LandScape+
User guide

Introduction - What is LandScape+?

LandScape+® is a web-based software tool that allows you to create a map of 'landscape types' in a survey area and use this model to assess geochemical outliers from survey data by landscape type. The landscape types are created from publicly available spatial data and are considered proxies for the physical regolith types in the model area. This provides a basic, first-pass tool for the interpretation of regolith landscapes and statistical outliers within your geochemical data.

Commonly all geochemical samples in a survey are considered as one population and the highest concentrations in the data set are considered outliers without consideration of landscape context. This ignores the underlying processes that may affect metal dispersion. For example, elevated metal concentrations may be readily identifiable as outliers in a geochemical data set where samples were collected over exposed outcrop or shallow residual materials of comparably more metal-rich parent materials. The same area would have a much weaker elemental signal in samples collected over moderately thick depositional landscape types such as a sand plain. It is therefore important to consider landscape context when evaluating soil geochemical results. All samples in a given landscape type are essentially sub-populations within the sample set that can be treated separately and assessed in a more appropriate spatial context.

Methodology – How does LandScape+ work?

1 Landscape model

LandScape+® uses a Python workflow that looks for clusters (groups) of spatial data with similar properties and gives each cluster a unique colour (here referred to as *landscape types*). The source data used for the models is summarised in Table 1. You can find the specific version of source data for your model in the *readme.txt* file in your *Data Package*.

The machine learning does not take any information on the geographical location or soil geochemistry into account. The spatial data layers that are used to cluster the landscape types are selected for their relationship to regolith features and provide information on landscape position, depth of transported cover, and parent and/or regolith materials. To minimise the introduction of human interpretation, LandScape+® does not include surface geology or regolith maps.

Table 1: Source data layers

Spatial input layer	Landscape information	Description	Spatial resolution
DEM	General landscape position	Copernicus GLO-30 Digital Elevation Model	~30 m
MrVBF	Estimate of depth of transported cover	Multi-resolution Valley Bottom Flatness index (Gallant et al. 2012)	~30 m
Radiometric data	Indication of parent material	Radiometric data, K pct, U ppm, Th ppm (Poudjom Djomani & Minty 2019)	~90 m
Regolith ratios	Regolith material	A product derived from ratios of bands from Sentinel-2 satellite data (adapted from Gozzard 2005[1])	~20 m

[1] Gozzard, J.R., 2005. Part 3: Regolith-landform mapping using remotely sensed imagery in IGES 2005. Workshop 1.3, Regolith mapping, workshop notes: Perth, Western Australia, IGES 2005, 73p.

The workflow extracts pixel values from spatial data layers for each model area and uses an algorithm, Uniform Manifold Approximation and Projection (UMAP) to reduce each sample point to three dimensions and applies an RGB (red, green, blue) colour scheme. This is referred to as dimensionality reduction. This provides a basis for visualisation and prepares the data for clustering (colouring by groups). An agglomerative hierarchical clustering algorithm is applied to the data with varying numbers of clusters (4 to 16) to separate out different landscape types.

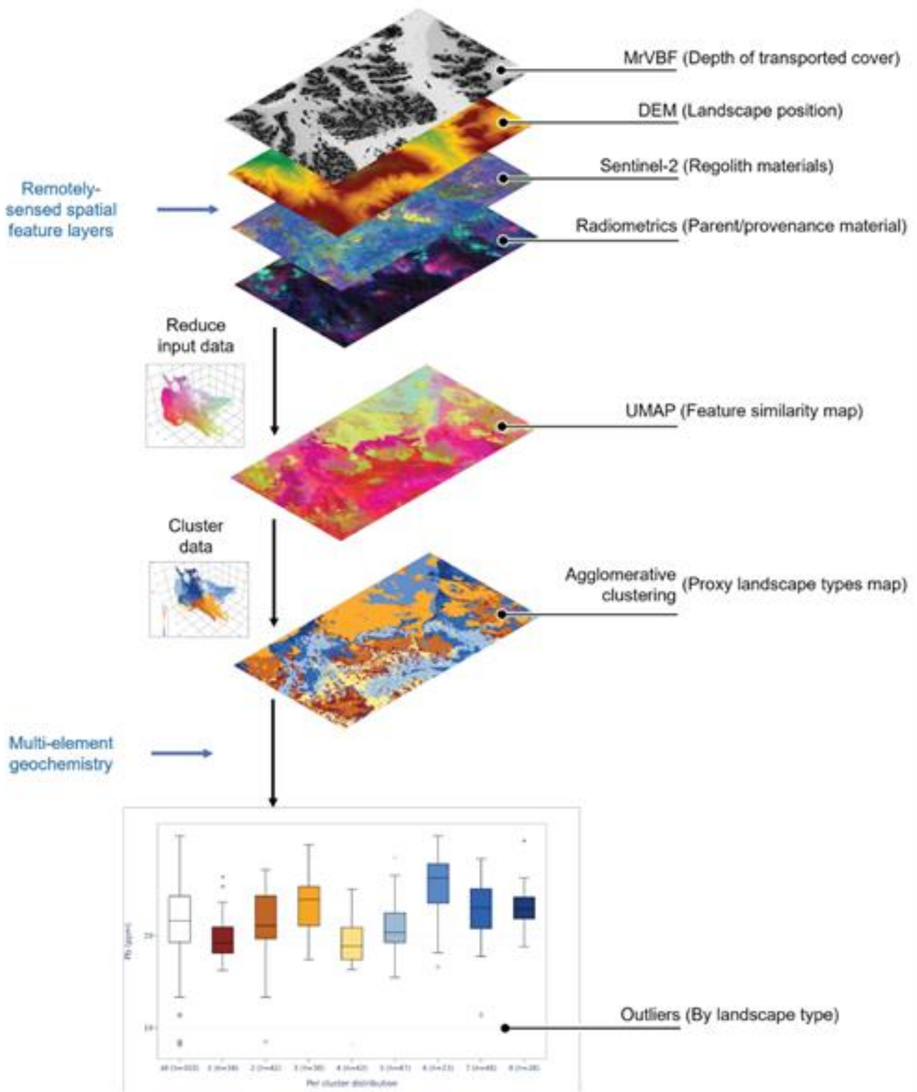


Figure 1: General workflow modified after Henne et al. 2023.

After the data is clustered, each pixel is reassigned its coordinates and projected onto a map coloured by the landscape cluster it belongs to. This is the basis of the clustering model. A 2D map of the pixels coloured by UMAP RGB value is also provided. Ideally the landscape clusters will resemble the UMAP RGB closely.

For a detailed description of the methods, please refer to the open access publications Henne et al. (2023) and (2025).

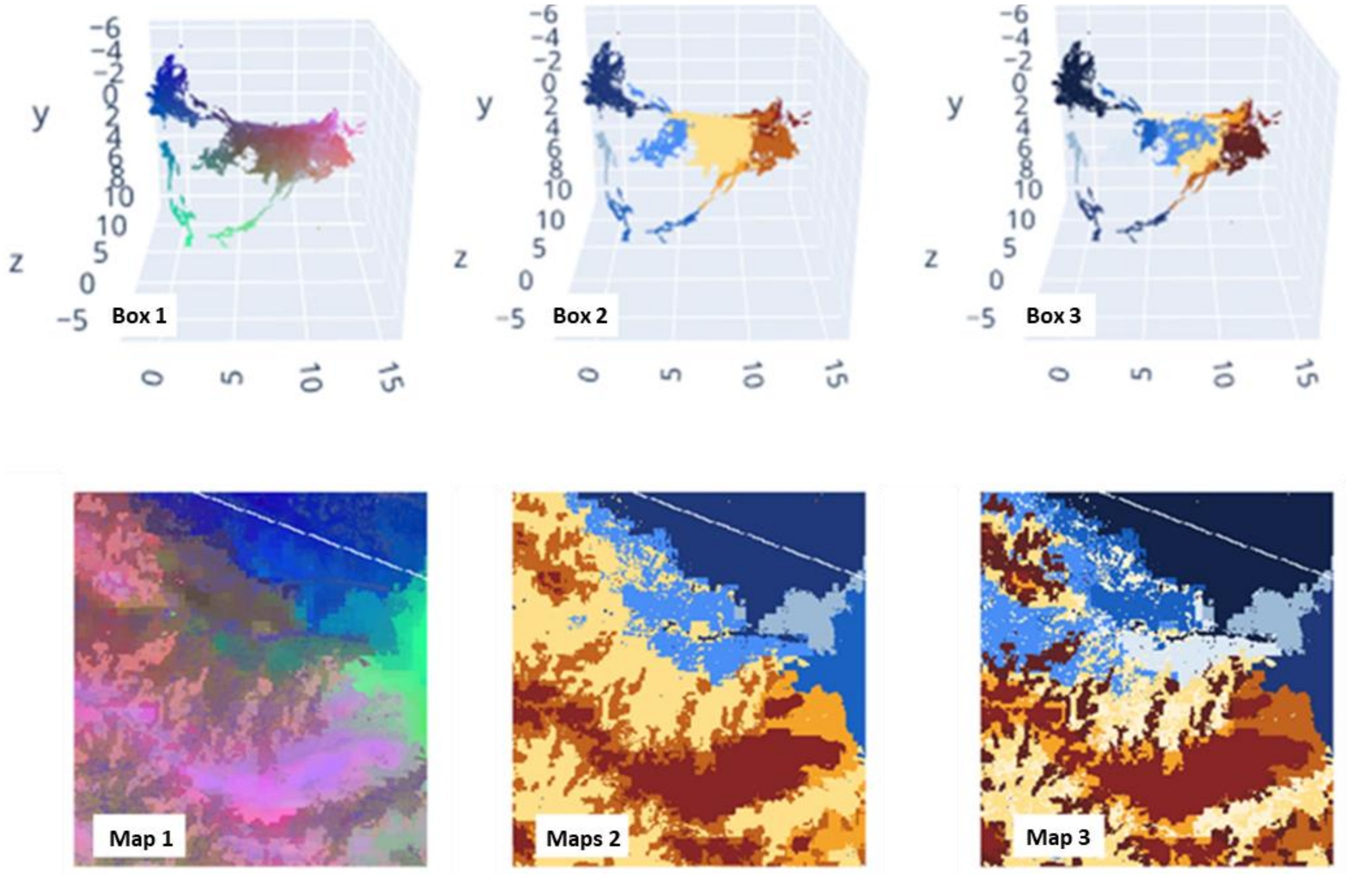


Figure 2: Top row: Examples of clusterboxes showing how pixels are associated in 3D space using an RGB colour scheme (UMAP, Box 1) and coloured by landscape type (Box 2 with eight landscape types, Box 3 with twelve landscape types). Bottom row: Pixels from the clusterboxes projected in 2D spatial context, creating proxy regolith maps with an RGB colour scheme applied (Map1) and with eight (Map 2) and twelve (Map 3) landscape types. The white diagonal line is a masked-out road and is not part of the clustering schemes.

2 Outlier calculations

Outliers are calculated for each element from the dataset uploaded into LandScape+. Elemental outliers are calculated on log-transformed data and are defined as values that are greater than 1.5 times the interquartile range. Geochemical data are also grouped by their corresponding landscape type and outliers are calculated for each of these landscape types. LandScape+ outputs these as boxplots (Tukey plots) for each element by landscape type and as GeoPackage, ESRI Shapefiles and CSV files.

3 Principal Component Analysis

The LandScape+ Data Package includes Principal Component Analysis. Principal Component Analysis does not consider landscape types for the results, only the chemistry of the entire survey in the modelled area. Your data must have more elements than principal components (>5). Principal components cannot be viewed in the online User Interface. They can only be accessed in the Data Package download. For more information refer to the *Principal Component Analysis* section at the end of this document.

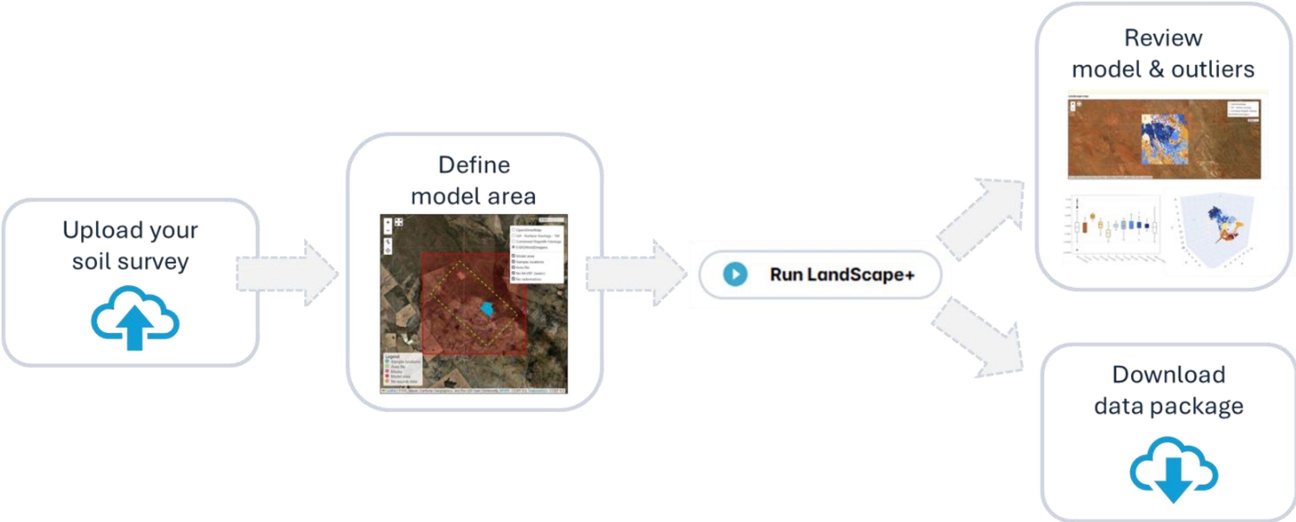
LandScape+ application workflow

1 Overview

The web-based LandScape+ application will guide you through the process of uploading soil geochemistry samples, creating a landscape model for these samples, and reviewing the outputs. Once the model is run, you can download the outputs as a *Data Package* at any time for the duration of your licence (1 year from licence activation). The online *User Interface* is designed to make review of the outputs as easy as possible. Once a model is run, any actions in the *User Interface* will not change the *Data Package*.

There are four main steps:

- Upload your samples
- Define your model area
- Run a landscape model
- Review and download your results



2 User Interface – Main features

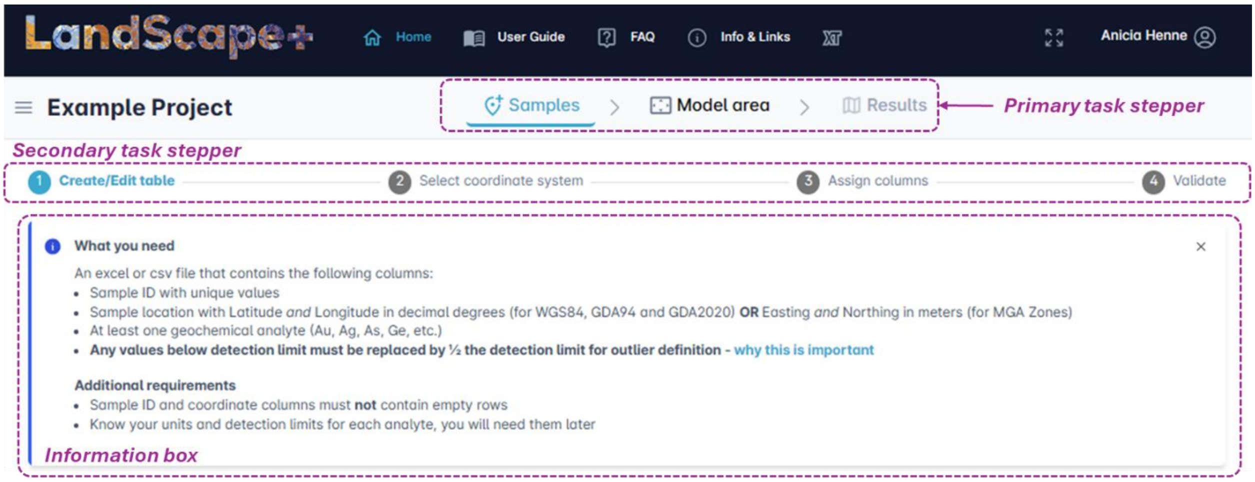
The interactive *User Interface* is designed to guide users through the LandScape+ workflow. The following are the main features you will encounter on every page.

Task stepper

Task steppers at the top of each page will prompt you to follow the process for a first-time user. *Task steppers* are located at the top of each page with two tiers (primary and secondary). The *Primary task stepper* at the top contains the major workflow steps (*Sample upload*, *Model area* configuration and reviewing *Results*). The *Secondary task stepper* (bottom) guides you through each task in detail. You can skip steps in the *Secondary task stepper* once you are familiar with the software.

You can return to the previous steps at any time until you run the model. Your settings will only be locked in once you hit "*Run LandScape+*" AND confirm the warning message that will pop up. Once the model has run and you are in the Results interface, the *Secondary task steppers* are cumulative, so any option from a previous step will still be available to you.

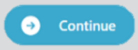
If, after review of your results, you wish to make minor adjustments to your model area, you can return through the Primary task stepper to adjust your settings. **You can run the model up to a maximum of three times.**



Continue button

Once you have completed each necessary step the *Continue button* will turn from grey to blue.

CSIRO blue indicates the next step



Information box

Information boxes provide help text. They appear by default and can be minimised by clicking on the grey cross. Click on *Information icons* to access additional help text.

Click on the ***Information icon*** to view help text

Minimise by clicking the cross

Screen scale and layout settings

The recommended display settings (scale and layout) for the LandScape+ user interface is 100 to 125 %. If your screen scale and layout is set higher, you may have to scroll down to access all page content. To change this on your PC go to Settings → System → Display → Scale and layout, or, if you are logged into LandScape+ in your browser, hit *Ctrl +/-*.

Progress interruption

The software saves you progress when you click on *Action buttons* (e.g., the *Continue button*). – If you (or your internet connection) get interrupted it will pick up near where you left next time you log in.

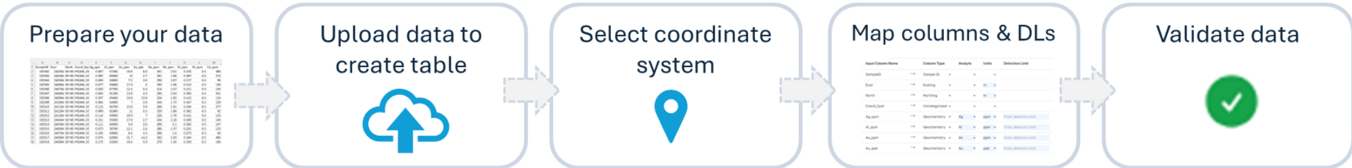
Step 1 - Sample upload

LandScape+ requires soil survey results to outline a model area. It has in-built prompts and validation steps to help the user upload their data in the required format.

There are five main steps:

1. Prepare your data
2. Upload your data and define table setting
3. Select the appropriate coordinate system
4. Assign columns and fill in details
5. Run a data validation in the software

The software saves you progress when you click on *Action buttons* (e.g., the *Continue button*). – If you (or your internet connection) get interrupted it will pick up near where you left next time you log in.



1.1 Prepare your data

Quality inputs = quality outputs

Good models can only be produced from good data. To make the sample upload process as smooth as possible, prepare your data prior to upload.

✓ *Data format*

The sample upload interface can input **CSV** and **Excel** files. All data must be contained in one CSV file or one Excel workbook sheet

Compulsory columns

As a minimum your data must contain the following 3 columns:

- Sample ID with unique values
- Sample location with Latitude *and* Longitude in decimal degrees (for WGS84, GDA94 and GDA2020) **OR** Easting *and* Northing in meters (for MGA Zones)
- At least one geochemical analyte (Au, Ag, As, Ge, etc.)

There is no upper limit on the amount of columns you can upload.

Additional file requirements

- All values below the detection limit must be replaced by ½ the detection limit
- Sample ID and coordinate columns must **not** contain empty rows
- Your file must contain a header row to label/describe the columns, e.g., Sample ID, Cu ppm
- All analytical results must be positive numerical values and cannot be zeros (0)
- Know your units and detection limits for each analyte (they can but do not have to be in the file you upload)

✓ *Values below the detection limit*

Any values below the detection limit in your geochemical data must be replaced by ½ the detection limit for outlier definition.

Not replacing values below the detection limit with ½ the detection limit **will** affect outlier definition. The user can choose to ignore these warnings at their own risk.

Background

The statistical definition of outliers by landscape type requires the imputation (replacement) of values below the detection limit as their exclusion can result in strong bias on statistical evaluation. Many imputation methods have been proposed; some of the most common introduce a positive bias (e.g., use of detection limit) or a negative bias (e.g. using 0).

This is why we recommend the replacement of all values below the detection limit with half the detection limit value. This is commonly used in the statistical evaluation of soil geochemical data and close to the simple substitution replacement factor suggested by Sandford et al. (1993) for low degrees of missing data. Where large amounts of analytical results are below the detection limit, this should always be considered when interpreting statistical calculations.

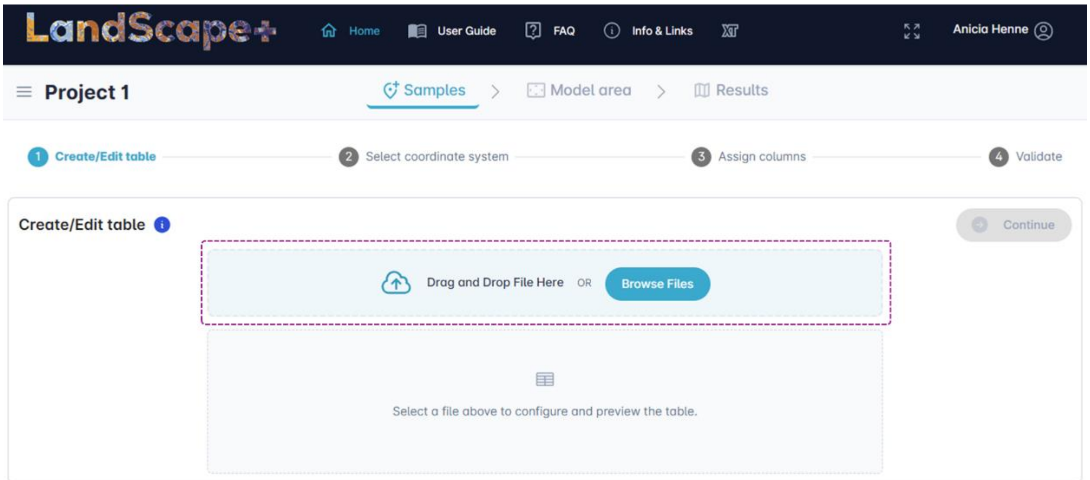
✓ Datasets with multiple detection limits

It is common for exploration surveys to be a combination of multiple laboratory analyses from different points in time. Detection limits for the same analyte may have changed since the first samples were analysed despite using the same analytical technique. This is especially pertinent when datasets contain historic data from previous exploration activities. The recommended action is to replace all values below the detection limit by ½ the lowest detection limit for each analyte. Extra care should be taken to replace the correct values as in this case the built-in validation of detection limit values (see last step) may not be effective. The validation process might flag values as erroneous.

1.2 Create table

✓ *Upload samples*

Upload your sample data by either drag-dropping the prepared file or using the *Browse Files* button. The app will create a copy of your data on the server. This copy will be used to run the model. Your original data file (local copy) will **not** be changed.



✓ *Adjust table settings*

The *User interface* allows you to adjust table settings to fit your data (left hand panel). A preview of your data is visible on the right-hand side (if your computer screen's scale and layout is very high (>125 %), you may need to scroll down as the *Table preview* might be below the *Table settings* pane).

The default settings accommodate most data files. Check if they apply to your data and see below for detailed explanations.

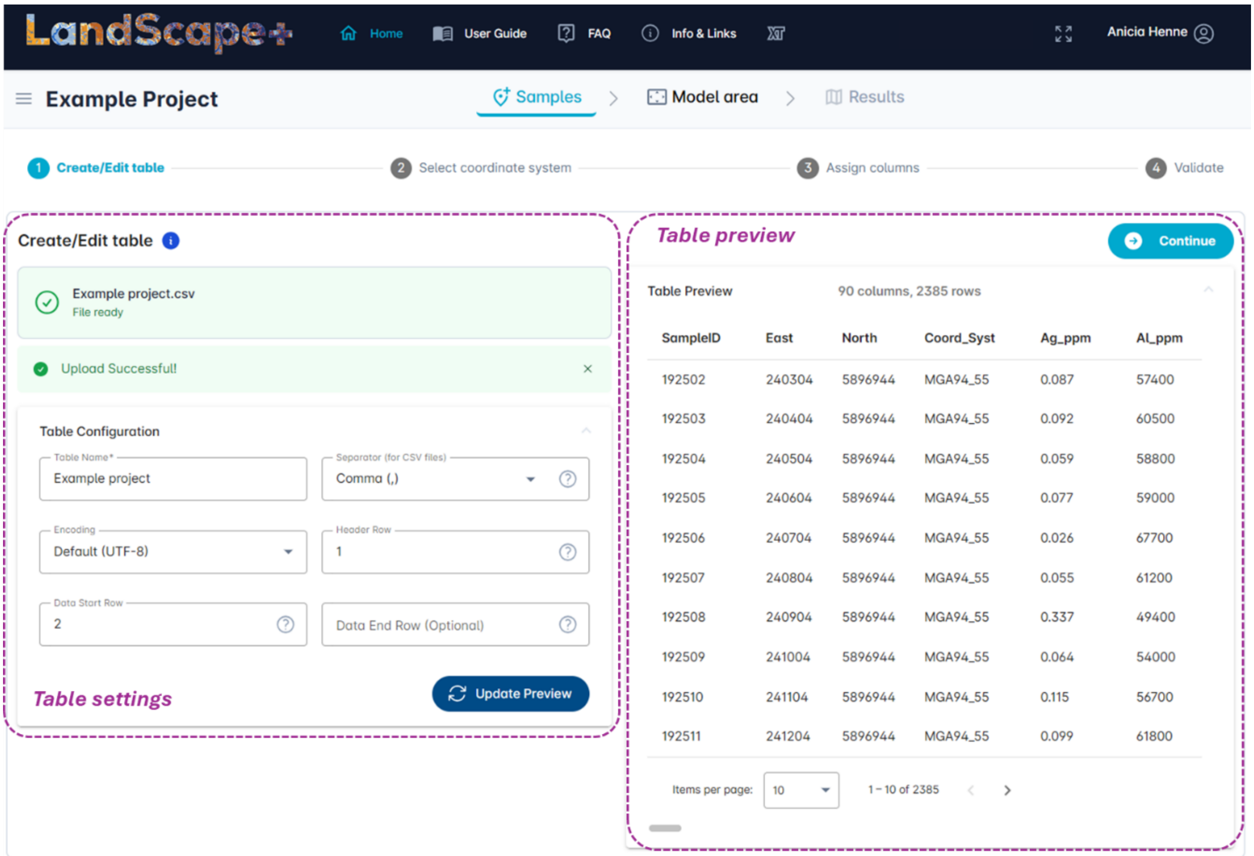


Table settings for CSV files

The *Table settings* are set to default formats that are most common. The upload also automatically recognises whether your file is a CSV or an Excel file. You can hit the *Update Preview* button at any point to view how the table settings affect your data read-in.

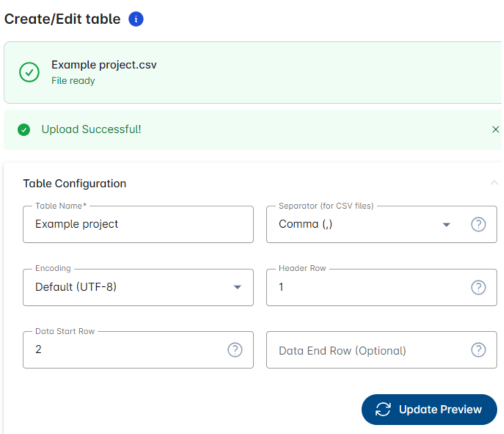


Table name is automatically read-in from the uploaded file name.

Separator default setting is *Comma (,)* (comma separated). You may choose a different format from the drop-down menu if your data looks incorrect in the *Table preview*.

Encoding default setting is *UTF-8*. You may choose a different encoding from the drop-down menu if your data looks incorrect in the *Table preview*.

Header Row default setting is 1 (the first row of your data sheet contains the headings for your columns). You can change this to any other row. The read-in will ignore any rows above the header row if it is not set to 1.

Data Start Row default setting is 2 (the second row of your data sheet). You may change this if your data does not start in this row (e.g., you may have detection limits and/or units in the row(s) below your header row).

Data End Row default setting is blank. By setting a data end row you can specify not to include all data.

Table settings for Excel files

The *Table settings* are set to default formats that are most common. The upload also automatically recognises whether your file is a CSV or an Excel file. You can hit the *Update Preview* button at any point to view how the table settings affect your data read-in.

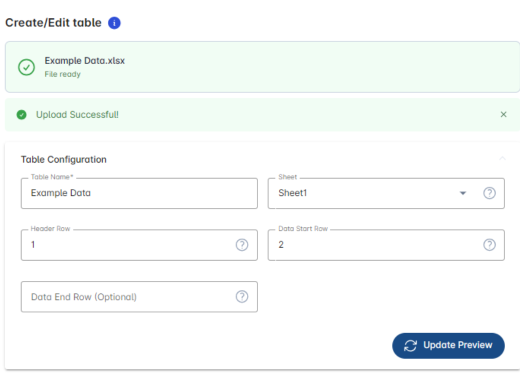


Table name is automatically read-in from the uploaded file name.

Sheet: If your Excel workbook contains multiple sheets, pick the correct sheet with your data from the drop-down menu.

Header Row default setting is 1 (the first row of your data sheet contains the headings for your columns). You can change this to any other row. The read-in will ignore any rows above the header row if it is not set to 1.

Data Start Row default setting is 2 (the second row of your data sheet). You may change this if your data does not start in this row (e.g., you may have detection limits and/or units in the row(s) below your header row).

Data End Row default setting is blank. By setting a data end row you can specify not to include all data.

✓ *Check settings and continue*

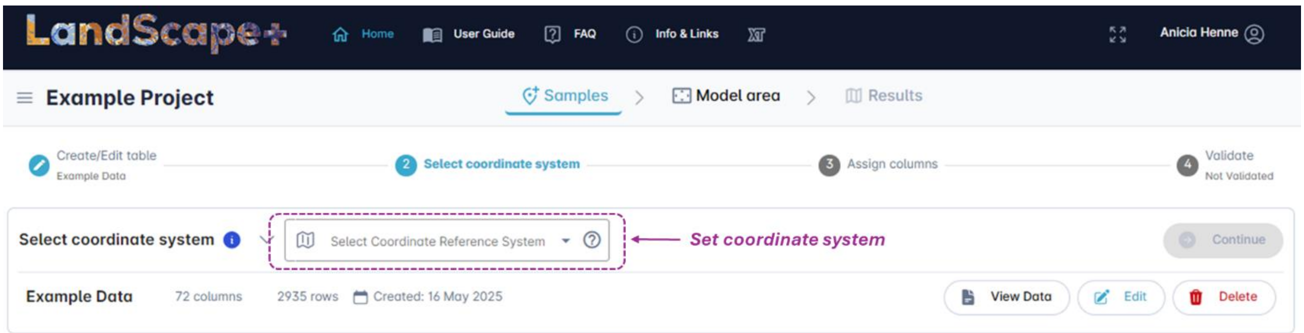
Once you are satisfied with your table settings, click on the *Continue button* to go to the next step.

1.3 Select the coordinate system

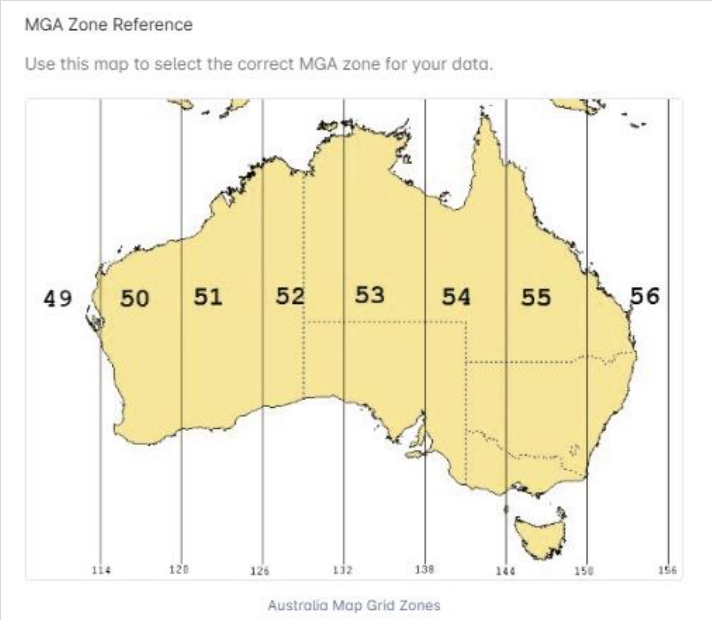
Select the correct coordinate reference system for your data from the drop-down-menu. You can click on the *View Data* button to double check the information in your uploaded data.

If you find you have made a mistake, e.g., you realise now that your data does not have coordinates, you can click on the *Edit button* to return to the previous menu. You can delete your data using the *Delete button*. Amend the source file and upload the new version.

Once you have selected the coordinate reference system the *Continue button* will turn from grey (Continue) to blue (Continue). Click on the *Continue button* to go to the next step.



If you are unsure of your coordinate system and want to double check, click the grey question mark icon in the drop-down menu to view an MGA Zone reference map. Note the longitude references at the bottom of the image.



If your samples spread across different MGA zones you must convert them to decimal degrees and repeat the upload.

1.4 Assign columns, units and detection limits

The software tool can map some information automatically. However, you must check that all columns have been mapped correctly and assign those the software cannot identify.

✓ *Assign Column Types*

Assign the appropriate column type from the drop-down menu. Options are Sample ID, Geochemistry, Easting, Northing, Latitude, Longitude, Uncategorised, and Soil Properties (for pH and EC).

You **must** assign the following to continue:

- A "Sample ID" column. All sample identifiers must be unique.
- Either "Latitude" *and* "Longitude" **OR** "Easting" *and* "Northing" (for MGA). If you have both, set one to *Uncategorised*.
- At least one geochemical analyte (e.g. Au, Ag, As, Ge).
- Any uncategorised columns will <u>not</u> be included in the model outputs.

✓ *Assign an analyte to each Geochemistry column type*

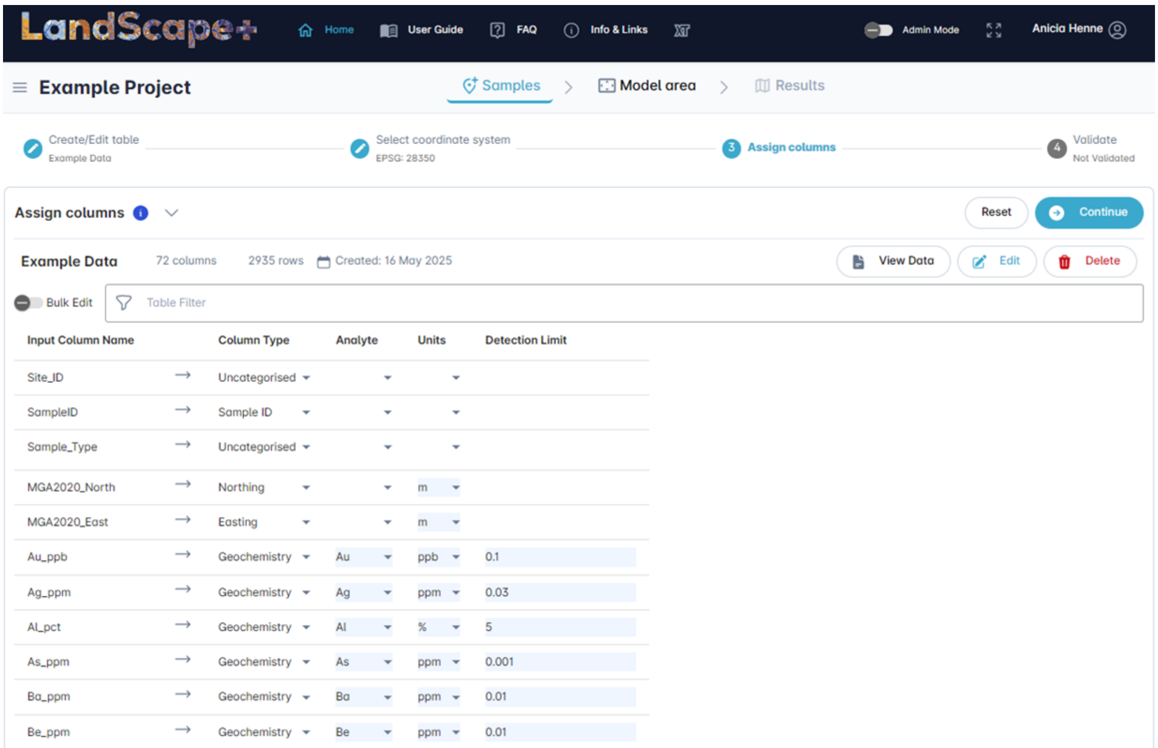
All columns mapped as *Geochemistry* must be assigned an analyte from the drop-down menu. Check all *Geochemistry* columns are mapped correctly. Change or map any missing analytes using the drop-down menu.

✓ *Assign Units*

Select *Units* for all columns except for "Sample ID", "Uncategorised" and "Soil properties". All *Analytes*, *Easting* and *Northing* must have a unit assigned to proceed. As long as the correct *Units* for each analyte are entered, there is no need to convert your values. The application will automatically convert all values in the background to ppm during the validation step.

✓ *Assign Detection Limits*

The sample upload requires you to enter the lowest *originally reported* detection limits for each element. This is used during the validation step so the software can flag when values have not been replaced with ½ the detection limit. It is an additional safeguard to help the end-user check their data is suitable for statistical outlier definition. The validation step relies on the user to enter the correct data.



✓ *Check data and continue*

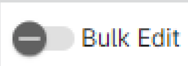
Ensure that all required fields (shaded light blue) are filled.

If you notice at this point that you do not have all necessary columns or that your table settings were incorrect, you can click on the *Edit button* to return to the first step.

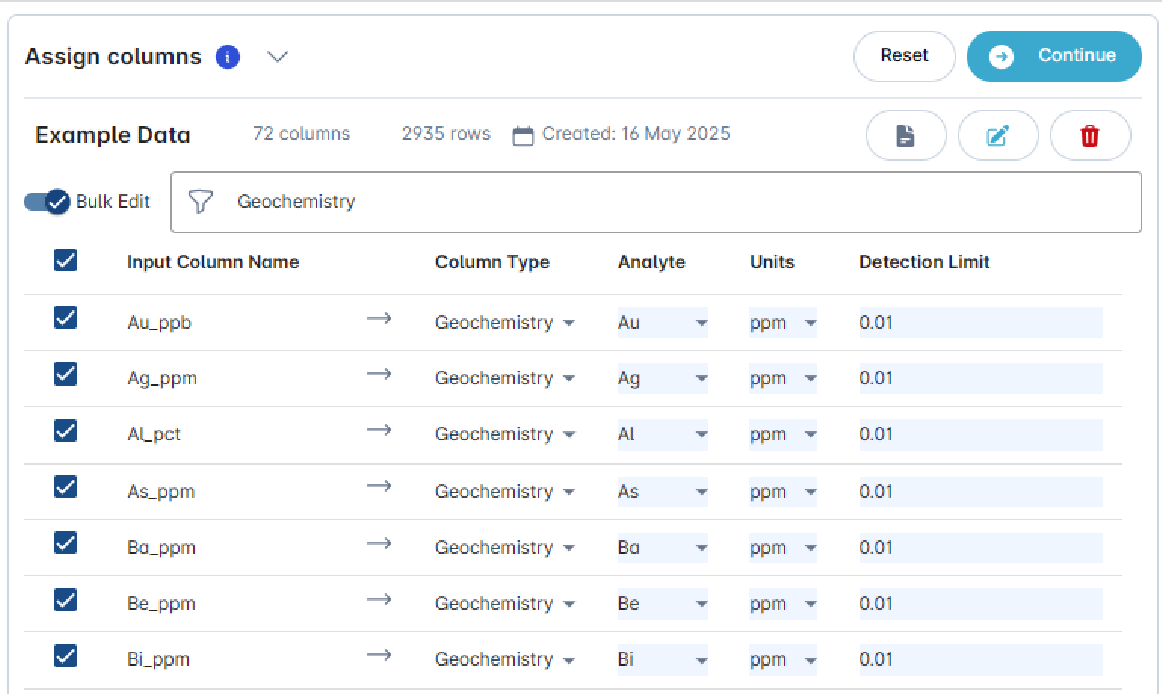
You can also delete the file if it is unsalvageable by clicking the *Delete button*.

The *View Data* button shows you a current preview of the table.

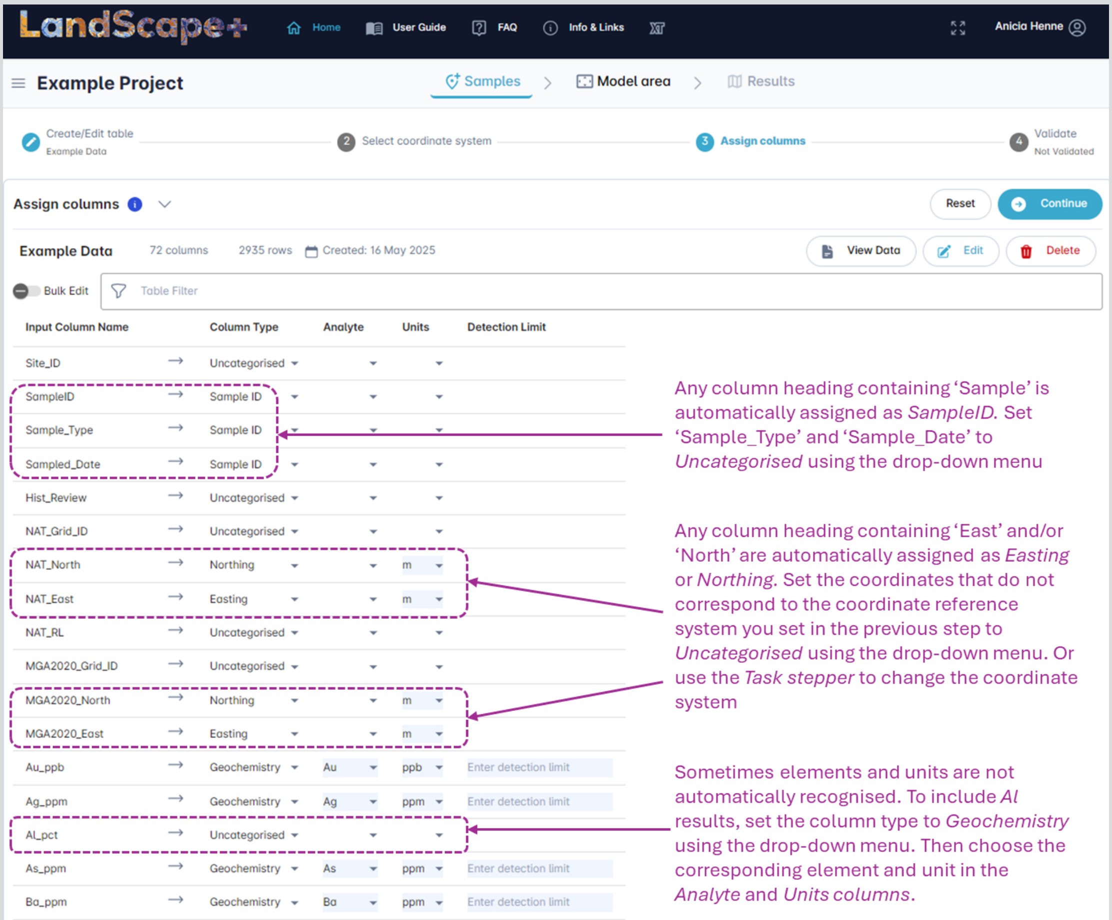
Once you have checked that all information has been mapped correctly, click on the *Continue button* to proceed to the table validation step.



Turn on ***Bulk Edit*** to edit multiple attributes at once and use the ***Filter*** function to select e.g., the column type *Geochemistry* only. Be mindful that all changes will apply to all line items that are ticked on the left-hand side (e.g., in this example all Units have been changed to ppm and all detection limits to 0.01. You can hit the ***Reset*** button (top right) to return all settings back to the default.



Examples of common data cleaning steps



1.5 Validate data

The table validation is an automated process triggered by clicking the *Continue button* in the previous step. The validation process is designed to aid the user to identify some erroneous or missing data points. This depends on the user mapping the correct data in the previous steps. Some validation errors can be ignored at the users' own risk. Others must be corrected to continue.

A *Table validation pop-up dialogue* will indicate whether the table has passed or failed validation.

✓ *Review action log/correct data*

Table is invalid

If the table has not passed validation, the *Table validation pop-up dialogue* will contain a red cross. A message will detail the error. You cannot proceed until the errors are corrected.

Close the dialogue box to edit the table mapping.

Example – Invalid Table

In the example below, the software has identified 366 samples in the uploaded data that are missing coordinates. In this specific case you will have to return to your original file and add the missing coordinates or remove any samples without coordinates.

Then click on the ***Edit*** button to upload the new version of your data. Proceed with Step 1.2.

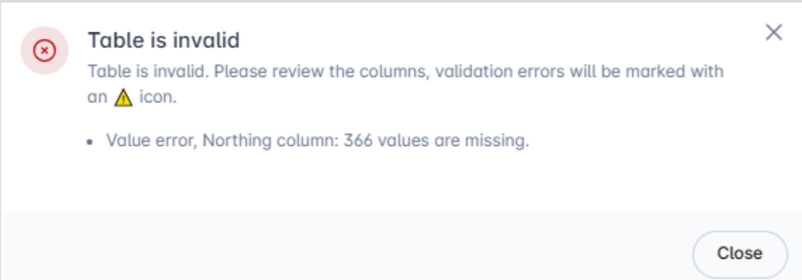
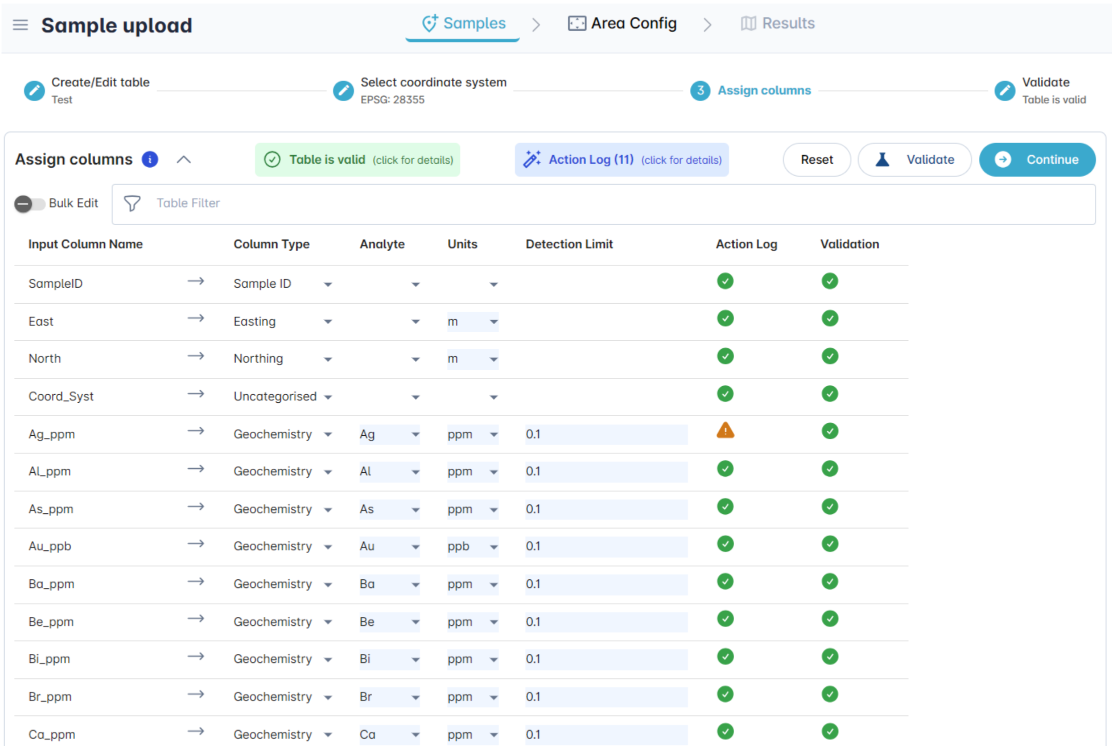


Table is valid

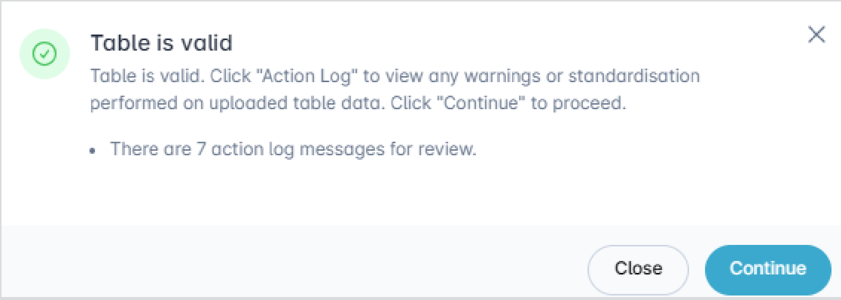
If the table has passed validation, the *Table validation pop-up dialogue* will contain a green tick. If the message refers to an *Action log*, close the dialogue box, to click on the *Action log*. A pop-up dialogue will detail any potential errors for your review. You can also hover over the orange *Attention icons* for potential errors that affect a specific row.



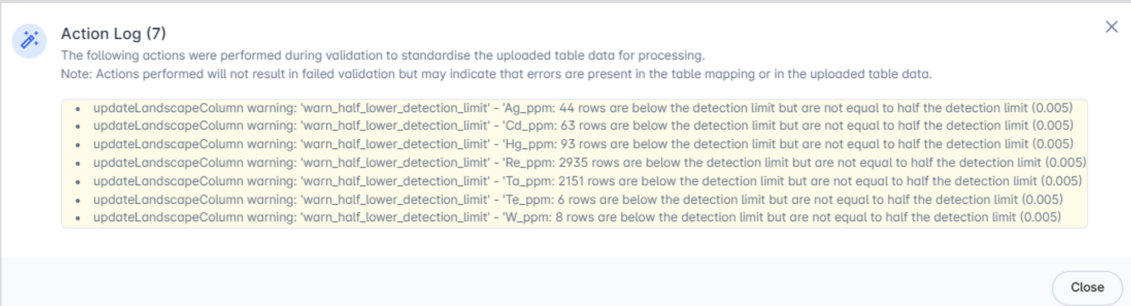
Input Column Name	Column Type	Analyte	Units	Detection Limit	Action Log	Validation
SampleID	Sample ID					
East	Easting		m			
North	Northing		m			
Coord_Syst	Uncategorised					
Ag_ppm	Geochemistry	Ag	ppm	0.1		
Al_ppm	Geochemistry	Al	ppm	0.1		
As_ppm	Geochemistry	As	ppm	0.1		
Au_ppb	Geochemistry	Au	ppb	0.1		
Ba_ppm	Geochemistry	Ba	ppm	0.1		
Be_ppm	Geochemistry	Be	ppm	0.1		
Bi_ppm	Geochemistry	Bi	ppm	0.1		
Br_ppm	Geochemistry	Br	ppm	0.1		
Ca_ppm	Geochemistry	Ca	ppm	0.1		

Example – Valid Table Action Log

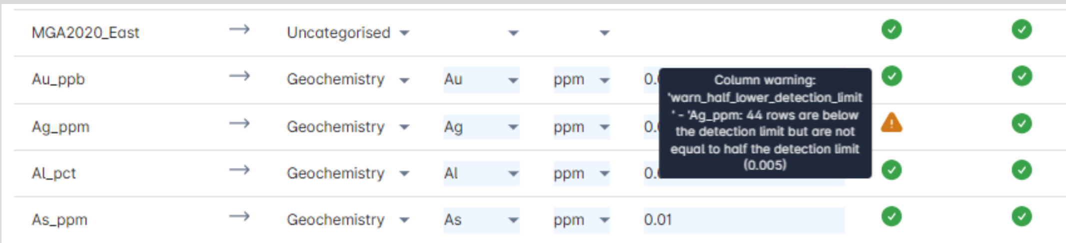
In the below example the table has passed validation, and you are able to proceed to the next step. However, the message notes that there are action log messages for your review. If you click ***Continue***, the log will automatically open. Alternatively, you can ***Close*** the message and click on the ***Action log***.



In this example, the ***Action Log*** describes the identification of flags around detection limits. For 7 elements, values were detected that are below the original detection limit that was entered in the ***Detection Limit column***. However, they are not equal to half the detection limit.



You can also hover over the orange exclamation icons in the ***Action Log column*** to view row specific error logs.



To address this error, check if the original detection limit entered for Ag is correct. If it is incorrect, change it and then click on  (next to the continue button in the main user interface) to repeat the validation. Reiterate this process until no more errors are flagged.

If the original detection limit is correct, check your data for below detection limit values that have not been replaced by half the detection limits. If you have intentionally chosen to use a different imputation method or have uploaded a data set with multiple detection limits for the same analyte, review that you are satisfied with your choice and click ***Continue*** at your own risk.

✓ *Confirm and Continue*

Once you are satisfied with your edits, click the *Continue button*.

Step 2 - Configure landscape model area

Once you have successfully uploaded and validated your samples, they will display in the *Model Area User Interface*. Here you can define the extent of your model area.

There are three main steps:

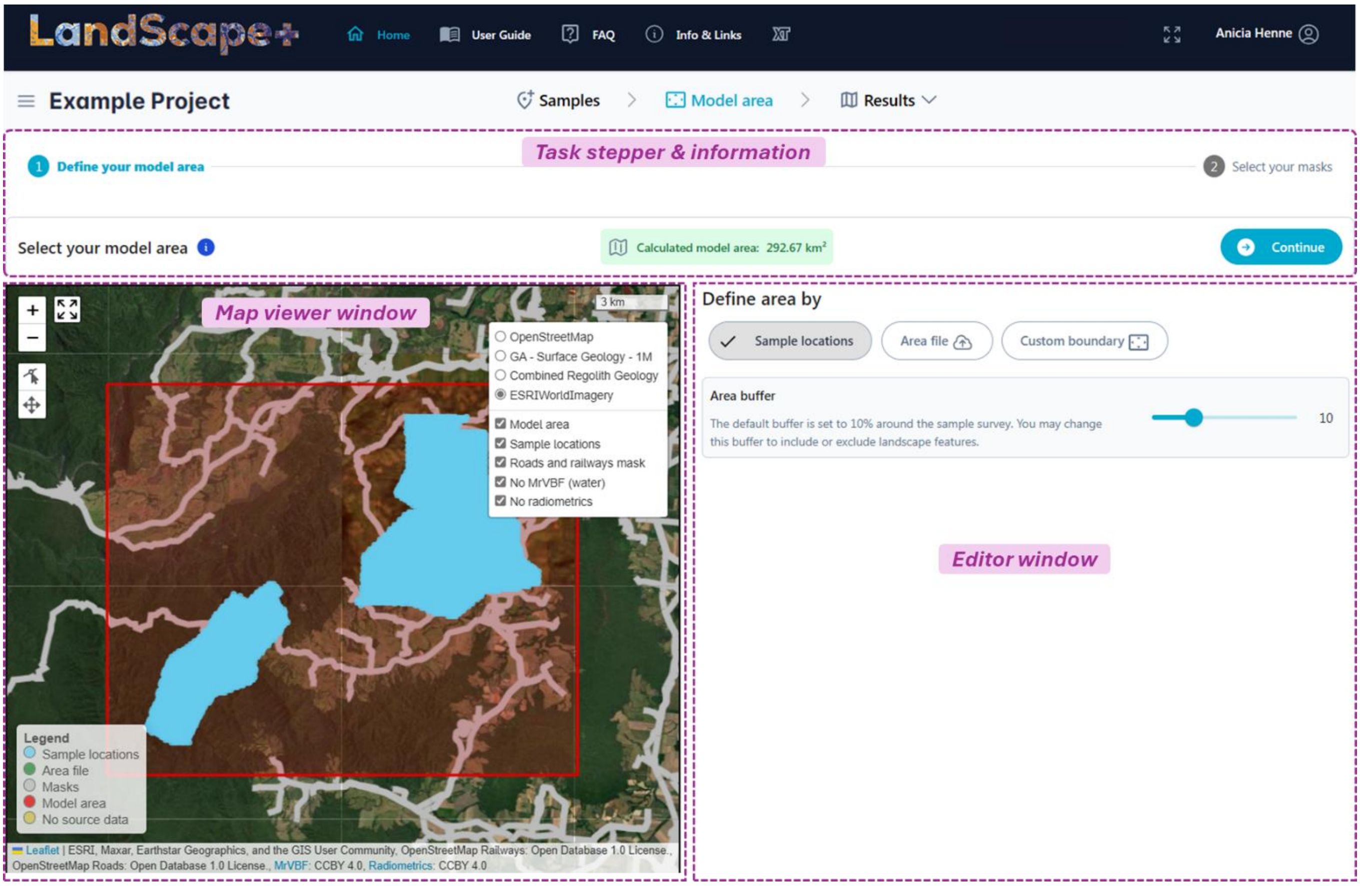
1. Define the model area extent
2. (Optional) Mask out features to exclude from the model area
3. Run LandScape+ model

The software saves your progress when you click on *Action buttons* (e.g., the *Continue button*). – If you (or your internet connection) get interrupted, it will pick up near where you left next time you log in.

User Interface - Model area

The *Model area user interface* contains three areas:

- Task stepper and information window
- Map viewer window
- Editor window



Task stepper and information

Task steppers at the top of the page will prompt you to follow the process for a first-time user. You can return to the previous steps at any time until you run the model. Your settings will only be locked in once you hit "*Run LandScape+*" AND confirm the information message that will pop up. The *Task stepper* is cumulative, so any option from a previous step will still be available to you.

Click on the blue *Information button* next to "Select your model area" to expand the help text. You can close this window using the cross in the top right corner.

The *Model area calculator* shows the calculated model area in km^2. This is automatically updated when you change the extent of the model area or add/remove a mask. This calculated area does not include areas of missing source data or masks. It will remain green if the area is within the enforced bounds (25 km^2 to 2000 km^2) and will turn red if it is too small or too large.

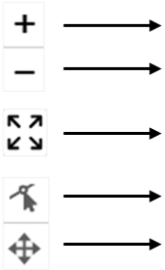
The *Continue button* is on the right-hand side of this window.

Map viewer window

The *Map viewer window* displays spatial data and is interactive (the model area can be edited within this window). It contains *Editing tools* in the top left corner, a *Legend* in the bottom left corner, and a *Spatial data toggle* in the top right corner.

Tools

The following tools are available:



- Zoom in
- Zoom out
- Maximise view to model area
- Edit model area extent
- Pan

Legend

Indicates the colour of different elements.

Legend item	Description
Sample locations	The uploaded samples will display as blue dots
Area file (optional)	Area files (e.g., tenement boundary) you upload will display as a shape with a green boundary. Area files are optional and will be uploaded by the user.
Masks (optional)	All masks (areas that are excluded from the model, either in-built roads and rails and/or the additional file you uploade) will be displayed as grey lines and/or polygons.
Model area	The model area will always be a square polygon around either your samples or your area file and will be outlined with a red border. The polygon will turn semi-transparent yellow when in edit mode.
No source data	Areas where radiometric measurements and/or MrVBF are not available are displayed as yellow polygons. Both source data are needed to run the model, and such areas will be excluded. MrVBF is not available for permanent waterbodies (as per the source SRTM DEM data at the time of generation).

Spatial data toggles

The *Spatial data toggle* allows you to toggle between 4 background layers in the top menu using *radio buttons*. Only one layer can be displayed at a time.

Background layers	Description
OpenStreetMap	Roads and railways sourced from OpenStreetMap
GA – Surface Geology – 1M	Surface Geology of Australia at a 1:1 million scale sourced from Geoscience Australia
Combined Regolith Geology	Regolith geology merged from the Regolith Map of Australia at a 1:5 million scale sourced from Geoscience Australia and the Regolith Geology of Western Australia sourced from the Department of Energy, Mines, Industry Regulation and Safety ; no data is available for TAS, ACT, NSW and VIC
ESRI World Imagery	Aerial imagery sourced from ESRI

The bottom menu contains six overlays that can be displayed or hidden. Hiding these data will not remove them from the model.

Overlays	Description
Model area	The current extent of the model area as defined by you
Sample locations	The locations of the samples as uploaded by you
Roads and railways mask	Roads and railways, sourced from OpenStreetMap, will automatically be masked from models unless deactivated in the *Editor window* (not advised).
No MrVBF (water)	Areas where Multi-resolution Valley Bottom Flatness is not available. Models cannot be run where this data is missing, and such areas will be shaded yellow within the *Map viewer window* and labelled “No source data” in the legend.
No radiometrics	Areas where radiometric data is not available. Models cannot be run where this data is missing, and such areas will be shaded yellow within the Map viewer window and labelled “No source data” in the legend.

Editor window

You can edit the model area in the Editor window. When you move through the *Task stepper*, more options will become available. The process is outlined in detail below. Remember, if in doubt, follow the blue *Continue button*.

2.1 Select model area

The model area is by default defined as a 10 % square buffer around your samples. This is because the landscape model is intended to be used for geochemical outlier calculation by landscape type from your dataset. You may wish to change this default setting, for example to satisfy minimum or maximum area extents or to include/exclude landscape features that are important for your exploration context or future soil sample planning.

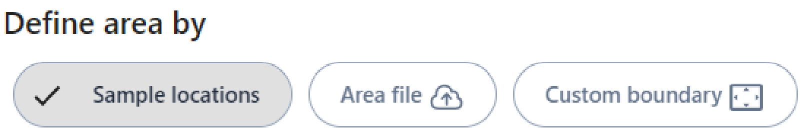
The total model area size must be between 25 km^2 and 2 000 km^2. A box above the area view tracks the area size and displays in green for models within the size range and red when the area size exceeds the size range or is too small. You will not be able to run a model if the box is red.

These limits are provided to guide the end-user to produce useful models with little knowledge input required. The lower limit has been set to prevent as much as reasonably possible, models in areas with little landscape variation where a landscape model is likely not useful for geochemical outlier interpretation.

The upper limit is dependent on the time and computational cost of running sites at a high resolution and the landscape variation that can be reasonably expected to be represented in a useful manner by a maximum of 16 landscape clusters.

✓ *Define model area input*

In the *Editor window*, you can choose between defining your model area by *Sample locations*, *Area file* or *Custom boundary*. You can change these options by clicking on the respective buttons in the *Editor window*. A tick will indicate which option is active.

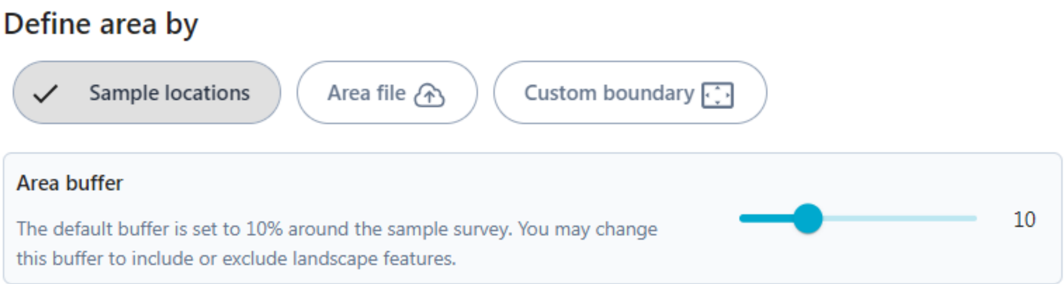


By sample locations (**Default setting**)

The default setting is by *Sample locations* with a 10 % buffer drawn around the samples.

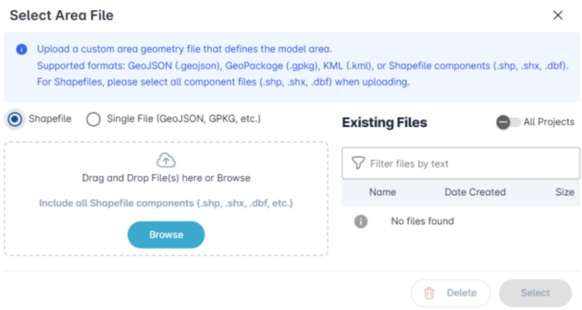
You can use the *Area buffer slider* to extend or reduce the buffer around your samples. This slider allows you to change the model extent to between -25 and 100 % of the survey area. The box will always be centred around the centre point of your survey area.

The model area (red box) in the *Map viewer window* and the *Area calculator* will automatically update.



By area file
You may choose to upload an area file such as a tenement boundary or other boundary you wish to use to define the outline of your model area instead of using the sample locations.

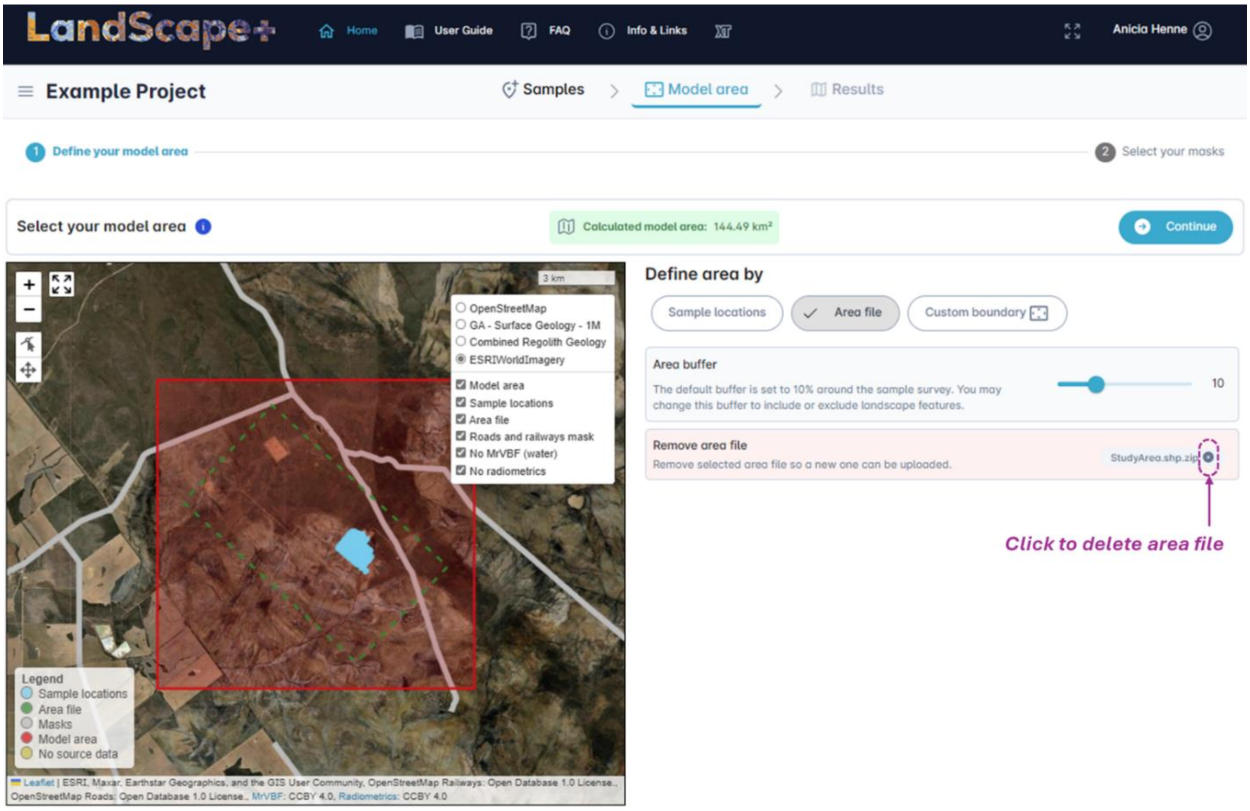
Click on the *Area file button*. A pop-up window will appear. Drag-drop or browse any area files you wish to upload. Note that the file has to be a GeoJSON, GeoPackage (.gpkg) or an ESRI Shapefile. If uploading an ESRI Shapefile, you must include all shapefile components (.shp, .shx, .dbf, and .prj files).



The *Area file* (dashed green polygon) will load into the *Map viewer window* and the *Model area* (red square) will update to fit around the uploaded shape. The *Model area* will always be a square. You can continue to use the *Area buffer* to change the extent of your *Model area*. It will now be centred around the *Area file*.

You can only display **one area file at a time**. If you wish to trial a different area, remove the current one, by clicking the small cross in the *Remove area file box* (see image below).

You can revert the model area extent back to “Sample locations” by clicking on the corresponding button.



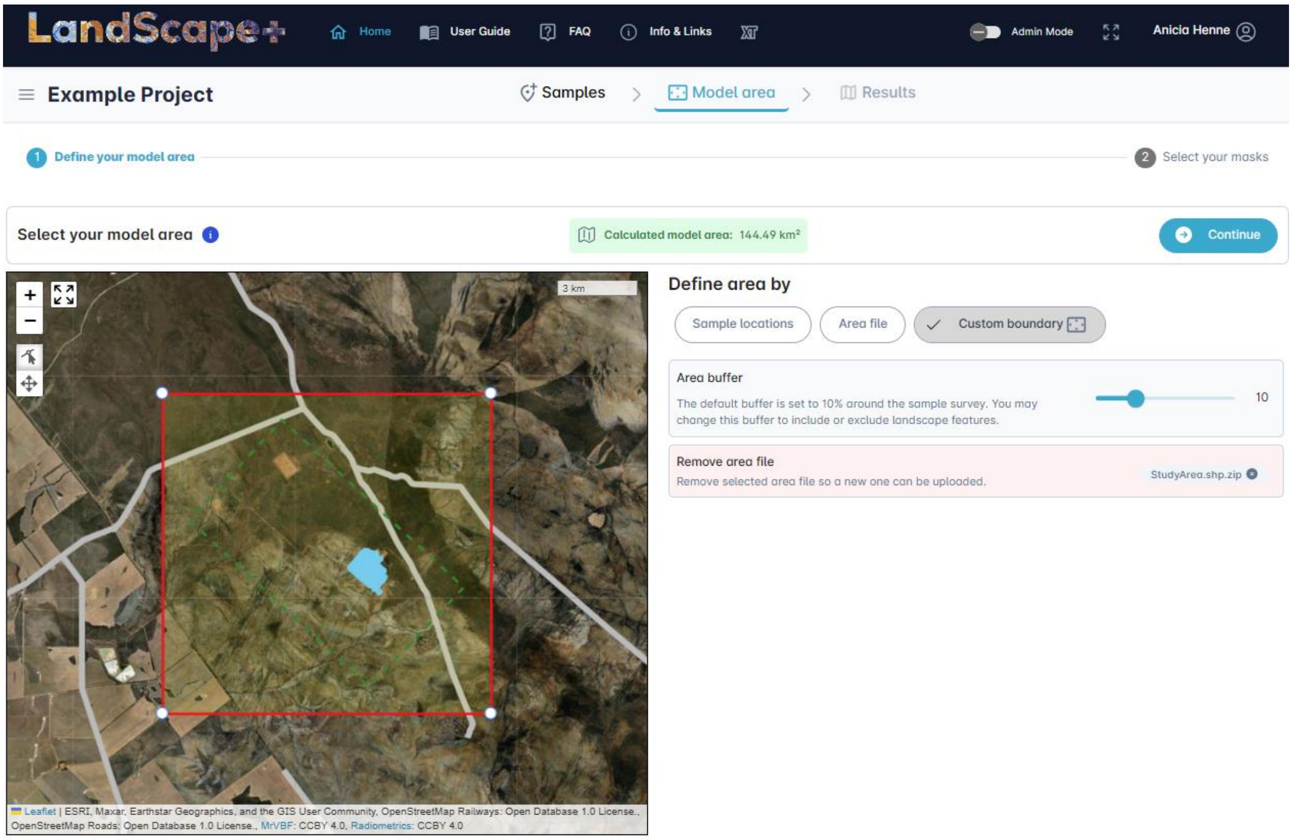
By custom boundary
You can customise the extent of the model area further using the *Editing tools* in the *Map viewer window* and/or by clicking on the *Custom boundary button* in the *Editor window*. The fill of the model area will turn yellow and white circles will appear in each corner of the model area outline box. See *Resize model area* in the next section for further details. The shape of the model area will always be a square box.

✓ *Resize model area*

The model area is displayed in the *Map view window* as a box with a red bold border and transparent red fill. Editing mode is activated by clicking on the model area in the *Map view window* or by clicking on the *Custom Boundary button* in the *Editor window*.

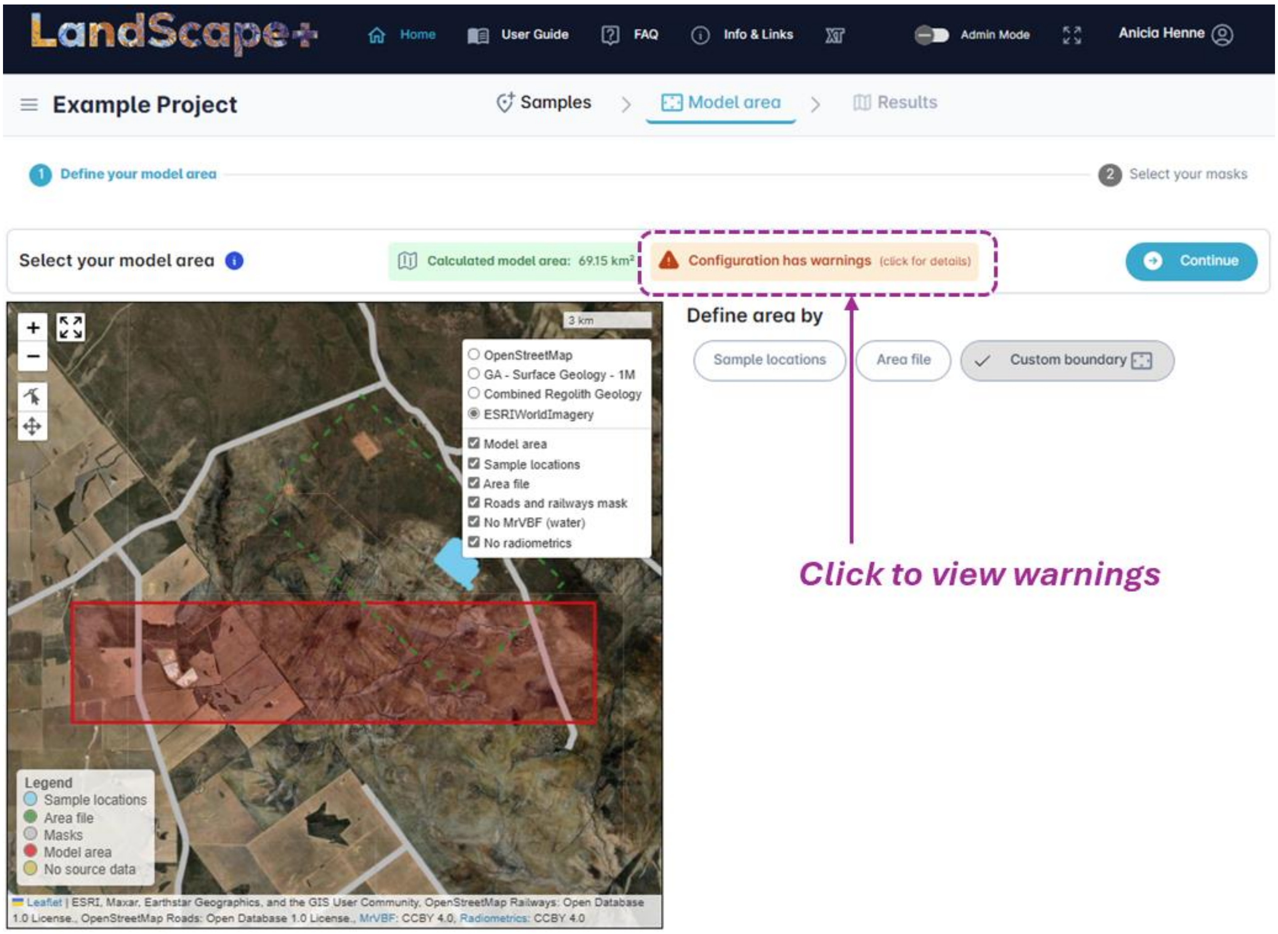
Map view window tools

The *Map view window* contains several functions to change the extent of your model area. A left mouse click will activate the *Edit layer function*. The area will turn semi-transparent yellow and *Round handles* will appear in the corners. Click and drag the handles to resize the box. Release the mouse to lock in the changes. This function can also be activated using the *Edit layer* button in the *Map viewer window*. Or by clicking on the *Custom boundary button* in the *Editor window*.

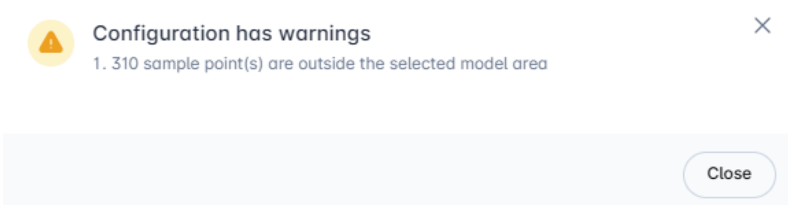


You may use the *Pan function* in the map window to drag your model area to a different location.

Note that any samples not within the model area will be excluded from the outlier definition and Principal Component Analysis. In such cases, an orange warning box will appear.



Click on the box to view details. Change the model area until you are satisfied that it captures the area you wish to model.



No matter the file or custom boundary you choose, the shape of the model will always be a square box.

- ✓ *Continue*

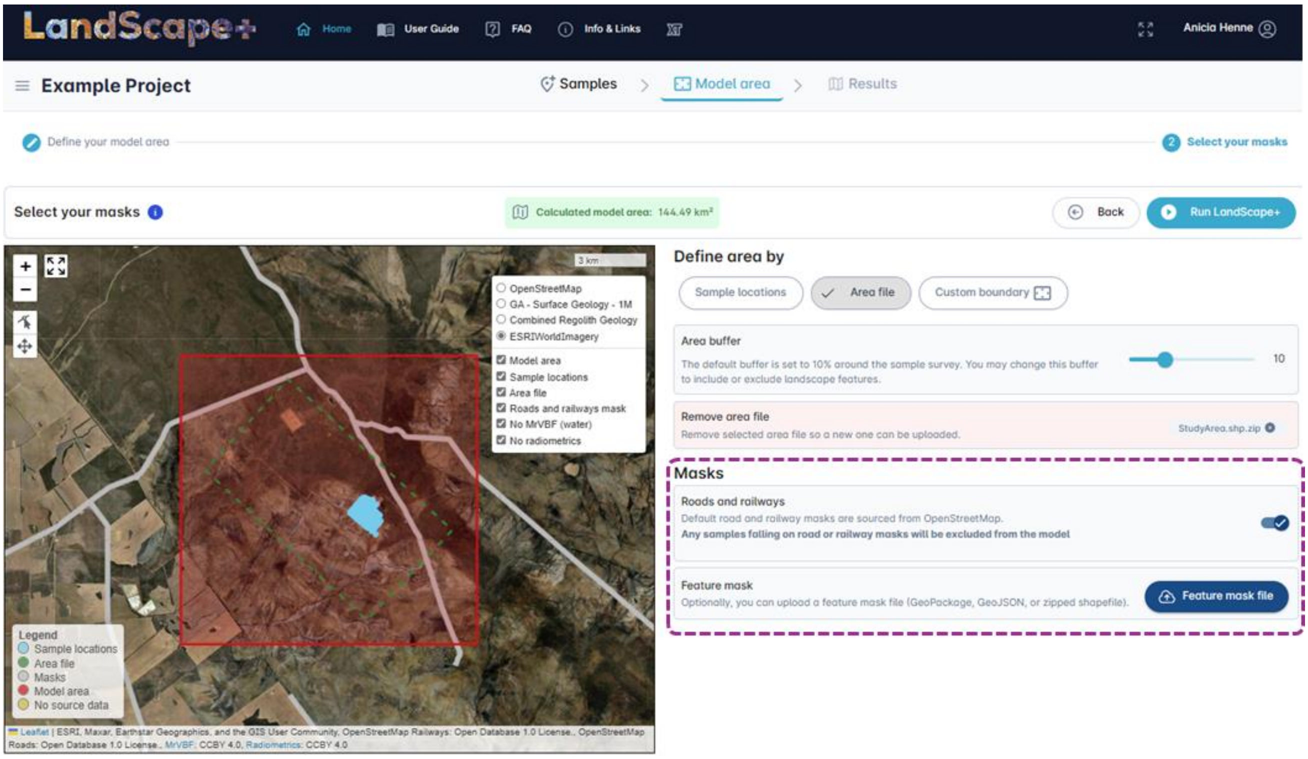
Once you are satisfied with the model extent, click the *Continue button*. The *Task stepper* is cumulative, so you will still be able to access all editing tools on the next page in case you need to change the model area extent after reviewing masked (excluded) areas within the model extent.

When defining a model area, look for major features in the survey area you wish to capture and avoid small areas with few or no samples. For more guidance, see some examples below.

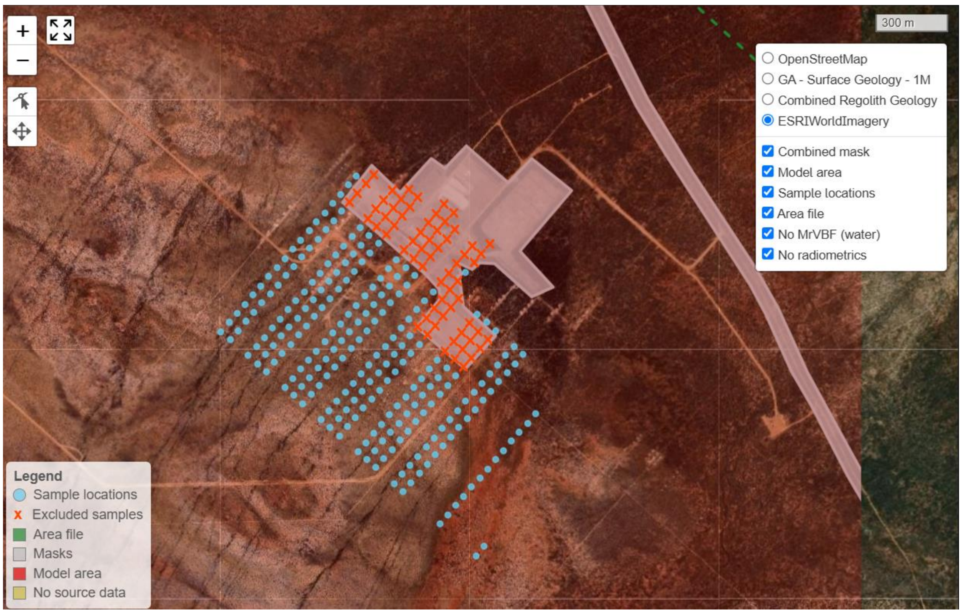
2.2 Select masks

LandScape+® generates models from data collected at the surface which is an expression of regolith and soil forming processes. However, the expression of these processes at the surface, and indeed the geochemical signature in samples collected in affected areas, can be disturbed by human activity. This includes major roads, railways, buildings and other infrastructure such as airports, mine sites, and tailings dams. We recommend masking such features in many cases.

All masks are displayed in grey and masked areas are automatically subtracted from the overall area size calculation.



Any **samples that fall into masked areas** will **not** be included in the model and no geochemical outliers or Principal Components will be calculated for excluded samples. Such samples will display as red crosses.



No source data

The input, or source, data for LandScape+® includes radiometric data and MrVBF (Multi-resolution Valley Bottom Flatness index). The radiometric data of Australia does not cover 100% of the Australian continent and where this is missing, no model can be produced. The MrVBF excludes permanent water bodies (as per when the product was generated from SRTM DEM data; it may not reflect all current water bodies correctly), and such areas will also not be included in the model. This cannot be changed. However, most rivers and temporary water bodies are not excluded. The missing area of such source data is indicated in yellow in the *Map viewer window*.

✓ *Roads and railways*

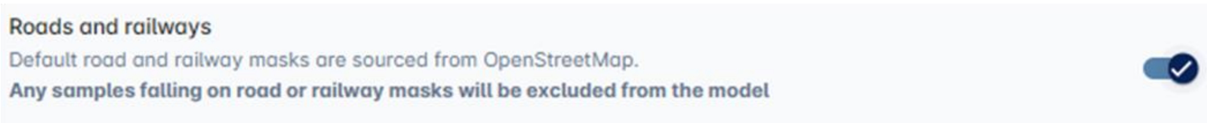
Default road and railway masks are provided and sourced from publicly available data (OpenStreetMap). Roads and railways are masked (i.e., the pixel that touches the road will not be included in the model) which is also generally a suitable approximate distance of sampling away from major roads due to the impact of contamination on soil samples.

The *Default setting* is that this mask is **ON** (roads and railways will be masked from the model area and appear as empty lines).

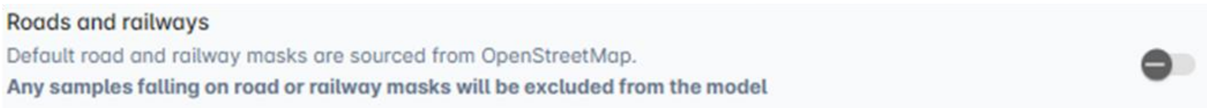
We have provided the option to turn these masks off in case your soil survey follows roads or your model area is heavily traversed. If roads were present at the time of sampling and you have sampled along them, including them in the model will reflect the real-life context of your soil samples. Less frequented roads may have little effect on the model.

To remove the road and railways mask, move the *Roads and railways slider* to the left.

Roads and railways mask turned ON (roads and rails will be masked):

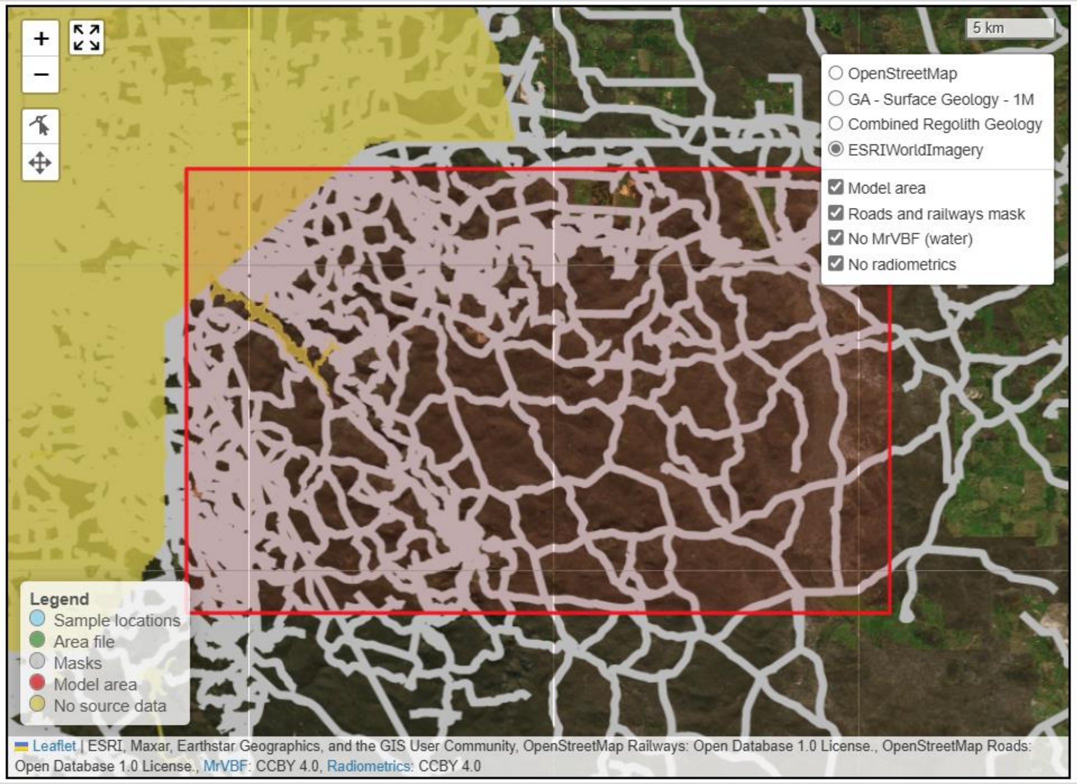


Roads and railways mask turned OFF (roads and rails will NOT be masked)



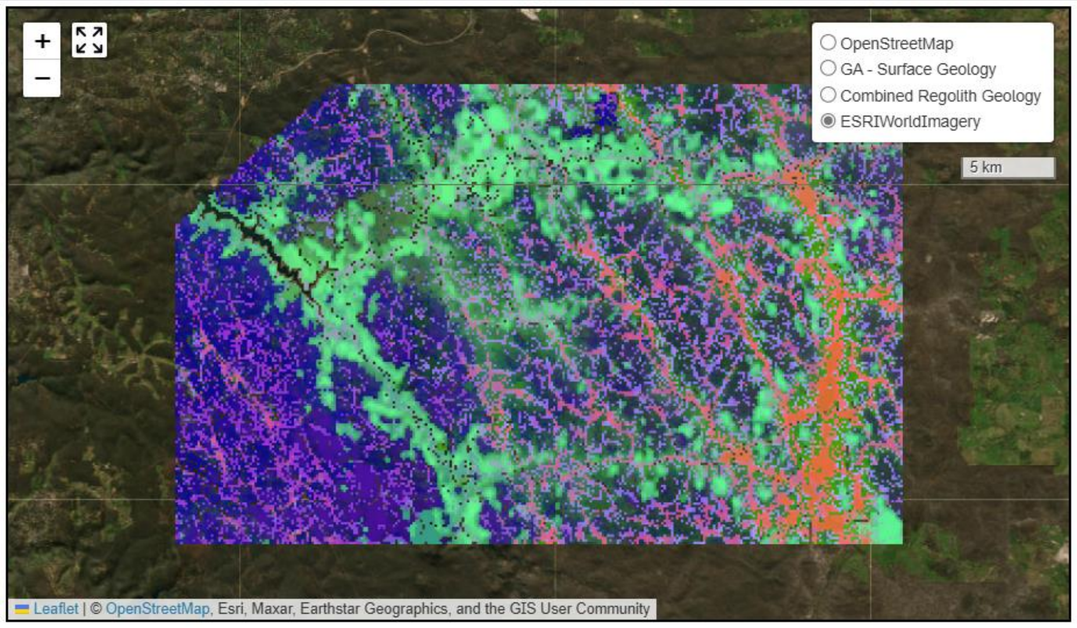
Example

In the below example from the Mundaring state forest in Western Australia, the top left corner of the model area is masked out because no radiometric data is available. A permanent water body (the Mundaring Weir) is also masked out due to missing MrVBF data. An extensive network of roads and 4WD tracks has also been masked. - **Note that the masks are intentionally bolded in the user interface for visualisation – only the pixels touching roads are excluded.**



The resulting model below has no infill for pixels in masked areas:

(continued from previous page)

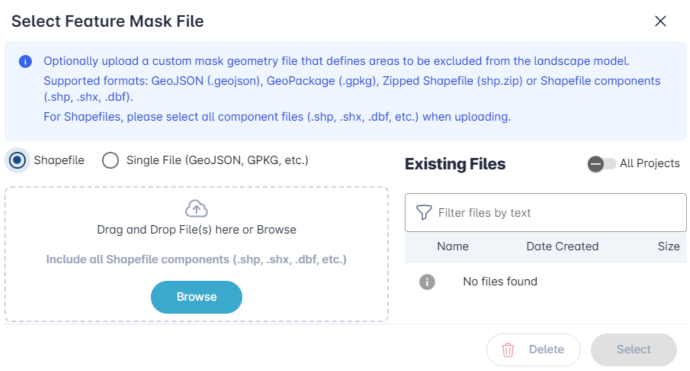


✓ *Feature masks*

You may choose to upload a custom *Feature mask* to exclude additional anthropogenic features, such as rivers, airports, mine sites, tailings dams, etc. from the model area. Mask data files can contain line or polygon geometries. All pixels that touch any mask geometry will be excluded. This is an optional feature and supported file types are GeoPackage (.gpkg), GeoJSON, or ESRI Shapefiles. If uploading an ESRI Shapefile, you must include all shapefile components (.shp, .shx, .dbf, and .prj files).

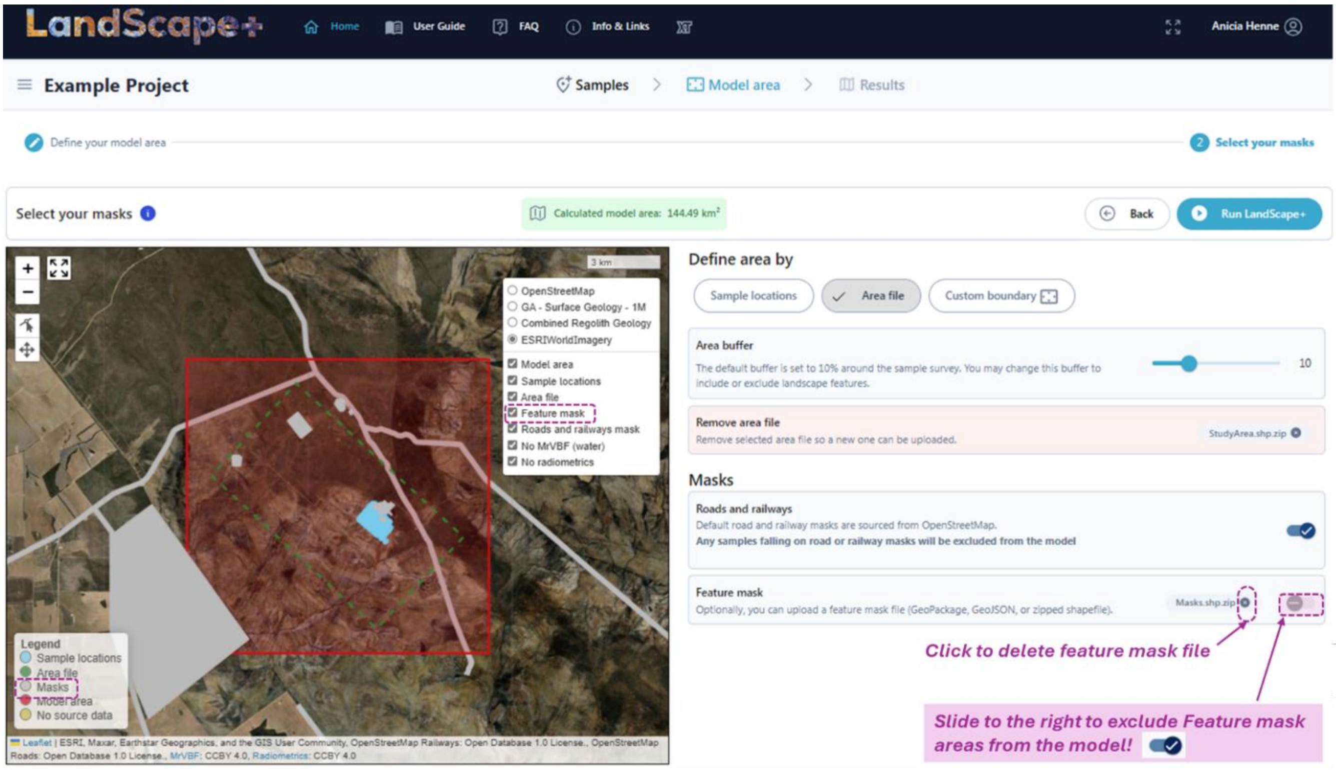
LandScape+ automatically saves your progress, so if you need to interrupt the workflow to draw a feature mask in a GIS software of your choice, you can close the software and return to where you left off at a later date.

Click on the *Feature mask file button* in the *Editor window*. A pop-up window will appear. Drag-drop or browse any feature mask file you wish to upload.

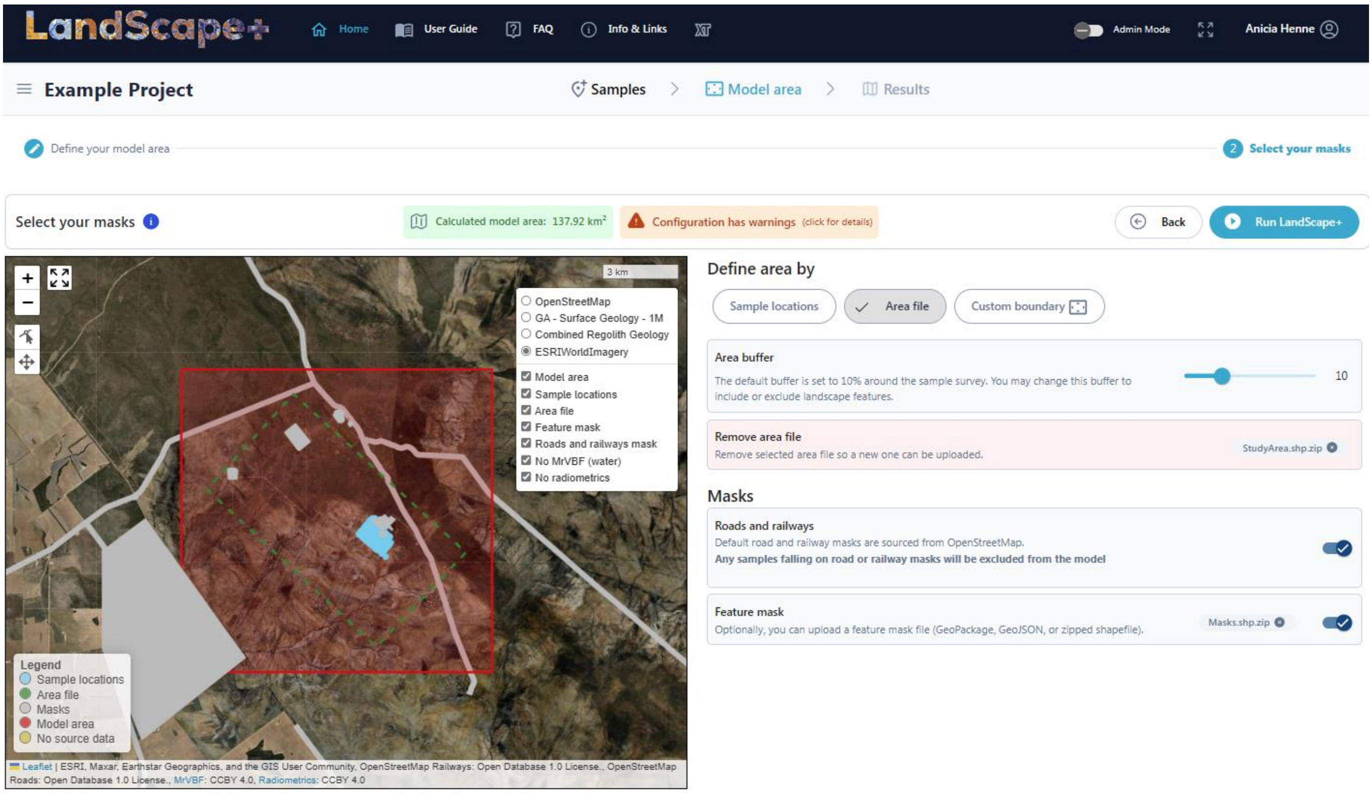


The Feature mask will display in the *Map viewer window* in grey (same colour as the road and railways mask) and has been added to the *Toggle menu*.

You can only upload **one** feature mask file. If you wish to trial a different mask file, remove the current one, by clicking the small cross in the *Feature mask box* (see image below). You **must** use the *Feature mask slider* to include or exclude the masked areas from the model (see image below).

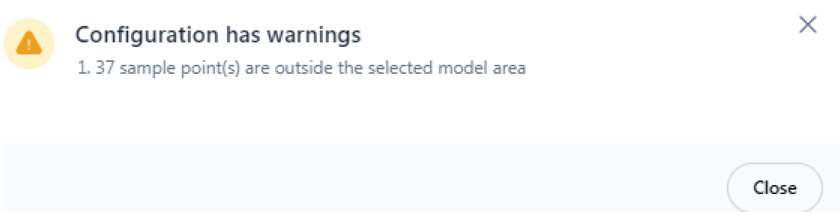


Once the mask is turned on, any samples that fall within masked areas will be excluded. Any excluded samples will display as red crosses. An orange *Warning box* will appear.



Click on the orange *Warning box* to view details:

In this example, 37 samples fall within the masked area and will not be included in outlier definition. You can choose to remove, alter and re-upload the feature mask to include these samples bearing in mind that these samples may have been affected by anthropogenic disturbance. Or you can exclude these samples. There can be value in either option. In this specific example, soil samples were collected over mineralisation. A mine site was subsequently built and is part of the masked features.



It does not matter if your custom *Feature mask* overlaps with an in-built mask or an area of no source data. The model build process automatically combines all masks and areas with missing source data into one.

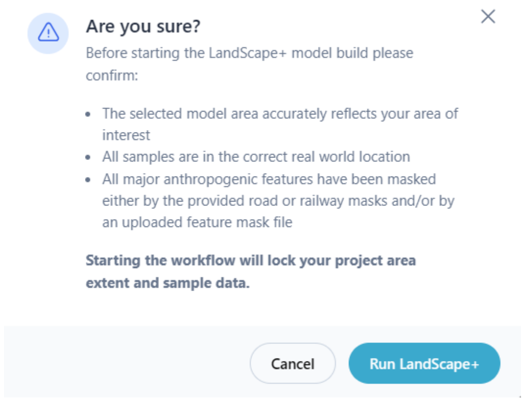
✓ *Continue*

Once you are satisfied with the model extent and masks, the last step is to run the model.

Step 3 - Run model

Once you are satisfied with your area configuration, you can proceed to run the model by pressing the *Run LandScape+ button*.

A pop up will open to ask you to confirm your area configurations.



Click *Cancel* to review your area and/or make changes. See *Examples* in the next section for some additional guidance.

If you are satisfied with your configurations, click *Run LandScape+* in the pop up to start the workflow.

Starting the workflow will lock in the extent of your project area and sample data details.

You will no longer be able to make changes to this model run.

The processing time of your model is dependent on the area size and may take up to 24 hours. Most models will run within 1 to 6 hours. You will receive an *Automated email* with a link to your results, once the model is completed. You can close the browser now. You do not need to stay logged into your LandScape+ account or be connected to the internet.

Once your model is completed and you have reviewed the results, you may return to the Model area configuration settings using the Primary task stepper to make minor adjustments and re-run the model. You can run this model up to a maximum of three times. The results of all three model runs will be available for the duration of your licence (for one year from licence activation).

Examples for picking model areas

In general, the usefulness of modelling any area depends on landscape variation - too little variation and modelling is not needed or appropriate, too much variation and 16 landscape types may not be enough. The usefulness of a model is therefore not necessarily dependent on the area size, although it can be reasonably expected that in many settings larger areas will contain more landscape variation. In addition, the scalability of model areas is also dependent on other factors, namely, input data properties and computing time. The intention of the landscape model is to provide landscape domaining for general mapping and to provide geochemical outlier definition by the mapped landscape types. The number of samples in an area should therefore also be considered. For statistical outlier definition we recommend a minimum of 50 samples be present in every landscape type.

Example - Impact of landscape variation on model area

Homogenous survey areas, e.g., areas of consistent sheetwash without variation, will not produce a useful model. In such instances geochemistry should be reviewed as a whole population.

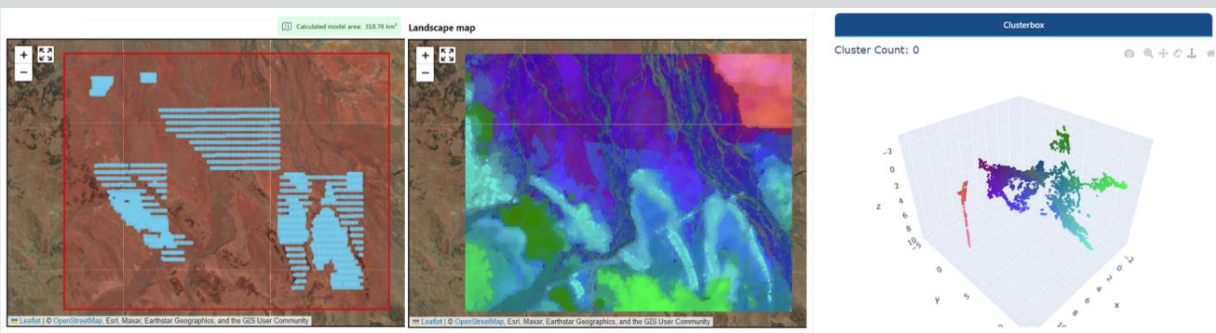
BACKGROUND

The dimensional reduction built into the workflow that creates the models summarises similarities in the input data to group similar pixels together. These groups will be clustered based on major differences to other groups. A prominent ridge in an otherwise flat landscape will likely be highlighted in any number of landscape types. A minor difference in soil properties in a large, strongly varied landscape may not be highlighted.

Example 1 – Landscape variation

A useful model

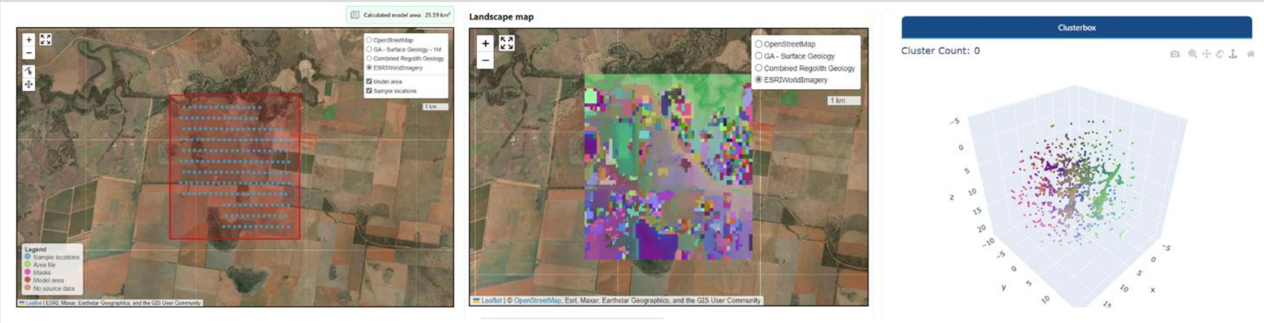
A useful model will cover an area with a minimum of four different landscape types and contain sufficient samples (see Impact of sampling density in the next section). In the example below multiple different features are part of the model area: active and ephemeral channels, sheetwash plains, elevation changes, strong radiometric features and even the surface expression of a fault line. Landscape variation is apparent in the 2D model with congruent features and little pixelation. The clusterbox (pixels of the map in 3D) shows distinct clusters (clouds of points). This is a useful model of regolith context in this area and to divide geochemical datasets into sub-populations for outlier definition.



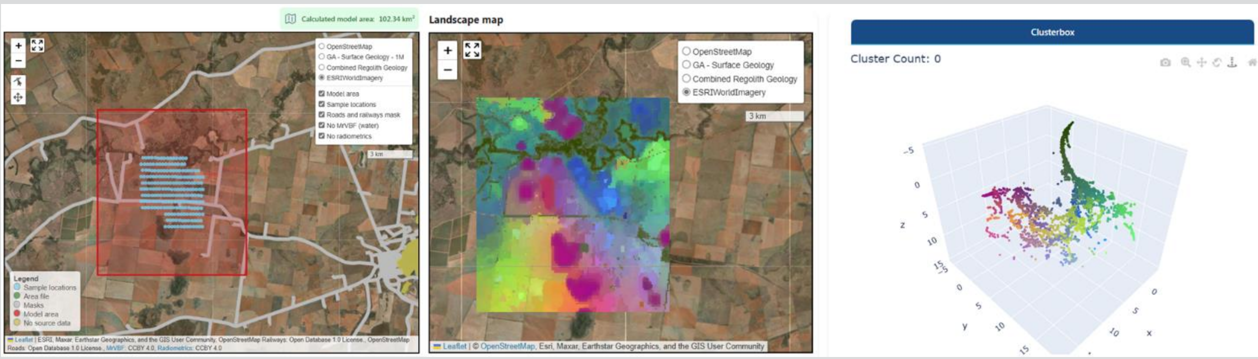
Example 2 – Landscape variation

A less useful model – little landscape variation in a small area

A small area with little landscape variation results in a pixelated (noisy), largely incongruent model in 2D (only a few features are defined such as channels in the north and outcrops in dark purple) with a fuzzy cloud of pixels in 3D and few distinct point clusters.



Extending the model area will, in most cases, include further landscape features/variation and produce a more useful model. However, location and position of samples should be considered (see Impact of sampling density in the next section) as there may be little value in interpreting outliers by sub-populations if most were sampled in the same landscape setting.

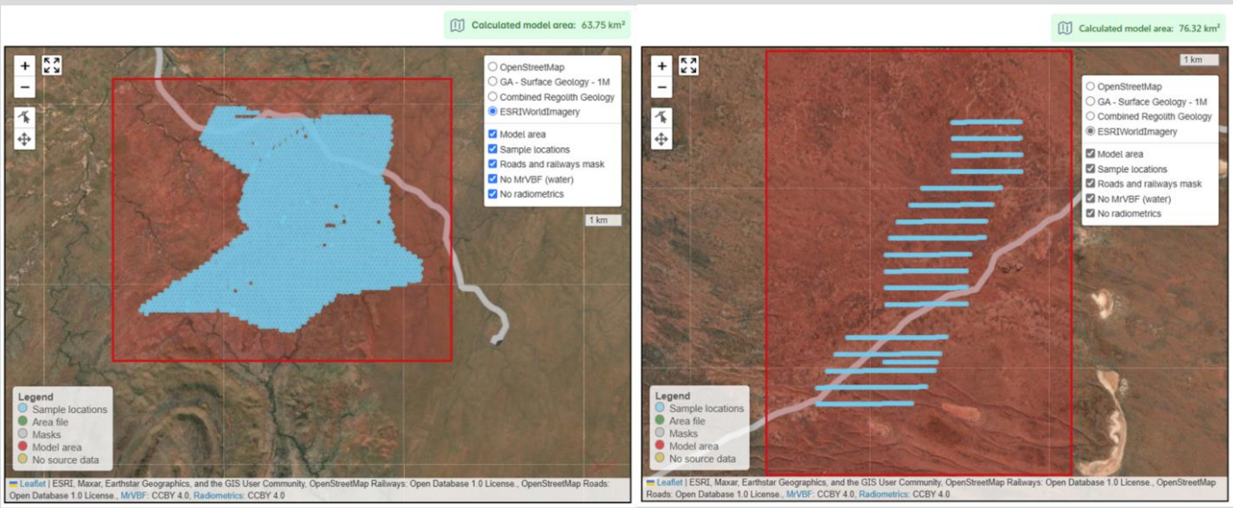


Examples - Impact of sampling density on model area extent

The purpose of the models is to put your soil surveys into context. In general, we recommend a 10 % buffer around a soil sample survey to ensure that landscape clusters are representative of the sample locations. For small areas, we recommend a larger buffer. The extent of your model area will affect statistical outlier definition due to the inclusion or exclusion of surrounding landscape features. For robust outlier definition, we recommend a minimum of 50 samples per landscape cluster.

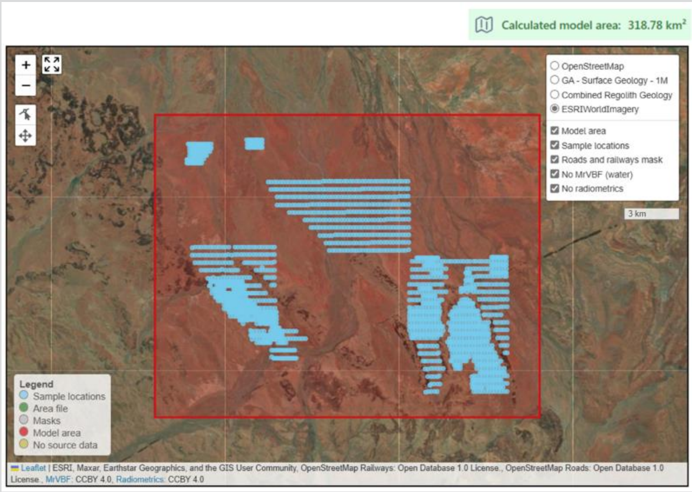
Example 1 – Appropriate sampling density

The two examples below show an appropriate number and spread of samples over varied landscape types and similar area sizes. In both example a buffer of 20-25 % was chosen. The left-hand example contains ≈2500 samples, the right-hand example ≈800. Both areas will produce useful landscape models. Larger sample numbers will provide more robust identification of outliers.



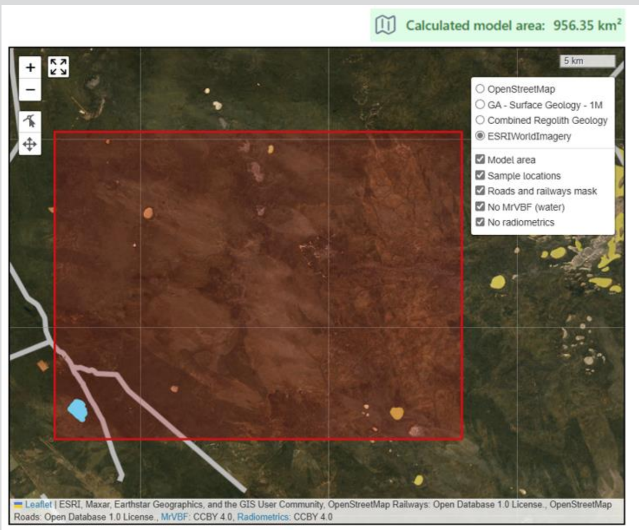
Example 2 – Appropriate sample density to compare multiple survey results in a regional area

The next example also shows an appropriate number (≈2000) of samples for the size of the survey area. The landscape variation is also appropriate with multiple prominent features (various channels, outcrops, sheetwash and residual settings). It is likely that a large number of landscape types will be needed to effectively capture the major landscape features. This is an appropriate model extent if the goal is to treat (what looks like) different soil surveys collectively as a single site for comparison.



Example 3 – Inappropriate sample density for outlier definition

The last example is unlikely to produce robust or useful outliers for geochemical elements by landscape type. The area is large and landscape maps produced from such a model will have value as a broad-scale landscape map. However, the entire ≈200 samples of the soil survey (in the SW corner of the model area) are located in what will likely be mapped as one single landscape type. As a result, the geochemical outliers by landscape type will have little value (they will be the same as looking for outliers within the whole data set, i.e., one population).



Example - Impact of anthropogenic disturbance on model areas

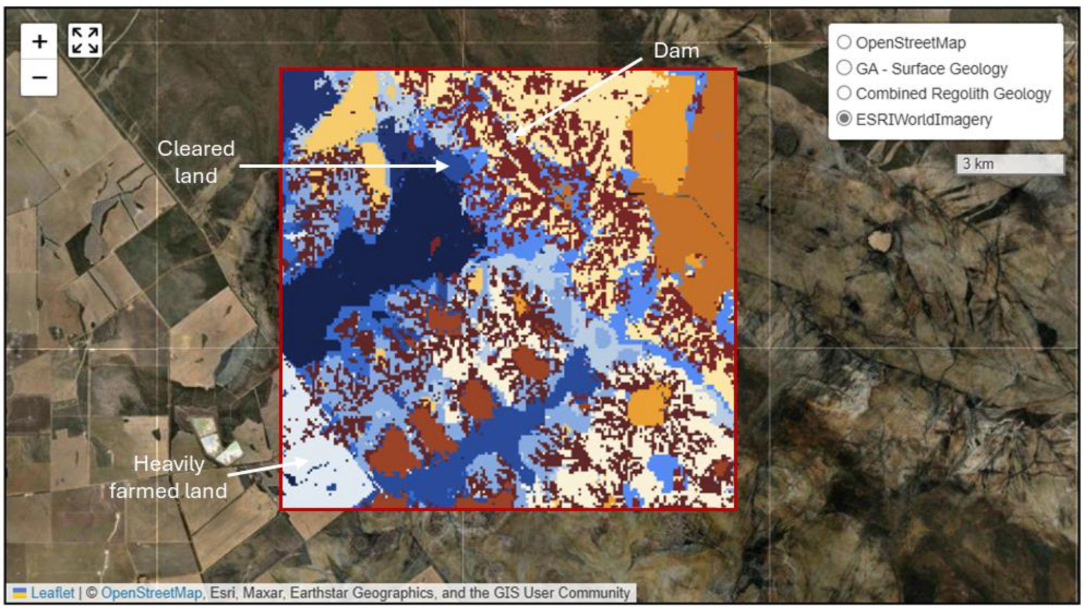
Anthropogenic activities may disturb the natural expression of regolith properties at the surface. For example, in an area with pine plantations that is otherwise undisturbed by human activity the model may assign separate landscape types for these pine plantations. You may choose to mask such areas or not mask them and ignore the subsequently modelled landscape type(s)/feature(s). This may become challenging in a large model if you have an airport, a tailings dam and a pine plantation in a large survey area when only 16 landscape types are available. Your model may assign three landscape types to different anthropogenic features, reducing the number of landscape types available for your regolith materials. In addition, a disturbed area may have been affected in a way that mimics other landscape types in your survey area and the model will combine these with such areas.

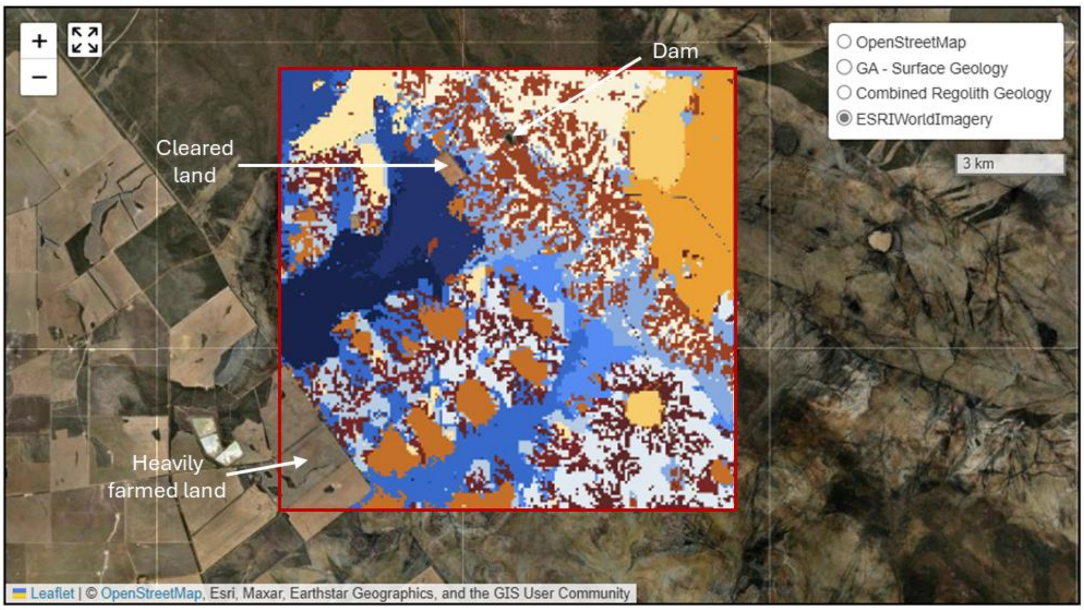
Example 1 – Isolated anthropogenic disturbances

In the below example both, the model with and the model without masks could be used, since no samples were sampled in the affected areas and 16 landscape types were sufficient to capture the variety of landscape types adequately.

In the maps below, whilst the colours of different landscape types have changed, the main landscape features within the model area have been mapped regardless of the presence or absence of masks.

(continued from previous page)





Step 4 – Review model outputs

Once your results are ready, you will receive an *Automated email* from no-reply@exploration.tools with a link to the *Results page*.

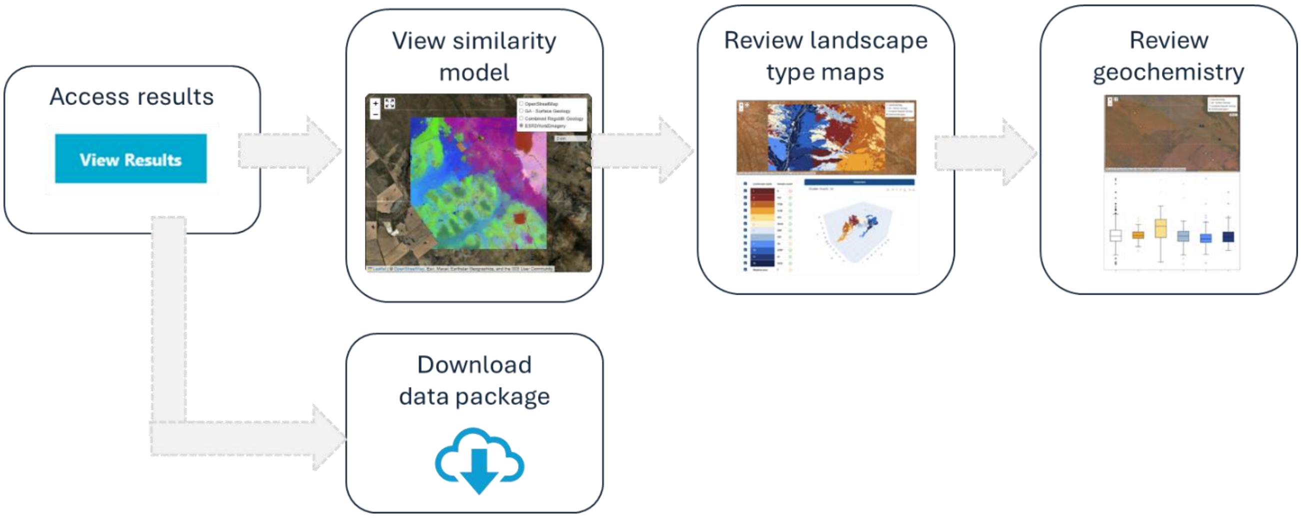
There are three main steps:

1. Review similarity model to familiarise yourself with what influences the landscape types in your model area
2. Review landscape maps with different amounts of landscape types
3. Review geochemistry and outliers by landscape type

We recommend that you review the results in the online *User interface* to decide which model is the best fit for your area. The interface will guide you through the decision-making process.

However, you can download all 12 landscape models as a *Data Package* at any time to use the outputs in a GIS software of your choice. Simply skip all model review steps in the user interface and go straight to download. **The content of the Data Package download will not be affected by any actions you take in the user interface** (except for the delete/remove entire project/results - which is not easy to do and has multiple warnings).

You can return to the user interface to review the model at any time for the duration of your licence (for one year after activation of the licence). ***Make sure you download the results before your licence expires.***



4.1 Access results

Click on the *Link* in the *Automated email* or simply log into your *LandScape+ account* and click on your *Project*. When your model is completed, it will automatically open to the *Results page*.

The results include a similarity model and 12 landscape models for the same area. Each model includes a landscape map with different numbers of landscape types (4-16) and the landscape models include outliers by landscape type for each element based on each map.

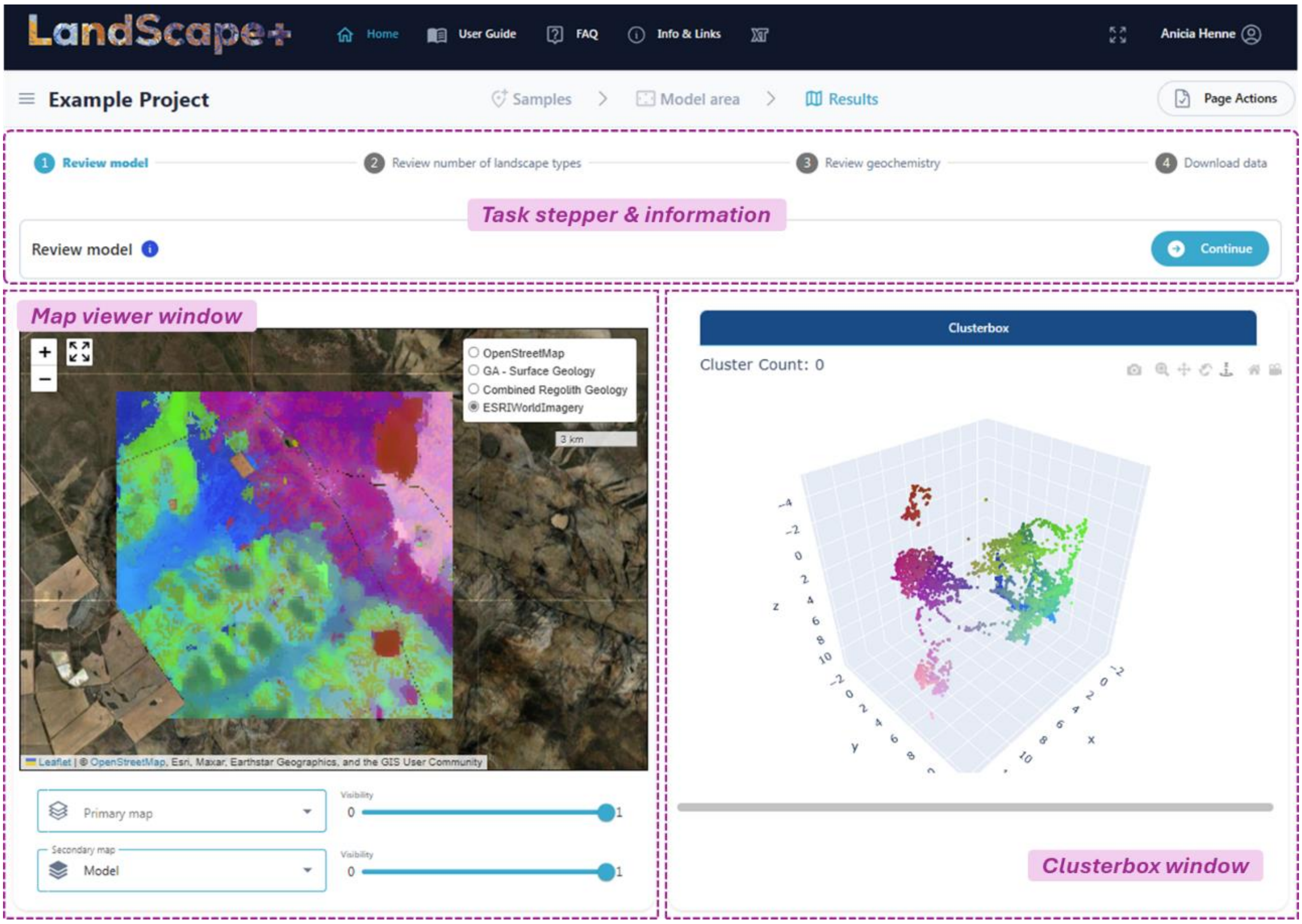
Once you access the *User interface*, a pop-up dialogue box will appear with some important information. Read it while your model is loading in the background.

4.2 Review similarity model

The first content you will be able to view in the user interface is the similarity model (the landscape model coloured in RGB based on UMAP which is used to reduce the spatial dimensions – see *Science Background* *box* below if you want to know more). This is the basis for your model and you can view this in 2D in the *Map viewer window* and in 3D in the *Clusterbox*.

This output is not categorised, i.e., there is no hard boundary between pixels that would combine a group as one landscape type. This step is to familiarise yourself with the underlying spatial model before you view the landscape maps, and the *User interface* functionalities.

The *Task stepper* is accumulative so you can skip ahead at any time and also come back to this step later.



Background

LandScape+ uses Uniform Manifold Approximation and Projection (UMAP), a dimensionality reduction algorithm, to transform the spatial input data for each project area to a three-dimensional representation of data. The method does not explicitly include any location information, spatial relationships, or spatial features (e.g., textures) as only the per-pixel values of each input layer are considered.

After UMAP has been applied to the input data, an agglomerative hierarchical clustering algorithm is used to group locations with similar data signatures. You can find more information on these algorithms here.

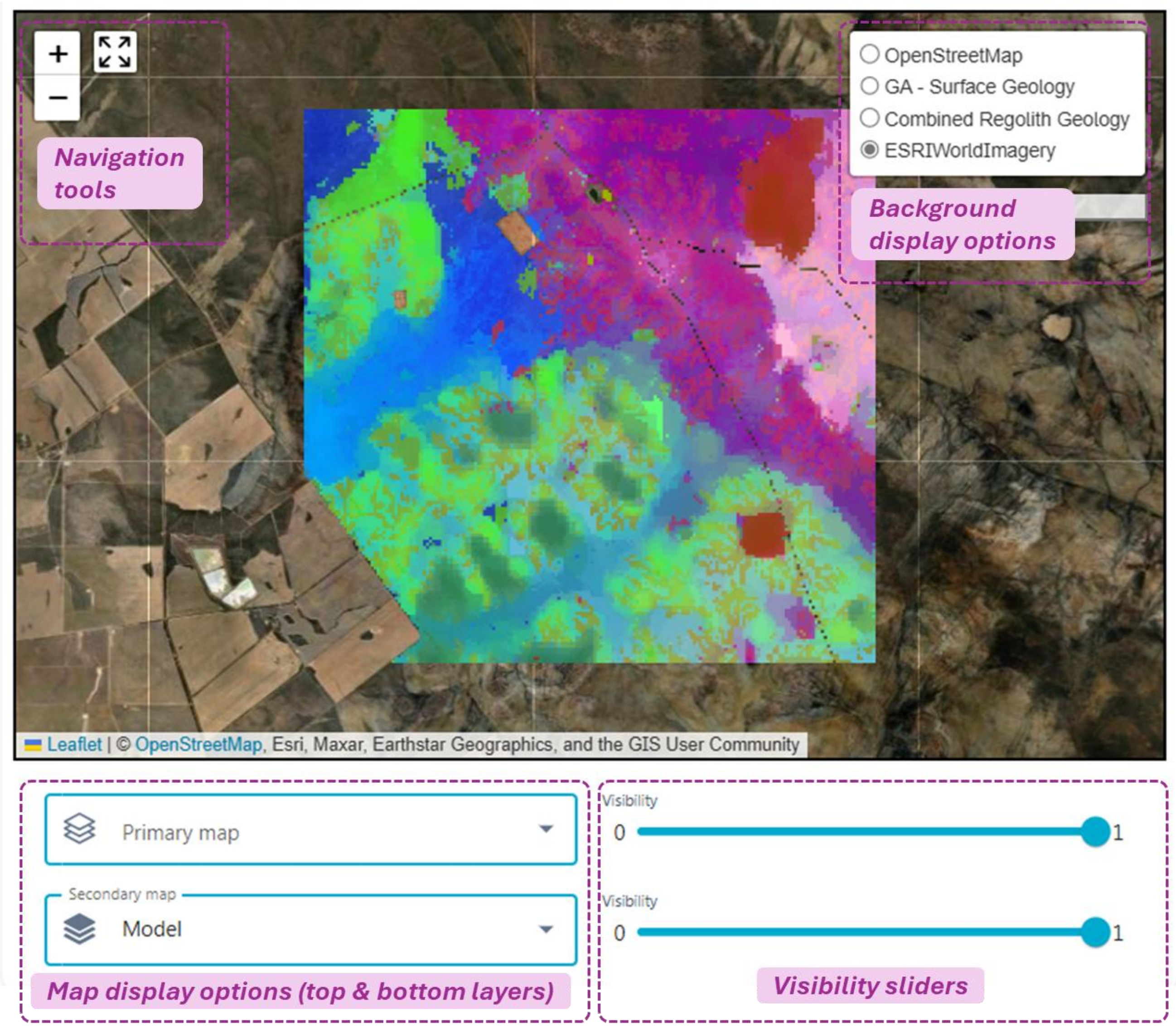
✓ *Compare input layers*

In the *Map viewer window*, select and compare the similarity model map to the different model input layers that have been used to create the model. Using the *Map display options* for the top and bottom layer, you can view two maps at the same time (e.g., the similarity model map and one input layer).

Select a layer in the *Primary map drop-down* and/or *Secondary map drop-down*. Options are the similarity model, Digital Elevation Model (DEM), Multi-resolution Valley Bottom Flatness (MrVBF), radiometric data, and Sentinel-2 ratios. You can choose either of these layers as both the *Primary* and *Secondary map*. Use the *Visibility sliders* to control the transparency of each map layer.

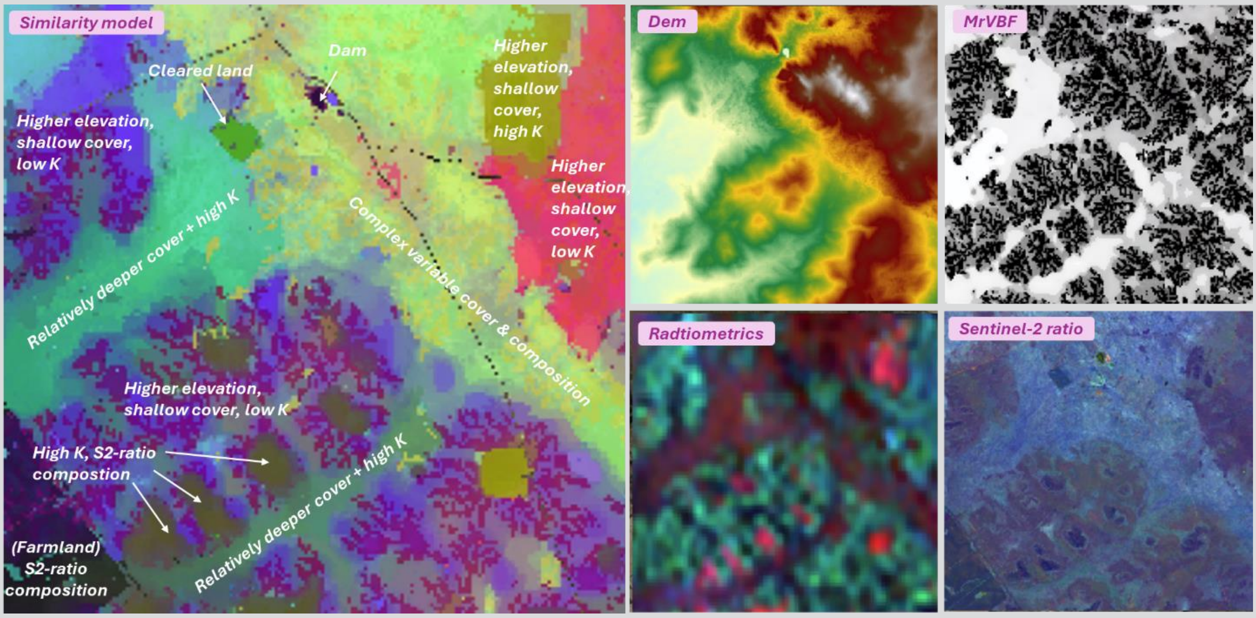
Use the *Navigation tools*, to *Zoom in*, *Zoom out* or return to the *Model extent*.

Change the *Background display* by using the *Radio buttons* in the *Background display options*. Available spatial layers are OpenStreetMap, Surface Geology, *Combined regolith geology (*a combination of GSWA and GA regolith geology layers) and *Satellite imagery*.



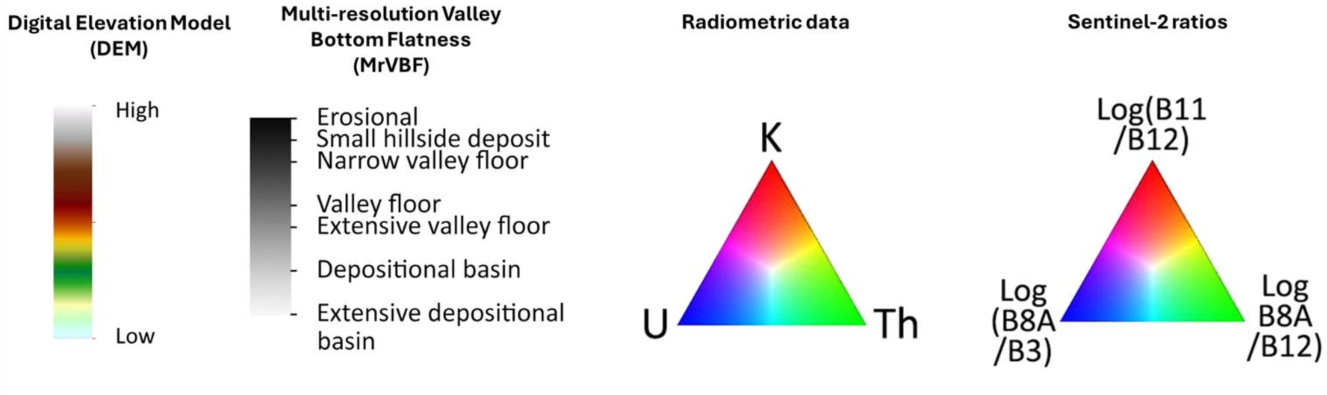
Example

The below example without any masks, shows a rapid overview of the main input features that influence the model. Identifying the main features, will help you decide how many landscape types you might want to choose in the next step.



Handy hint: Place the model at the top (***Primary map***) and the input layers at the bottom (***Secondary map***), then use the ***Visibility sliders*** to fade the model in and out to see which features are influenced by which input layer.

Input layer display legends



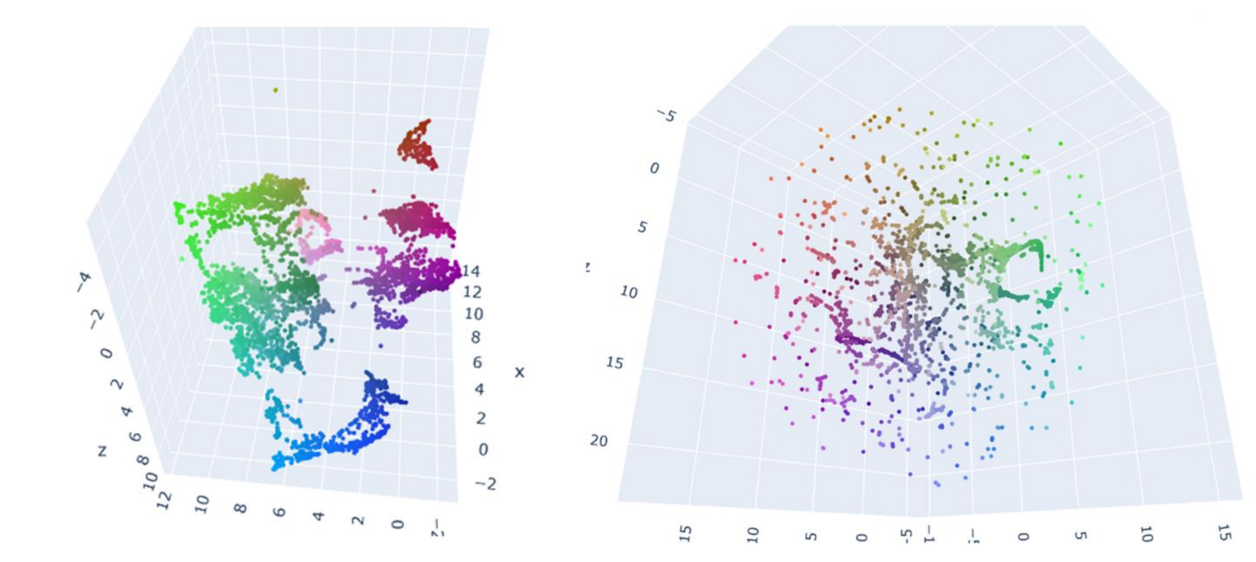
✓ *Review Clusterbox*

The *Clusterbox* is a 3D representation of the pixels on the map. The closer pixels are clustered together the more similar their modelled input layer attributes. If most points are equally close together (forming a uniform “cluster” or “cloud”) there is likely not much landscape variability in the modelled area. Similarly, if you have a dispersed “cloud” of pixels that do not form “clusters”, this also indicates little landscape variation. Points that form distinct clusters, especially offset from other point “clouds” indicate that you have distinct landscape types.

Click and hold your mouse over the *Clusterbox* to rotate and view each cluster in more detail. You can also *Zoom in* and *out*. The colour of each pixel in the *Clusterbox* is the same as its colour on the map.

Example of a useful and a less useful model (Clusterbox)

A useful model displays multiple distinct clusters (left) and indicates distinct landscape types with different properties. A less useful model will display no or few distinct clusters and/or a fuzzy cloud of individual pixels (right) indicating little landscape variation within the chosen model area, or variation that cannot be captured with the model input layers. If this is the case for your model, you may modify your model area to capture greater landscape variation in your next model run.

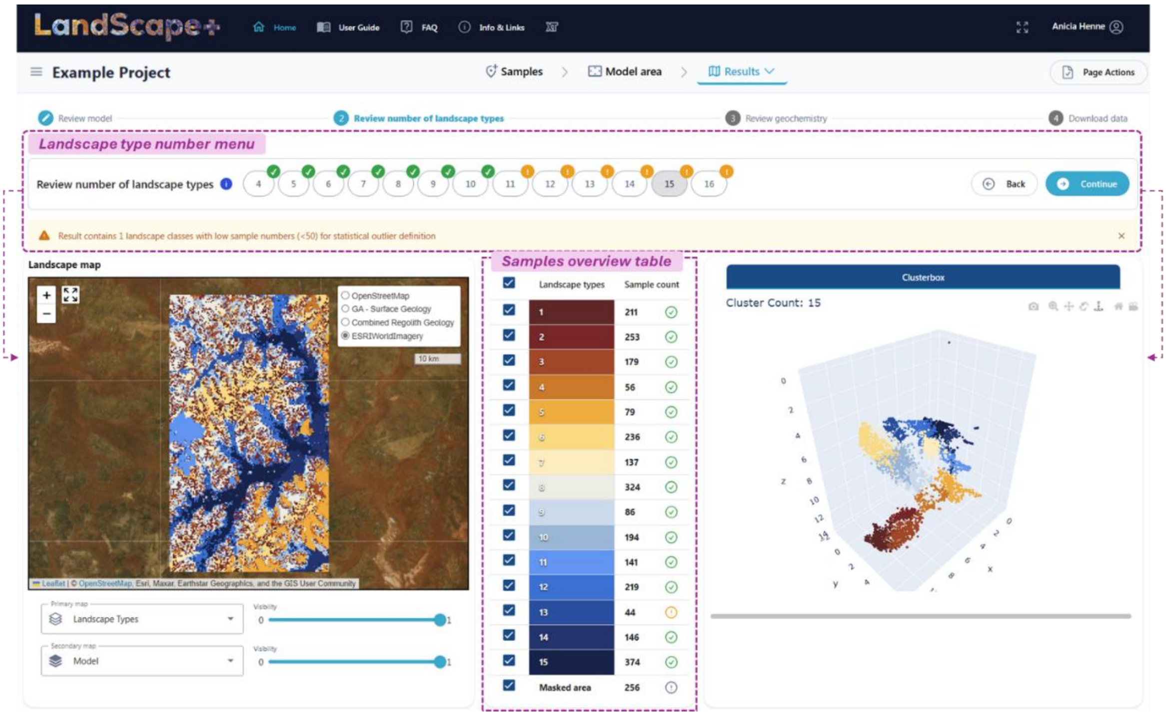


- ✓ *Continue*

Once you have familiarised yourself with the model, click on the *Continue button* or click on *Select number of landscape types* in the *Task stepper* to proceed.

4.3 Review landscape type maps

The *User interface* will now have two additional functionalities. The *Landscape type number menu* underneath the *Task stepper* at the top of your screen and the *Samples overview table*.



The *Landscape type number menu* contains 16 *Landscape number buttons*. The number on each button corresponds to the number of landscape types in the map, i.e., clicking on the button with the number 4 will show you a landscape map with four landscape types. Clicking on one of the buttons will update all content on the page. - Both the pixels in the *Map viewer window* and in the *Clusterbox* will be coloured by the number of landscape types you choose with the *Landscape number buttons*.

✓ Review maps and clusterboxes

Review the available number of landscape types by toggling between the *Landscape number buttons*. The map in the *Map viewer window* and the colours of the pixels in the *Clusterbox* will change accordingly. Ideally each colour represents a distinct cluster of pixels in the *Clusterbox* and a distinct landscape type on the map in the *Map viewer window*.

The model forces pixels into separate groups to satisfy the amount of landscape types you choose. The highest number of landscape types therefore does not always produce the best model.

You will already be familiar with your model from the last step and will have an idea of how many different landscape types you would like to see in your map. The best *Landscape type map* will resemble the similarity model.

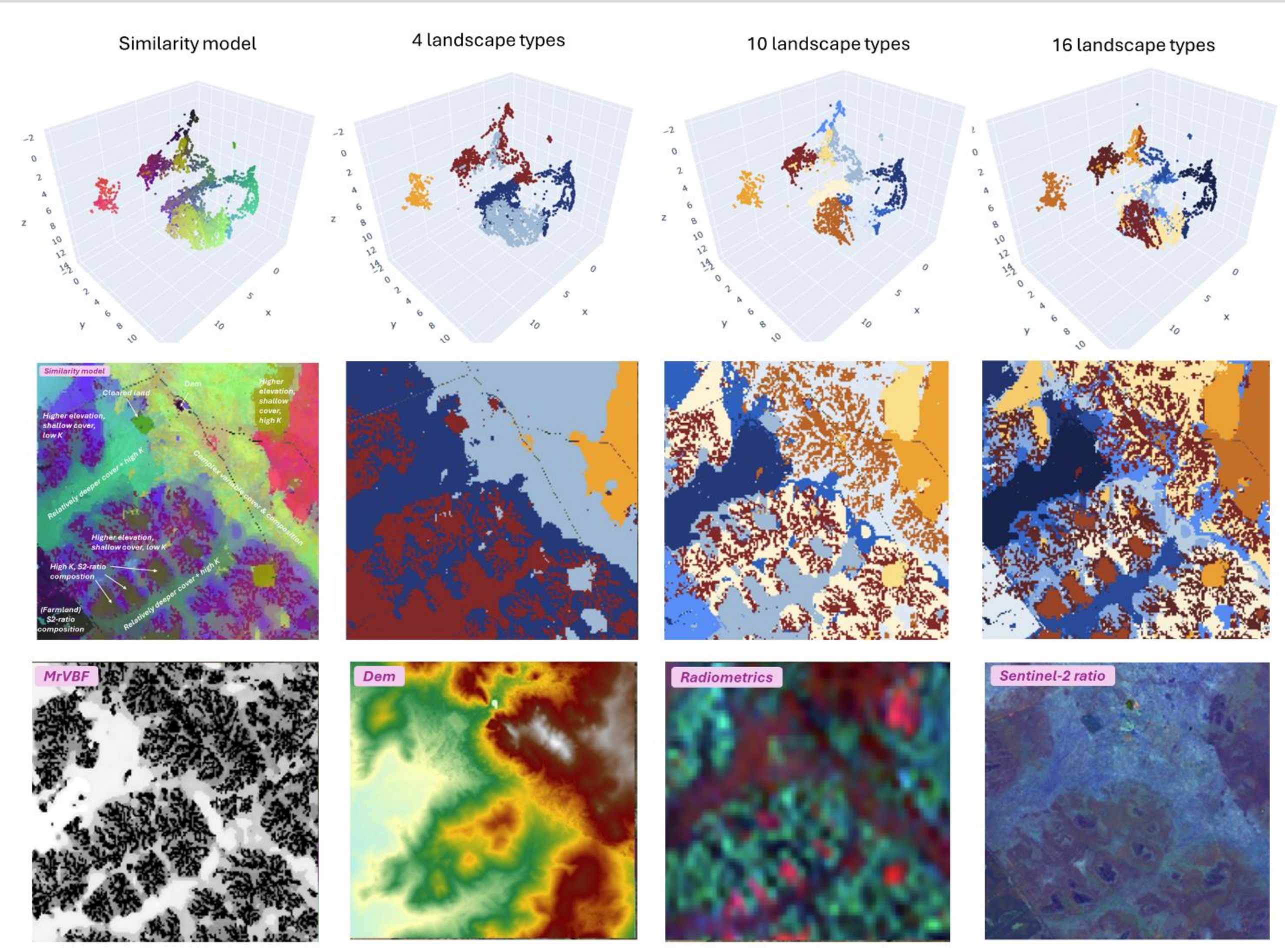
✓ Compare landscape maps to other spatial layers

Once you have narrowed down how many landscape types best fit your survey area, select *Model input layers* (the Digital Elevation Model (DEM), Multi-resolution Valley Bottom Flatness (MrVBF), radiometric data, and Sentinel-2 ratios) in the *Map display options* from the *Primary map drop-down* and/or *Secondary map drop-down* to compare to the different *Landscape type maps* just like in the previous step. Use the *Visibility sliders* to control the transparency of each map layer and use the *Navigation tools*, to *Zoom in*, *Zoom out* or return to the *Model extent*.

Make use of available *surface* and *regolith geology* maps by changing the *Background display* in the *Map viewer window* using the *Radio buttons* in the *Background display options*. These can be helpful in some settings.

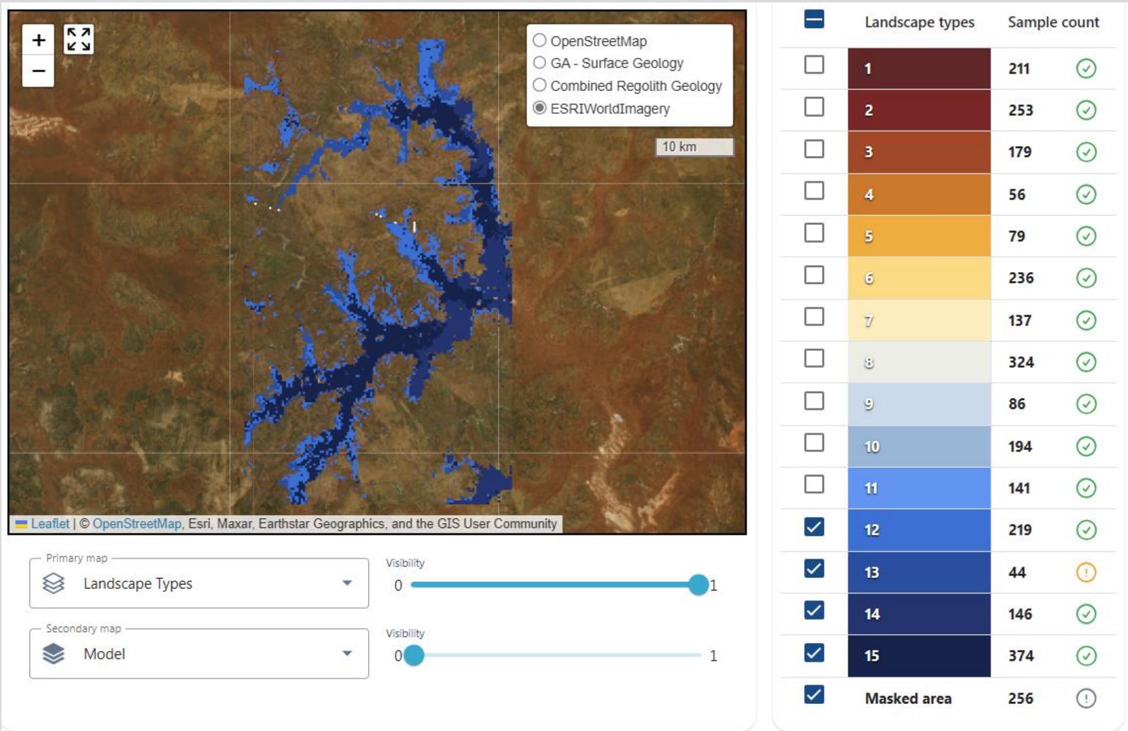
Example

In the example below the ***Similarity model*** shows several distinct clusters of pixels. These are reflected in the 2D map. The map with four landscape types draws out deeper cover in dark blue, areas of shallow cover in dark red-brown, areas of higher elevation in light blue, and the low K radiometric signature in the NW (orange), but it does not represent other features identified in the last step. The map with 10 landscape classes has added more detail (e.g., farmland in mid blue in the SW) and a radiometric K high in the NE (light yellow). If, for example we were interested in picking out the other radiometric anomalies in the south-west as a separate landscape type, these are present in the landscape map with 16 clusters. However, this model divides the clusterbox pixels into groups that are hard to separate visually and adds potentially unnecessary complexity to the map. If the radiometric K highs are important, the ideal output would be more than 10 and less than 16 landscape types. Quickly toggling between the different landscape numbers has narrowed down the options sufficiently to move on to the next step (reviewing the number of samples per landscape type).

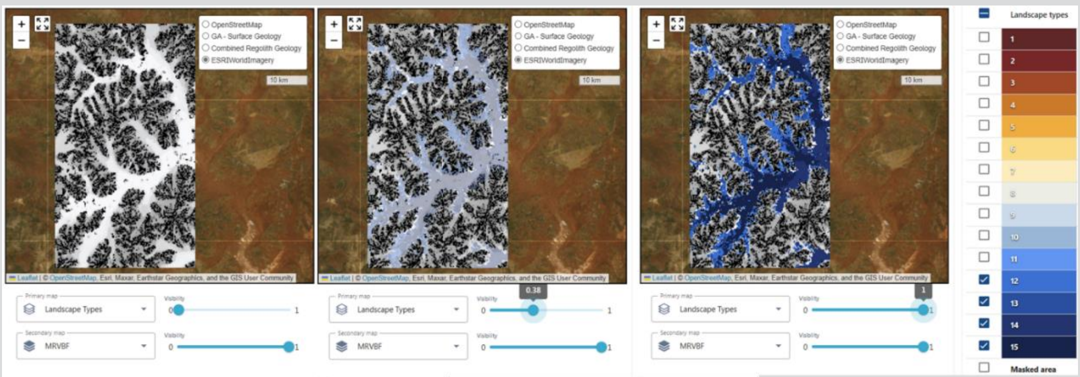


Handy hint:

You can ***show/hide*** different landscape types by ticking/unticking the boxes to the left of each landscape type in the ***Sample overview table***. This can be very helpful to interpret the different landscape features that the mapped units represent. In the example below only dark blue landscape types are displayed.



By displaying the MrVBF in the ***Secondary map*** (bottom) and the landscape types in the ***Primary map*** (top), it becomes obvious that these landscape types are in depositional settings of deeper cover. An appropriate descriptor might be *Channel materials* with lighter blue shades at higher elevations and more shallow cover. Overlaying these over the other input layers may enable more thorough definitions on their relative composition or radiometric signature.



A note on landscape map colours

In general, all landscape types are ordered by cover thickness (MrVBF) from dark red-brown to dark blue.

The MrVBF scale is relative (based on the lowest and highest mean MrVBF values in your area), and low values indicate shallow depth of cover, whilst high values indicate deeper cover. The lowest mean MrVBF values in your model are assigned a dark red-brown colour, and commonly indicate outcrop/subcrop/residual cover grading into orange/yellows/light blues for side-slopes, channel deposits, shallow cover. The highest mean MrVBF values in your model are assigned dark blue and often indicate areas of deep transported cover. The darker the blue the deeper the cover usually is on a relative scale within your model area.

However, since the MrVBF is relative, occasionally, the presence of elevated lowlands (e.g., depressions in higher elevation regions of your survey) or outcrops in otherwise deep cover (e.g., silcrete outcrops on salt lakes) can influence the order of landscape colours you might expect. The colours based on MrVBF provides a useful visual ranking but does not indicate the *magnitude* (or form) of differences between landscape types. Especially for models with higher numbers of landscape types, it is common to have multiple landscape types with similar mean MrVBF values which differ in other properties.

Each site is processed individually, and **landscape types cannot be compared to other sites**. What is yellow at a study site in the Pilbara is not at all comparable to what is yellow at a study site in the Yilgarn, or even on the adjacent tenement package if it was processed separately.

Background

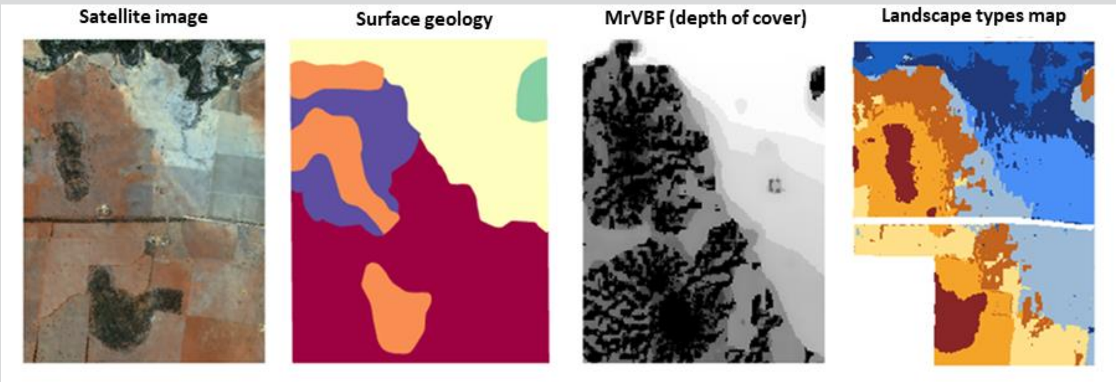
The process of clustering data into proxy landscape types inherently puts a hard boundary on the data (the computer is forced to make a decision) even though, in many cases, there is a gradual change and the exact location of the boundaries is somewhat arbitrary. This is no different to other geological maps, although less biased than human interpretation, and analysis of samples near cluster boundaries should be considered with caution.

MrVBF - A key difference to traditional regolith maps

A key difference between the machine learning derived landscape type maps and a more typical regolith surface map is the inclusion of the MrVBF layer. For surface exploration it is essential to evaluate anomalies distinctly based on the thickness of cover and the MrVBF provides an estimate of this feature. Some areas might look different to traditional regolith maps because of this. This can be particularly pronounced for alluvial channels which might switch colours on the maps. This is because the overall landscape cover thickness has changed with the MrVBF. Hence, although the samples may be in a similar geomorphological setting (fluvial channel) the depth of cover has changed. This is important information in the context of surface geochemistry, since the elemental signatures in deeper cover are likely to be weaker or perhaps unlikely to be detected with surface geochemical techniques. This enables better discrimination and consideration of the geochemical data. A similar observation occurs in sheetwash plains and sand plains as the landscape proxies extend across paleochannels. On the surface, they may look similar, at depth there might be a lot of change with potential paleochannel features and this is not generally captured in a traditional regolith or surface geology maps.

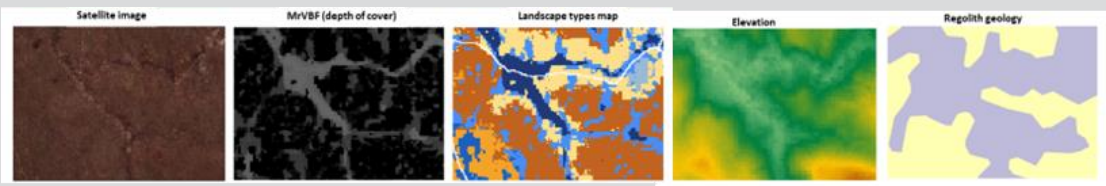
Example 1 - Similar regolith material with changing depth of cover

In the below example, the *satellite image* shows a river on the northern (top) edge. The mapped *surface geology* notes the yellow unit as an alluvial plain and the dark red-brown unit as a sand plain. The *MrVBF* shows changing depth of cover (from relatively shallow in the south to relatively deeper in the north) that are not reflected in the surface geology. However, in the landscape model the alluvial plain is broken into three shades of blue and the sandplain into three landscape types of pale yellow, brown and grey blue. These six *landscape types* show the gradual thickening of the cover as both the sandplain and the alluvial plain stretch away from the upland regions. From a geomorphological perspective, the *surface geology* provides valuable information, from a 3D exploration perspective it does not include key information for the surface explorer.



Example 2 - Alluvial channels with different colours

The example below shows drainage channels in an upland setting of ≈ 2km^2. Whilst the general drainage area is evident in the satellite image and in a more general but broader sense in the elevation and regolith geology maps, the landscape type map shows the drainage channel as mid blue and dark blue. On the surface this region is similar, but the MrVBF shows the broader, flatter depositional nature in the channel, which triggered the model to separate the channel into two landscape proxies.



✓ Review the numbers of samples in each landscape type

The *Sample overview table* provides a summary of how many samples fall within each landscape type and will also update by clicking the *Landscape number buttons*. This number will change with changing amounts of landscape types depending on sample locations. The green ticks and yellow, red or grey exclamation mark *Icons* in the *Sample overview table* relate to the robustness of outliers that are calculated for each element in each landscape type. We recommend a minimum of 50 samples in each landscape type for meaningful statistical outlier definition. The *Icons* give some guidance to allow you to make informed decisions. A landscape model with less landscape types than is ideal for your project area might be preferable because the main aim of the LandScape+ outputs is to provide context for your geochemistry.

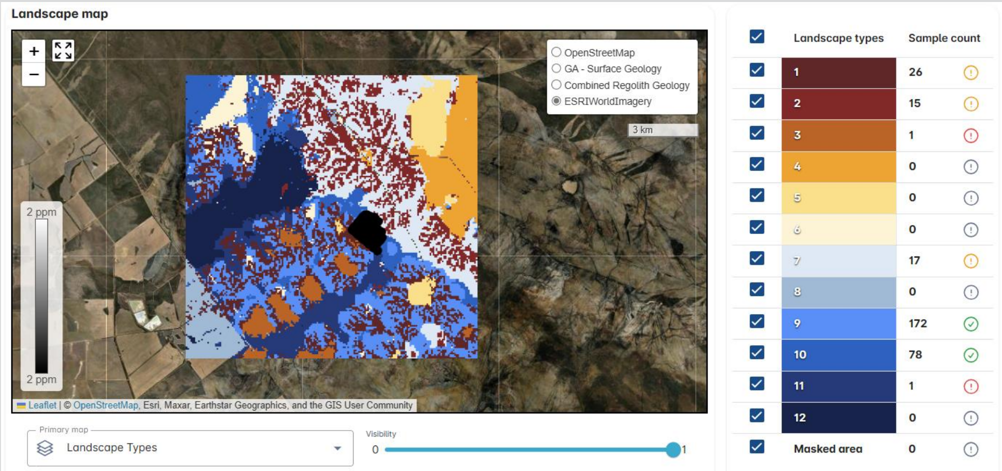
- Over 50 samples are located in this landscape type
- Less than 50 samples are located in this landscape type; assess any outliers with caution
- Less than 15 samples are located in this landscape type; assess outliers with extreme caution
- There are no samples located in this landscape type

The *Yellow information box* below the *Landscape number buttons* provides a summary of what to look out for.

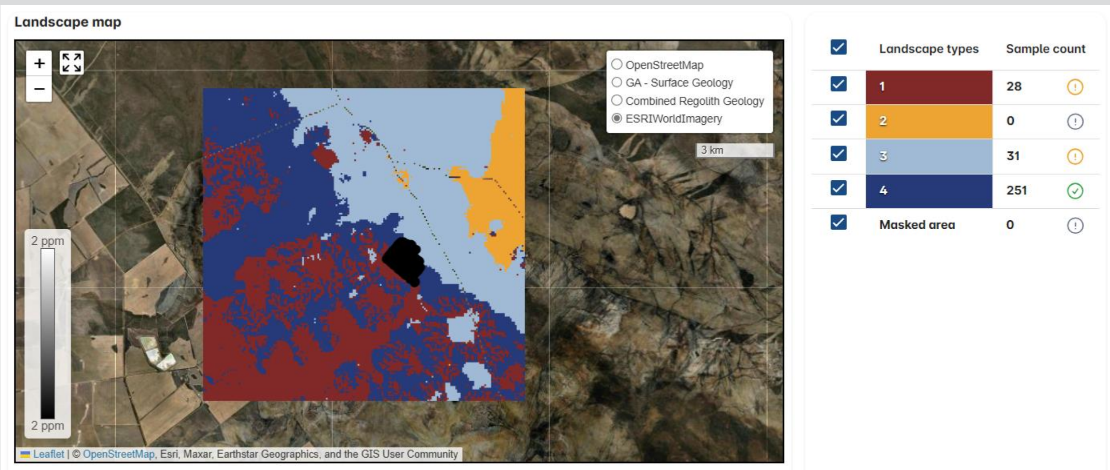
Selecting an appropriate number of landscape types is a balance between representing the major landscape types in your area whilst also enabling meaningful interpretation of geochemical data.

Example – Landscape variation vs. number of samples

The example below has been purpose run using fake geochemistry to illustrate the issue of running a large model area for a small, concentrated survey of less than 300 samples. When the map with the ideal number of landscape types is chosen, only landscape types 9 and 10 contain sufficient samples. **Whilst the map can be very useful to plan extension soil surveys, it is less useful for outlier definition.**



Even the map with 4 Landscape types allows only for review of the geochemistry in landscape type 4 with high confidence, and landscape type outputs 1 and 3 have been flagged as containing low samples. Whilst it is prudent to review these samples as a whole population in addition to grouping them by landscape type, the map does provide landscape context – landscape type 4 (dark blue) is a depositional regime with relatively deep cover, whilst landscape types 1 and 3 are distinguished by higher elevation. As such, samples in these settings are up-slope whilst samples in landscape type 4 are downslope – the model still provides context for interpretation of the results.



✓ *Naming your landscape types*

The model identifies patterns and structures within the input data independent from human input such as regolith or surface geology maps. This means that the resulting map units are not labelled and have no description.

There are no naming conventions, and you do not have to give your landscape types a name, but you generally need to know whether you are looking at **depositional, residual or erosional landscapes** and review the *MrVBF layer* in the *Map viewer window* to identify whether the landscape types represent **none, some, moderate or thick cover**. Some landscape types may be a mix such as residual soils trending into colluvial soils (erosional side slopes).

Put general notes or terms against the colours to help fine-tune your understanding of what the colours commonly represent. You and your exploration team will know your area, and it is likely that you will associate certain landscape settings with a distinct colour of the modelled landscape types.

Use the *Navigation tools* in the *Map window viewer* to *Zoom in* on clearly identifiable features and look for similarities that will explain why these data are grouped.

For example, *Zoom in* on areas with outcrops or residual hills (note the colour) and look for these features elsewhere. Is it the same colour? Commonly, you will start to recognise the same colours corresponding to similar landform features. Some obvious landscape settings to observe are active fluvial channels, river terraces, deltas and large linear dune fields. Some features may be less clear – erosional mid slopes trending into foot slopes, toe slopes, and pediplains/sheetwash plains can overlap in the different landscape proxies.

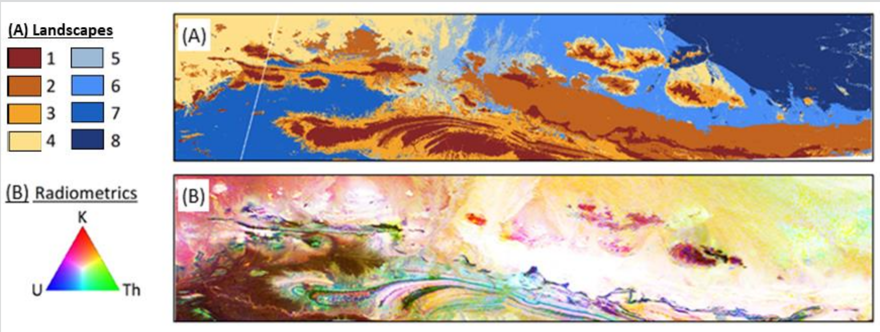
Example workflow for naming landscape types

It is likely that most landscape types identify landforms/regolith materials that you are already familiar with, such as outcrops or drainage channels. You know your survey area - utilise any other high-resolution data and your on-the-ground observations by downloading the *Data Package* and dragging a GeoTIFF (.tif) of your *Landscape types* map into a GIS software.

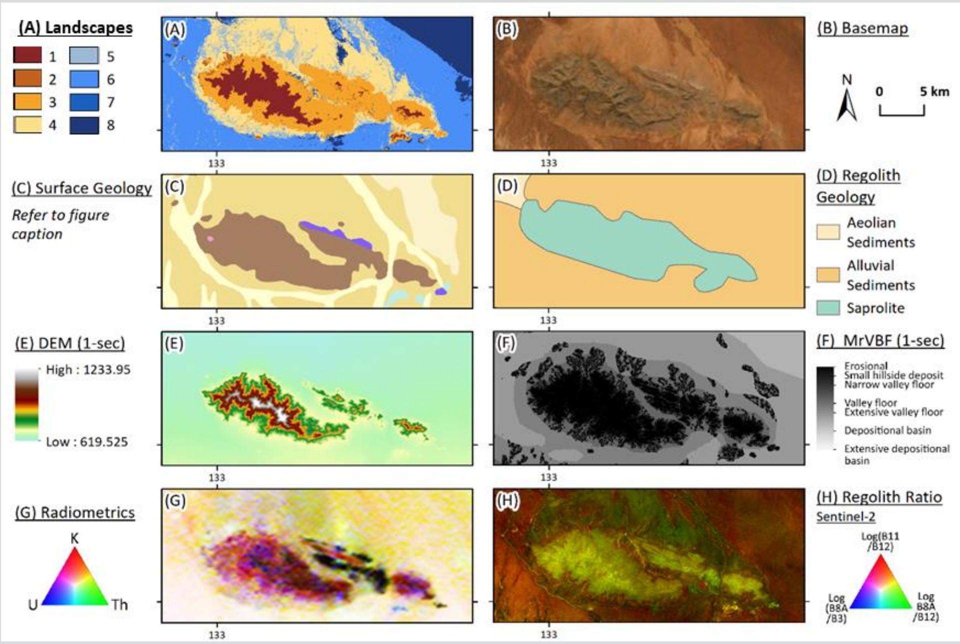
Whilst the below examples illustrate working in GIS software for ease of summary figures, you can complete the same reviews in the *User interface* in a more interactive fashion.

Example 1 - Naming landscape types

In the below example from the MacDonnell Ranges in the Northern Territory, those familiar with the general area will readily identify the dark red-brown and light brown landscape types in (A) as elevated ranges, and the radiometric data (B) points to differences in parent material between the two. An appropriate 'name' for these landscape types may be *outcrop/subcrop* (dark red-brown, 1) and *residual* (light brown, 2) landscapes.

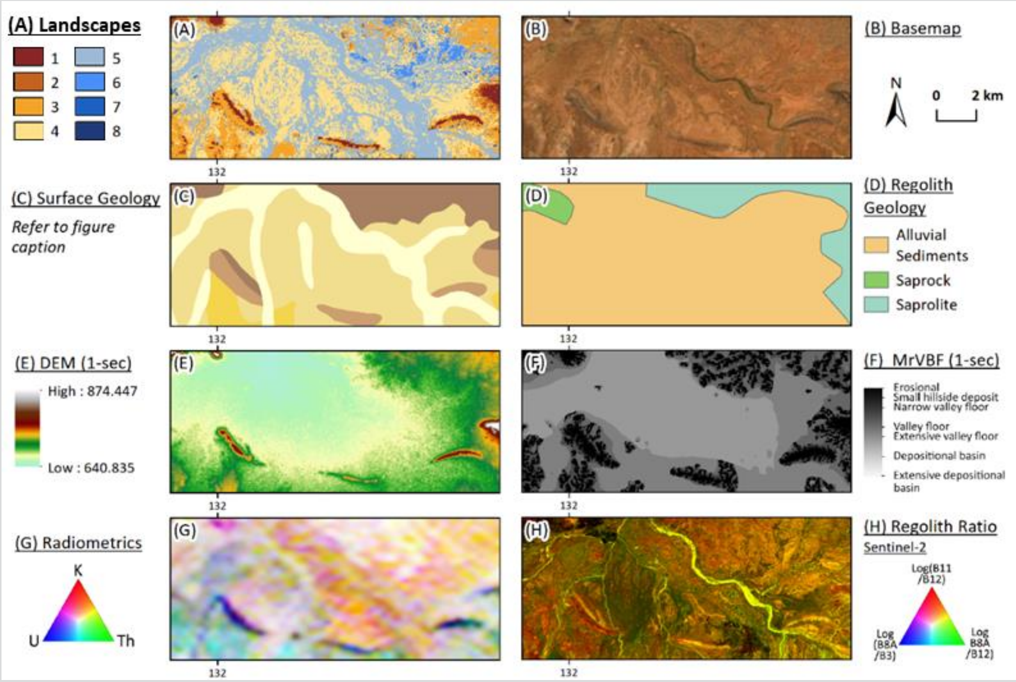


Zoom in on prominent features, such as outcrops, to label further landscape types:



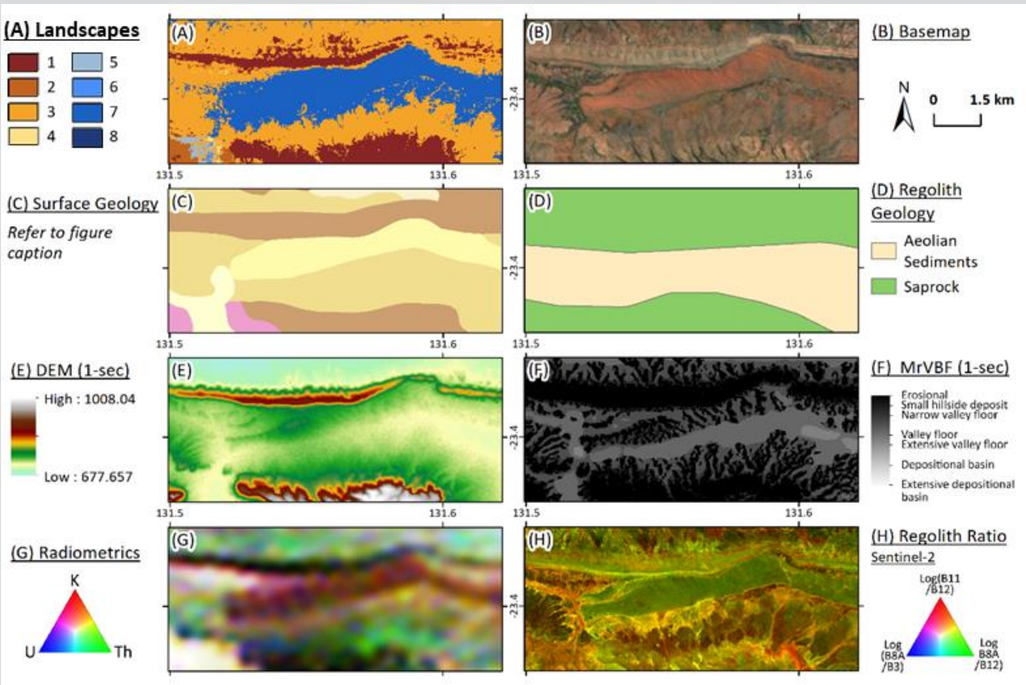
The satellite image (B) shows a pronounced outcrop, the DEM (E) and MrVBF (F) indicate differences in landscape position and depth of cover between dark red-brown (1) and orange (3). An appropriate 'name' for landscape type 3 (orange) may be *side slope*. Further downslope (light yellow, 4, and blue colours, 6 and 8), the Sentinel-2 ratio map (H) indicates a change in regolith materials, which is strengthened by a change in the radiometric signal (G) as well as the depth of cover (F). Both, light yellow (4) and mid blue (6), are mixed shallow cover, which may be termed *shallow cover 1* and *shallow cover 2*; the main distinction being the difference in depth of cover. The dark blue landscape type (8) represents the relatively deepest cover and may be termed *deeper cover*.

Zoom in on other prominent features such as channels, to label further landscape types:



The light grey-blue landscape type (5) is in a generally low landscape position (E) with deeper cover (F) and the Sentinel-2 ratios (H) show different regolith materials which match distinct river channels in the satellite image (B). We may 'name' this landscape type *alluvial channels and/or floodplains*, and the spatial association of this landscape type with surrounding colours also provides more information about the pale yellow landscape type (*mixed shallow cover 2*).

Zoom in on the remaining unnamed mid blue landscape type (7):



The mid blue landscape type (7) is wedged between *side slope material* (orange) shedding from *outcrops/subcrops* (dark red-brown) to the north and south. The setting is a relatively lower landscape position (E) with deeper cover (F) and a distinct colour difference in the Sentinel-2 ratios (H; indicating a change in regolith materials). The satellite imagery identifies a sand dune. Through confirming these same features in other parts of the model area, we may name this landscape type *aeolian cover*.

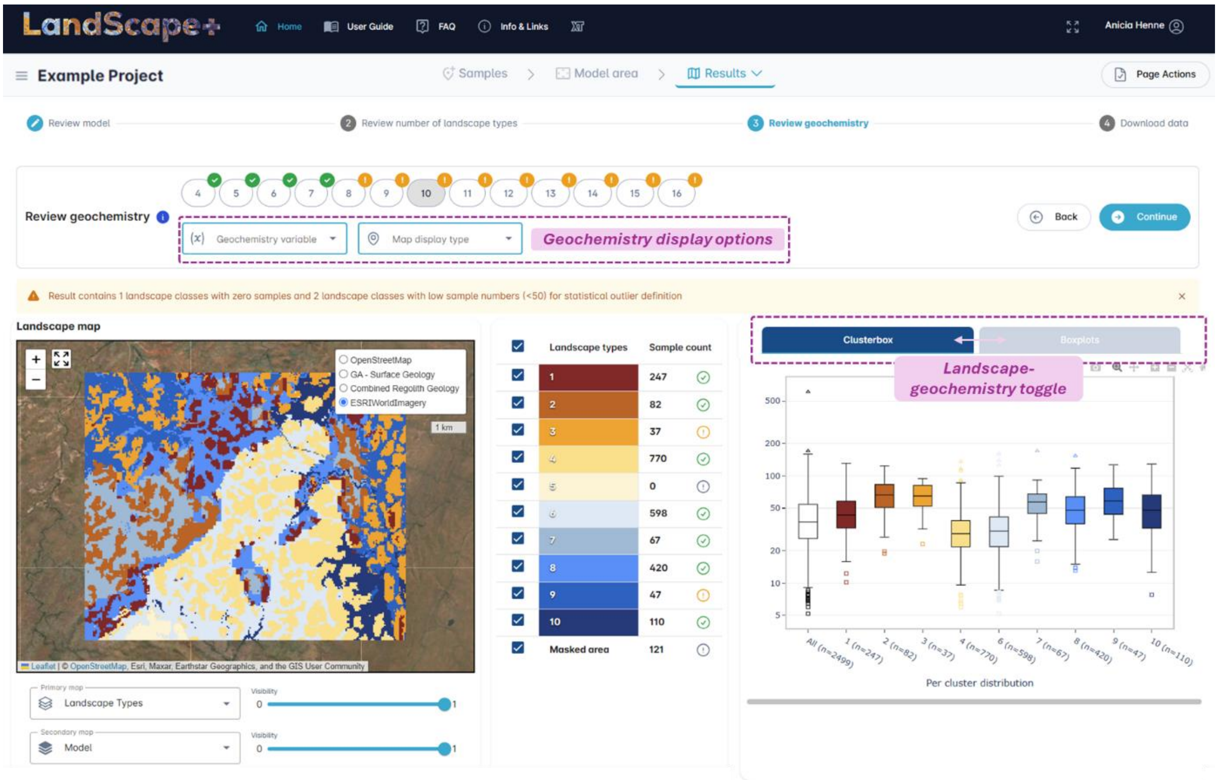
The below interpretation provides sufficient context for geochemical outliers which have been calculated for each landscape type.

Proxy regolith landscape types

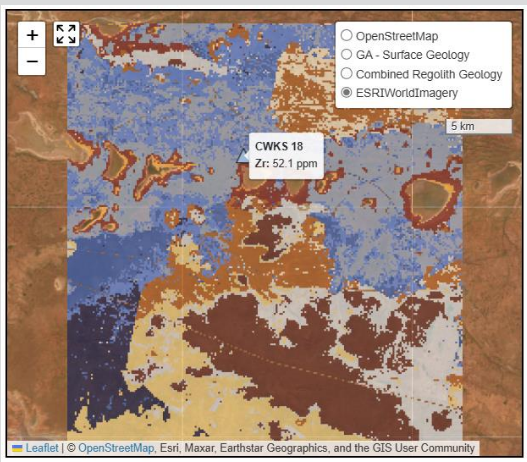
- 1 Outcrop/subcrop (*residual*)
- 2 Residual (*residual*)
- 3 Side slope (*erosional*)
- 4 Mixed shallow cover 2 (*depositional*)
- 5 Alluvial channels, floodplain (*depositional*)
- 6 Mixed shallow cover 1 (*depositional*)
- 7 Aeolian cover (*depositional*)
- 8 Mixed moderately deep cover (*depositional*)

4.4 Review Geochemistry

The *User interface* will now have two additional functionalities. The *Geochemistry display options* underneath the *Landscape number buttons* at the top of your screen and the *Landscape-geochemistry toggle*. The *Geochemistry display options* allow you to view soil samples by geochemistry in the *Map viewer window*. The *Landscape-geochemistry toggle* allows you to toggle between *Boxplots* for each element and the *Clusterbox* for the model. The *Boxplots* will automatically display once you choose an analyte in the *Geochemistry display options*.

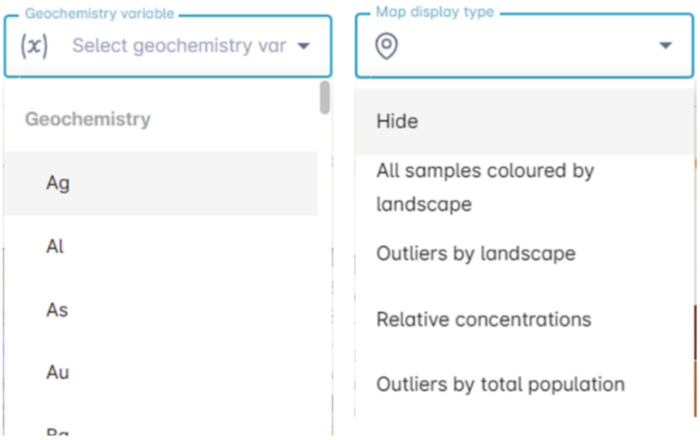


Handy hint: ***Hover*** over your outliers in the boxplots or on the map to display the *Sample ID* and *concentration* of your chosen analyte.



✓ *Review analytes and boxplots*

Choose from the available analytes in the *Geochemistry variables drop-down* to display the analyte of your choice. The *Boxplot* will automatically populate and replace the *Clusterbox*. Samples will only appear on the map once you choose a *Map display type* from the *Map display types drop-down*.



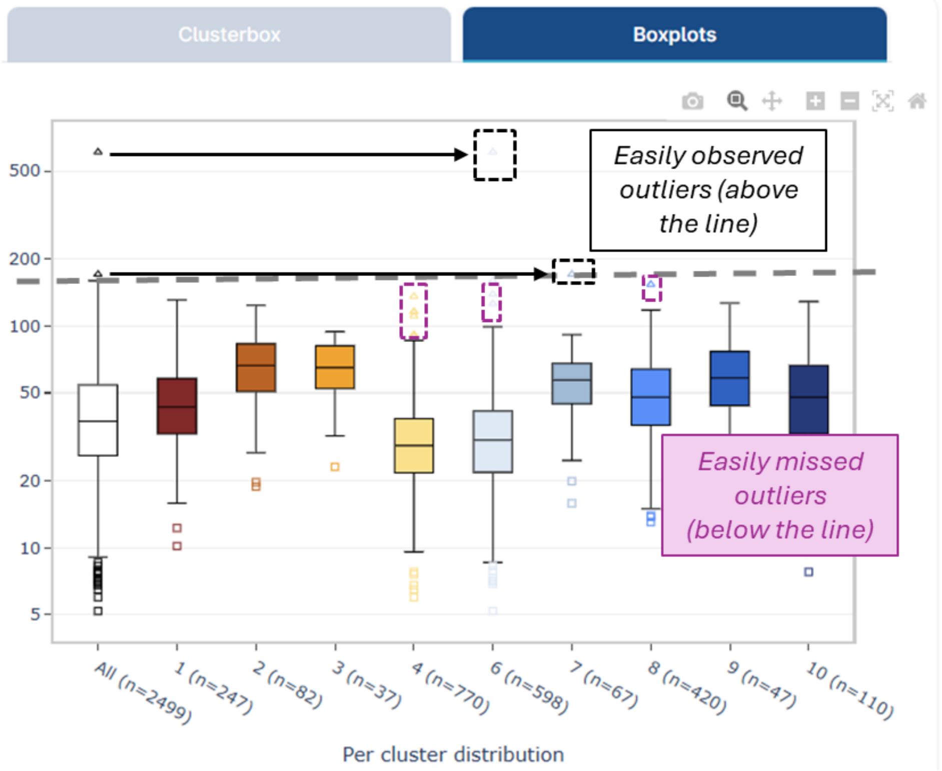
The following Map display types are available:

Map display type	Description
All samples coloured by landscape type	Displays all samples as coloured *circles* and outliers of the analyte chosen from the *Geochemistry variables* as *triangles*
Outliers by landscape type	Displays only the outliers of the analyte chosen from the Geochemistry variables as *triangles*
Relative concentrations	Displays all samples as *circles* in a grey scale from lowest to highest concentrations of the analyte chosen from the *Geochemistry variables*, outliers calculated from the entire survey data (not by landscape type) will display as *triangles*
Outliers by total population	Displays only outliers calculated from the entire survey data (not by landscape type) as *triangles*
Hide	Hides all samples from the Map viewer window

How to read Boxplots

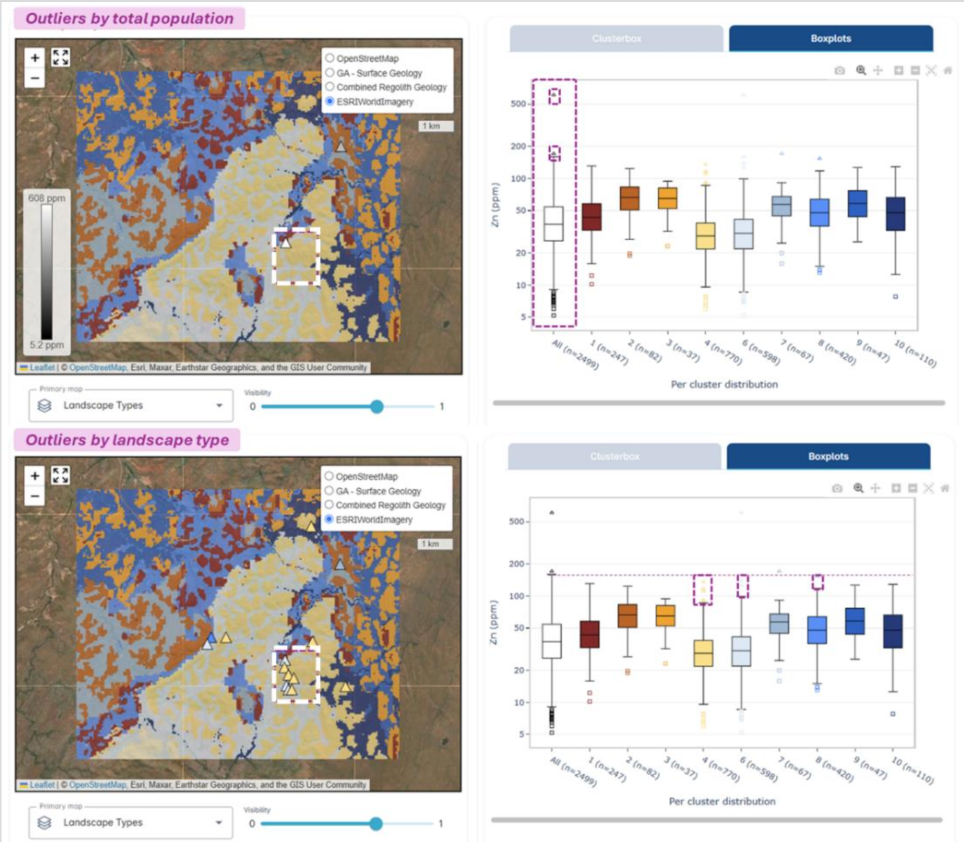
The *Boxplots* are *Tukey plots* and show outliers by sample population. They are linked to the *Geochemistry variable drop-down*, i.e., when you display Au on the map, the boxplot will also change to Au. Boxplots display outliers by total population (white box on the left-hand side) and outliers for each landscape type (coloured boxes). Outliers are indicated as *triangles* in the boxplots and on the map. Outliers may differ depending on the number of landscape types you choose.

The triangles above the white box are always the highest overall relative concentrations of the chosen analyte (1.5 times the interquartile range above the 3rd quartile of the whole population). Outliers calculated for sub-populations in each landscape type are also the highest concentrations in their specific landscape type. However, they may be significantly lower overall. If we draw a line at the 1.5 times interquartile range (the dashed line extending from the top whisker of the white box in the figure below) this becomes more evident. Such outliers, below the line (see triangles in the dashed purple box in the figure below), would be easily overlooked or not identified without landscape context.

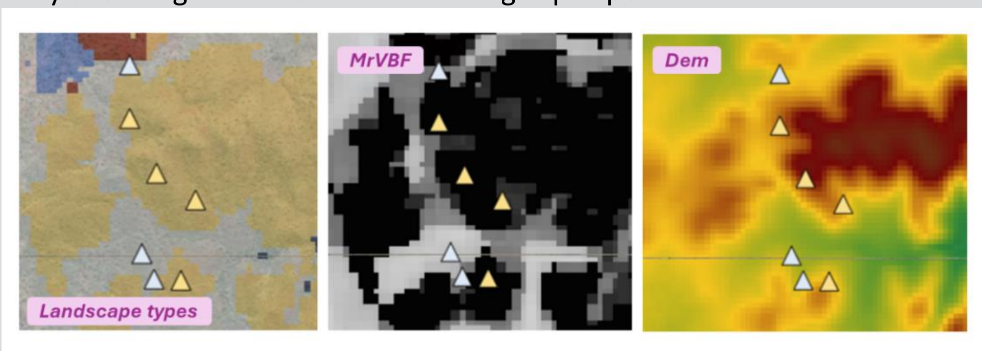


Example – comparison between total outliers and outliers by landscape type

In the example below, when outliers for zinc are calculated on the whole dataset, only two values display as outliers in the map and above the white boxplot (see white *triangles* in both the map and boxplot). When the same data is separated into 10 landscape types and outliers are calculated for each, more outliers are identified (see coloured *triangles* in both the ***Map viewer window*** and the ***Boxplots***). This extends a potential area of interest in the south providing more confidence in this elevated concentration.



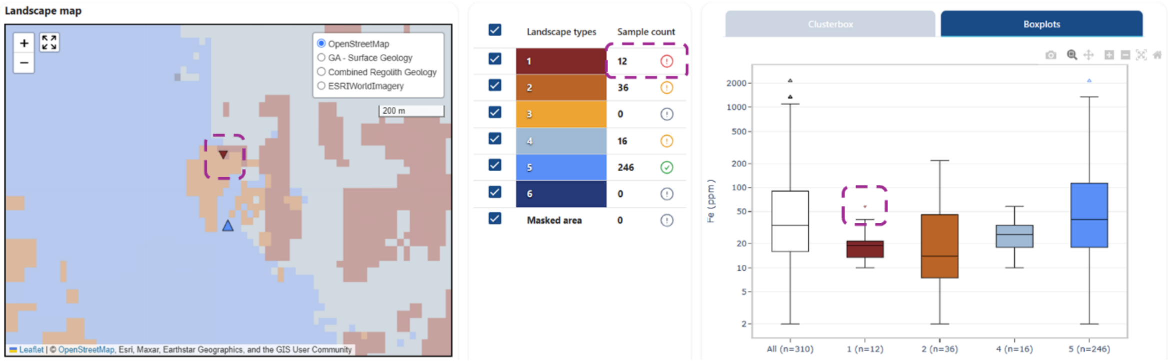
The *User interface* allows us to *Zoom in* on these outliers. Outliers are located in two landscape types. By toggling different input layer features we can easily identify that the light-yellow landscape type is in residual cover at relatively higher elevations, whilst the grey-blue landscape type is in lower landscape positions and likely deeper cover. It is reasonable to interpret the light blue materials to be channel materials which are potentially shedding from the residual settings upslope.



Critically low sample populations

When sample populations contain critically low amounts of samples (less than 15), statistical outlier definition may not be robust and such outliers must be treated with extreme caution. This is indicated by a *Red exclamation mark icon* in the *Sample number overview table*. Such outliers will display as *Upside-down triangles* in the *Boxplots* and in the *Map viewer window* with a smaller font size.

Yellow exclamation mark icons indicate low numbers of samples and these should be reviewed with caution.



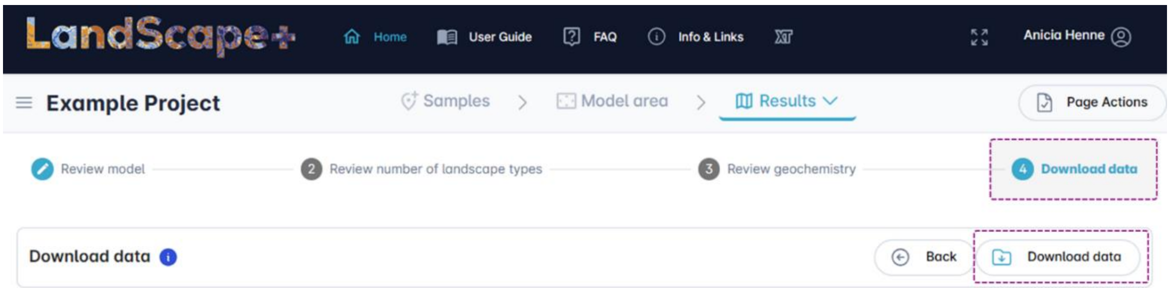
Even if you don't have the recommended number of samples for statistical outlier definition (50 or more) in a given landscape type in your favourite landscape model, you can compare outliers across multiple *Landscape maps* with different numbers of landscapes. If outliers remain outliers across different landscape types, this should provide more confidence that this outlier presents an unusually greater concentration of the chosen analyte than expected for this setting.

4.5 Download results

Depending on your area size and number of samples, your data package may be large (multiple GB), so ensure you have enough space on your disk and a good internet connection for the download.

Click on the *Download data button* to download your data. This will download all landscape models (12 models with 4-16 landscape types) and their corresponding outliers. This is your 'hard copy'. You can return to the online *User interface* for the remainder of your licence (for one year after licence activation).

In addition to the online *User interface*, you will be able to view Principal Component Analysis results in your data package.



Data package contents

The data package download is independent of any actions you take in the *User interface*. You can skip all model review steps in the *User interface* and go straight to download all of your data. The data package contains the following:

- 1 Landscape Type Models and Outliers
- 2 Outliers Independent of Landscape Types
- 3 Principal Component Analysis
- 4 Dimensionality Reduction Model
- 5 Documentation

1 Landscape type maps and outliers
This folder contains the outputs by landscape type for all landscape models.

It contains 12 folders. The number in front of the folder pertains to the number of landscape types in the respective model. Use the outputs from the folder that contains the amount of landscape types you believe best suits your survey area. If you are unsure, it is easiest to review this data in the online *User interface*.

Each folder contains:

- Geochemistry by landscape type
  - Boxplots for each element by landscape type (.png)
  - Outliers by landscape type for each element (.gpkg, .shp)
- Landscape maps
  - Landscape map (.png, .tif)
  - Clusterbox with pixels in 3D coloured by landscape type (.html)
- Geochemistry as CLR values and all outliers (.csv)

2 Outliers independent of landscape types
This folder contains outlier calculations of the uploaded geochemistry without consideration of landscape types. These outliers are calculated from the whole population and are useful for comparison to outliers calculated from the sub-populations (outliers calculated for each landscape type).

The folder contains:

- Outliers by landscape type for each element (.gpkg, .shp)

3 Principal Component Analysis
Your data package includes Principal Component Analysis of all elements for all samples you uploaded with the aim to provide a first-pass multi-element geochemistry interpretation tool to identify potential areas of interest.

The folder contains:

- Spider diagram and spatial plots (.png)
- Principal Components for each sample (.shp, .csv)
- Eigenvalues (.csv)

4 Dimensionality reduction model

This is the dimensionality reduction (UMAP) model of your area that is the basis for the landscape type maps. We called it *Similarity Model* in the *User Interface*.

Background

Uniform Manifold Approximation and Projection (UMAP) is a dimensionality reduction algorithm used to transform the spatial input data for each project area to a three-dimensional representation of data.

This folder contains:

- 2D UMAP of the model area coloured in red-green-blue (.png, .tif)
- 3D UMAP of pixels in the model area in a clusterbox (.html)

5 Documentation

The documentation folder contains useful information that will help you interpret and work with the model outputs. The folder contains:

- A readme.txt file that includes model inputs and parameters and source and version controls
- The online *User Guide* as a PDF

Sharing your project

You can share your model and all outputs with other people by giving them access to your *Project* in LandScape+. Each user you share this project with, has **editing rights**. This means, if you have not run your model yet, any person with this access can change the settings of your model and can run a model (which will use up your licence after a maximum of 3 model runs). If you share the model once it is completed, they can view and interact with the results in the *User interface* and/or download the *Data Package*; they **cannot** delete the project.

To share your *Project*, click on the *Hamburger menu* in the left-hand corner of your *Task stepper* menu. A side panel will open. If you have multiple *Projects* (model areas), select the *Project* you wish to share from the drop-down menu.

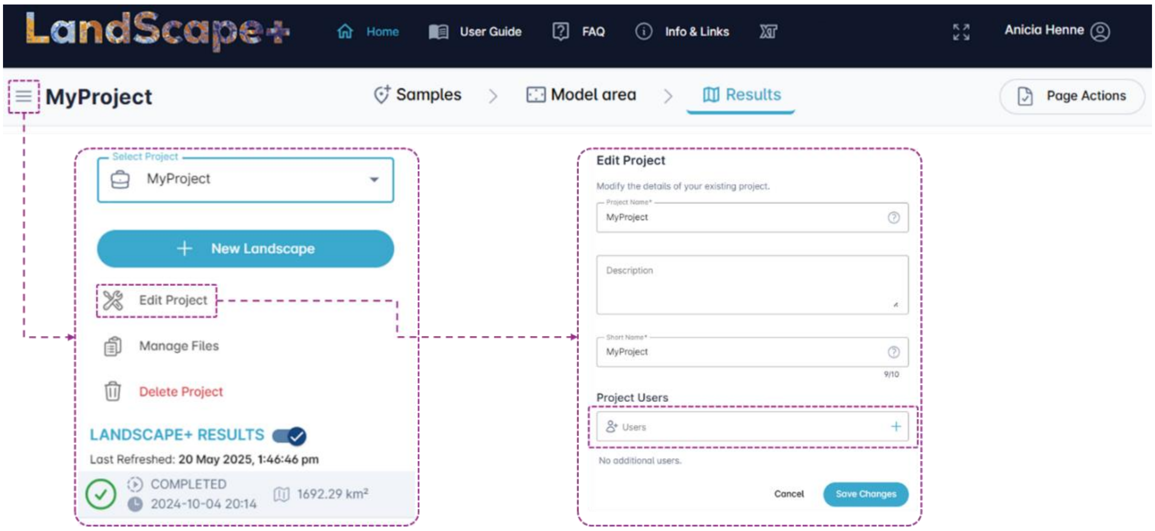
Then click on *Edit Project*. An *Edit Project dialogue box* will pop up.

Enter the email address of the person you wish to share access with into the *Project Users* box.

Press Enter on your keyboard. The user will appear below the *Users field*.

Click on *Save Changes*.

For privacy reasons, we cannot send an automated email with an access link to the email address you entered. You will have to inform the **recipient** and ask them to **create a LandScape+ account**. They do **not** need to acquire a licence. Once they have set up an account, they will have full access to your *Project*.

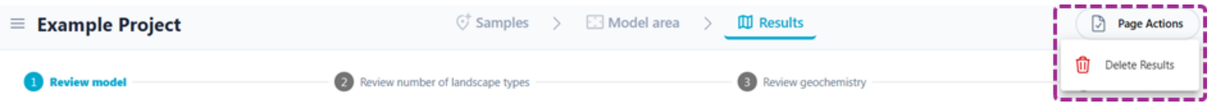


Deleting your results

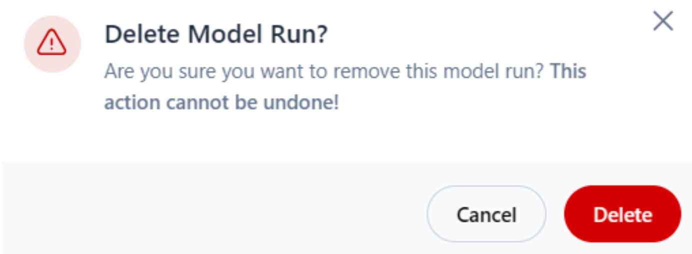
Your results and all associated data will be deleted once your licence expires. The CSIRO will **not** retain a copy of your results, any files you have uploaded or the table you created on the server in the first steps. ***This cannot be undone.***

You will receive a reminder email 30 days before your data will be deleted, in case you have not downloaded your results.

You can also, at any time, delete your results. Click on the *Project Actions button* in your *Task stepper* bar. A *Delete Results* option will appear. Click on *Delete Results*.



A warning message will appear, asking you to confirm. Click *Cancel* if you do not wish to delete the model and all results. Click the red *Delete button* if you wish to delete all results and uploaded files. ***This cannot be undone.*** The CSIRO will not be able to recover these data once they are deleted. **Ensure that you have downloaded your data package BEFORE you delete these results.**



Principal Component Analysis

Your *Data Package* includes Principal Component Analysis (PCA) of all elements* for all samples* you uploaded with the aim to provide a first-pass multi-element geochemistry interpretation tool to identify potential areas of interest. Note PCA does not consider landscape types for the results, only the chemistry of the entire survey in the modelled area. Your data must have more elements than principal components (>5).

Overview

Principal Component Analysis is a statistical technique that reduces the variables (also called dimensions) of a geochemical dataset. All dimensions (elemental results) are transformed into a new coordinate system to describe the data with fewer dimensions (only five in this case) whilst preserving as much of the information (variation) in the dataset as possible. This makes the data easier to visualise and can be useful to identify patterns in a geochemical dataset using multielement trends.

Disclaimer

Principal component analysis is a useful technique for first-pass identification of geochemical trends in a dataset with large numbers of variables and samples. A dataset with few analytes, few samples, or where a large proportion of data is at or below detection, may not necessarily be suitable for PCA and the user should be aware of this during interpretation of their results.

*PCA calculations

Principal Component Analysis is performed on centred-log-ratio and quantile-normalised transformed geochemical data for each sample. Each data point is reduced from n dimensions (number of analysed elements) into five principal components (PC1 to PC5). If more than 10 % of analyses for a given element are missing (e.g., as would be observed where the element has not been analysed) this element is not included in the PCA calculations. If less than 10 % of analyses of a given element are missing, the element is included in the PCA analysis, but no principal components are calculated for the affected sample that has that data missing.

PCA outputs

Spider diagram and spatial plots

LandScape+ produces a spider diagram and spatial plots as a PNG file. The spider diagram shows the loadings (weightings) of each element for each of the five principal components. This illustrates the general geochemical affinity of each principal component. The spatial distribution of each of the principal components by sample is also displayed in the PNG file below the spider diagram.

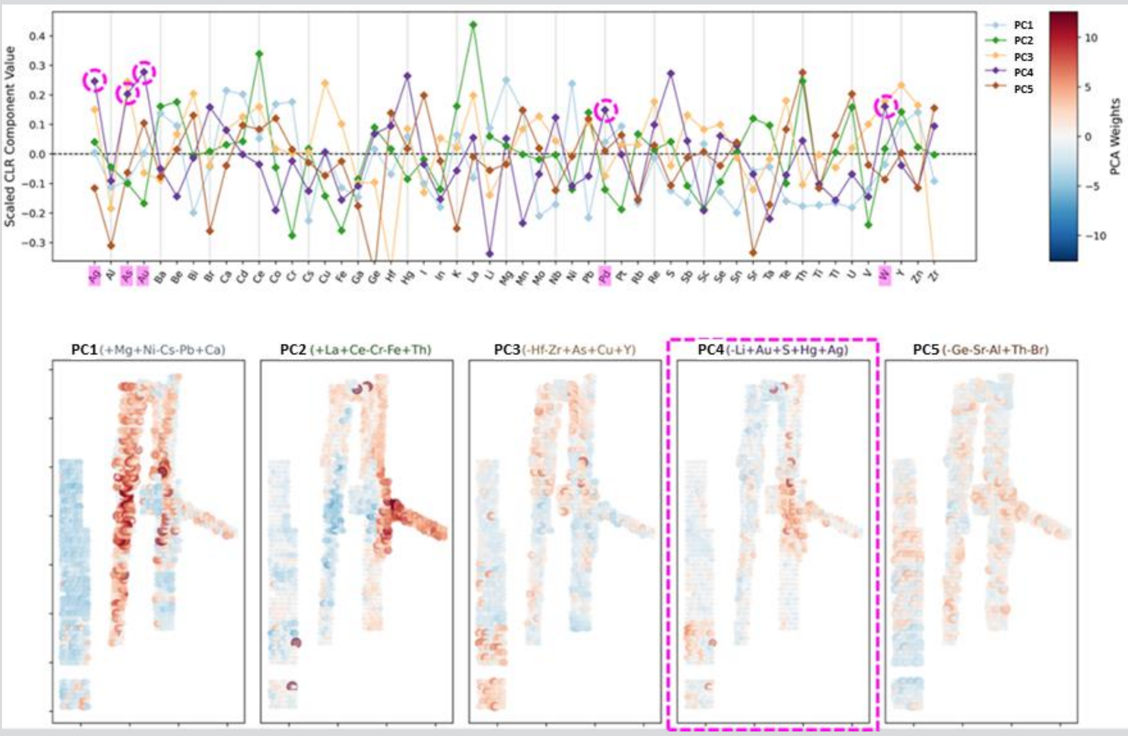
The spider diagram displays loadings for each principal component by a coloured diamond for each element. These are relative and show whether the principal component is positively or negatively associated with this element. All elements included in the PCA analysis are sorted alphabetically from left to right with visual guiding lines every 4 elements (in grey). The further a diamond plots away from the 0 line (dashed black line), the greater the loading for (influence on) the specific principal component.

The boxes below the spider diagram show the spatial distribution of each of the five principal components weighted by both colour and symbol size. The top five elemental loadings (the greatest influence on this principal component, whether positive or negative) for each principal component are indicated in the headings above these boxes. The colour red indicates a positive component weight (association); the colour blue indicates a negative component weight (association). The larger the symbols the stronger the association.

The explained variance of the principal components decreases from PC1 to PC5. Lower principal components will, therefore, often pick out large-scale lithological variation as the major component(s). Principal components with exploration potential commonly explain less variance in the data than the major elements related to geological influence. Hence, principal components that explain less variance (e.g. PC4 and 5) may represent relevant components in the context of mineral exploration. In some cases, signatures relating to mineralisation will constitute relatively low proportions of variance and thus will not be represented by principal components.

Example - Spider diagram and spatial distribution plots of principal components

In the below an interesting principal component for potential mineral exploration targeting with relatively high, positive loadings for Ag, As, Au, Pd and W is indicated in pink circles and shading (PC4; purple diamonds connected by a purple line)



Principal component shapefile and CSV files

Principal components for each sample are available as a shapefile for use in GIS software and as a CSV file.

Eigenvalues (CSV file)

Eigenvalues indicate how much of the variance in the data is explained by each principal component, as a percentage for each principal component.

Example – Federation and Dominion Prospects

In the example below of the first three principal components from the Federation Project site (read the full report here), the first principal component (PC1) explains 46 % of the variability within the dataset and, unsurprisingly, highlights shallow geology and climatic influence (slope aspect or elevation) and matched the landscape proxy types/colours in the background. PC2 explains 18 % of the variability within the dataset and is associated with As, Au, Ta and Tl. PC3 only explains 15 % of the variability within the dataset, but also shows exploration potential with association of Ag, As, Cd, Cu, Hg, Ni, S, W and Zn. It therefore has positive loadings with some target and pathfinder elements such as Ag, Cu and Zn, but not with Au or Pb. PC3 also is strongly influenced by Mn in the samples. In retrospect, it appears that PC2 highlights the Federation mineralisation (downslope from and immediately to the northwest of the Federation prospect) and PC3 highlights the Dominion prospect (positive PC3 scores coincide broadly with the Dominion prospect and are positive; they are more subdued around Federation).