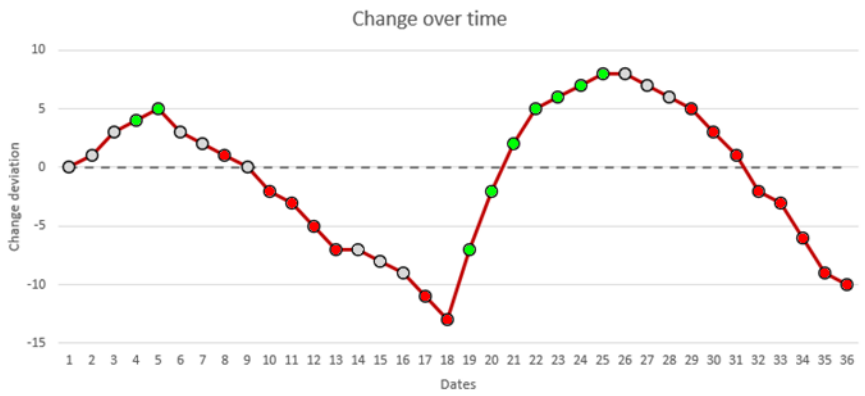
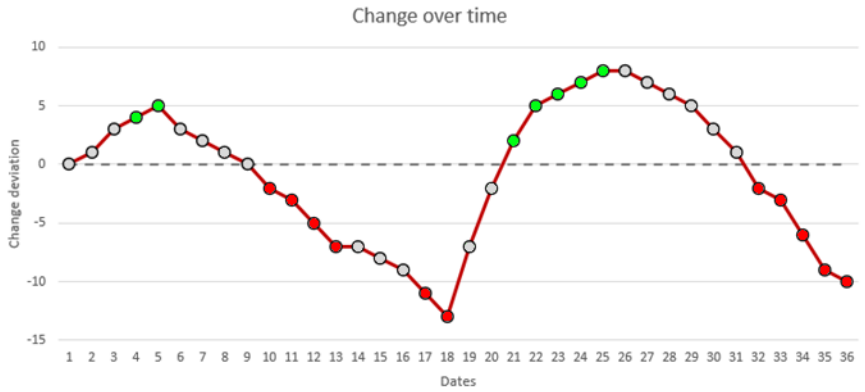
As more vegetation data is appended to the change database, the dynamic method will continue to recalibrate the harmonic regression model if vegetation deviates.

As a general rule of thumb, this method is optimal when detection of gradual or severe impacts in stable or unstable vegetation is of focus.

11.7 Direction of change for trigger alert

Incline only (any). Any incline (growth) detected in any zone is a candidate for alert if it breaks the ruleset.	
Decline only (any). Any decline detected in any zone is a candidate for alert if it breaks the ruleset.	

Incline only (+ zones only). Incline (growth) detected in + zone only may be a candidate for alert if it breaks the ruleset.	
Decline only (- zones only). Decline detected in - zone only may be a candidate for alert if it breaks the ruleset.	

Incline or decline (any). Any Incline (growth) or decline detected may be a candidate for alert if it breaks the ruleset.	
Incline or decline (+ / - zones only). Incline (growth) detected in + zone or decline detected in – zone may be a candidate for alert if it breaks the ruleset.