



# How to use *UltraFine+<sup>®</sup> Next Gen Analytics* - Outliers by landscape type data

## What are outliers by landscape type?

Outliers in soil geochemical data are commonly calculated on all samples (the whole population) collected over an area of interest, regardless of their landscape context. Landscape types are essentially sub-populations within the sample set that can be assessed separately. The common “all samples” approach ignores the underlying processes that may affect metal dispersion. For example, high metal concentrations may be readily identifiable as outliers in a geochemical dataset where samples were collected over mineralisation beneath exposed outcrop or shallow residual materials. The same mineralisation 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 geochemical outliers. The UltraFine+<sup>®</sup> Next Gen workflow provides a basic, first-pass interpretation of geochemical samples by proxy regolith type to identify otherwise “overlooked” potential anomalies.

## Where do I find the outliers for my landscape types?

You will find your outliers by landscape type in your *Data Package* under *3 Landscape Models* (kmeans4, agg8 or agg12) → *2 Outliers by Landscape Type*. In this folder you will find boxplots for each of the analysed elements (under *Boxplots*) and maps of these outliers for each analysed element (under *Maps*). The outliers are also available as shape files (under *GIS*) with associated layer files that will display your outliers as triangles coloured by corresponding landscape type. All outliers for each

element by sample are also available in a CSV file, which you can find in your *Data Package* under *1 Landscape Analysis*.

**TIP:** Instead of reviewing each map and boxplot for outliers by landscape type for all 52 elements and comparing them to outliers of the whole population, you can also use the DSO to quickly toggle between outliers.

## How are outliers by landscape type calculated?

Tukey outliers are calculated on log-transformed data. The outliers are defined as values that are greater than 1.5 times the interquartile range added to the 3<sup>rd</sup> quartile value. The interquartile range is the difference between the third and first quartiles that contain 50% of the data. You can find a description with images and data [here](#). All concentrations are treated as log normal transformed data even if the distribution is approximately normal with the raw values (common for major elements, rare for trace elements). All values below the detection limit are replaced with half the detection limit value for each respective element and included in these calculations. The data is grouped by their corresponding landscape cluster and outliers are calculated for each of these clusters.

## Do I need to name my landscape types before I look at the outliers?

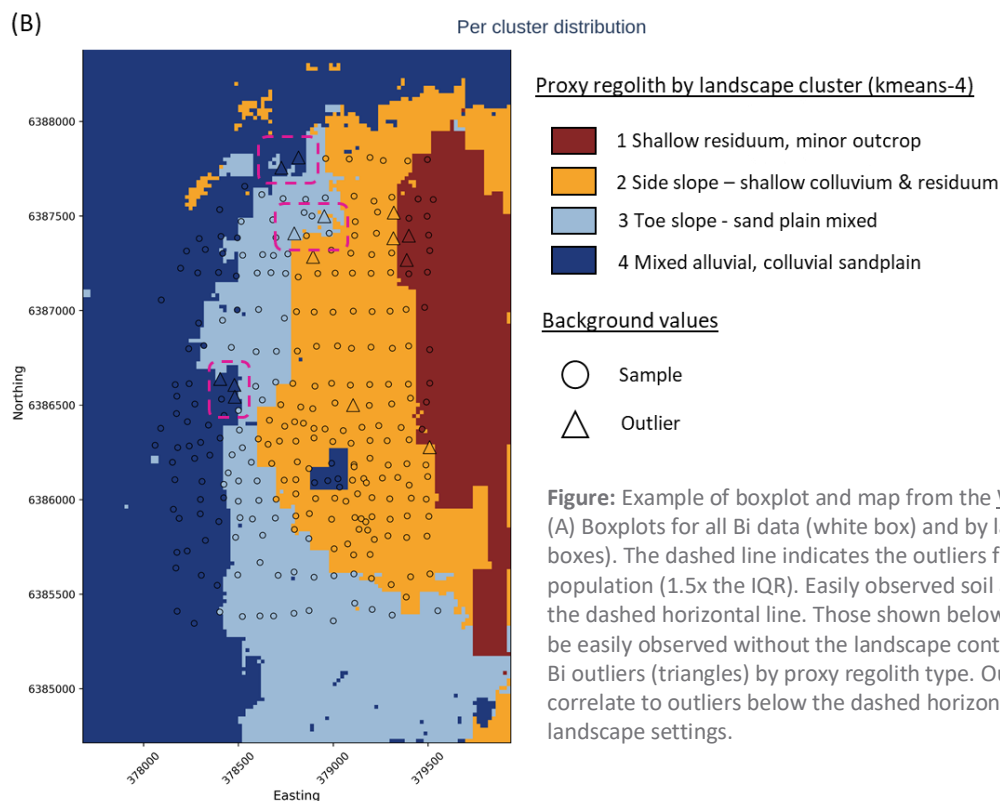
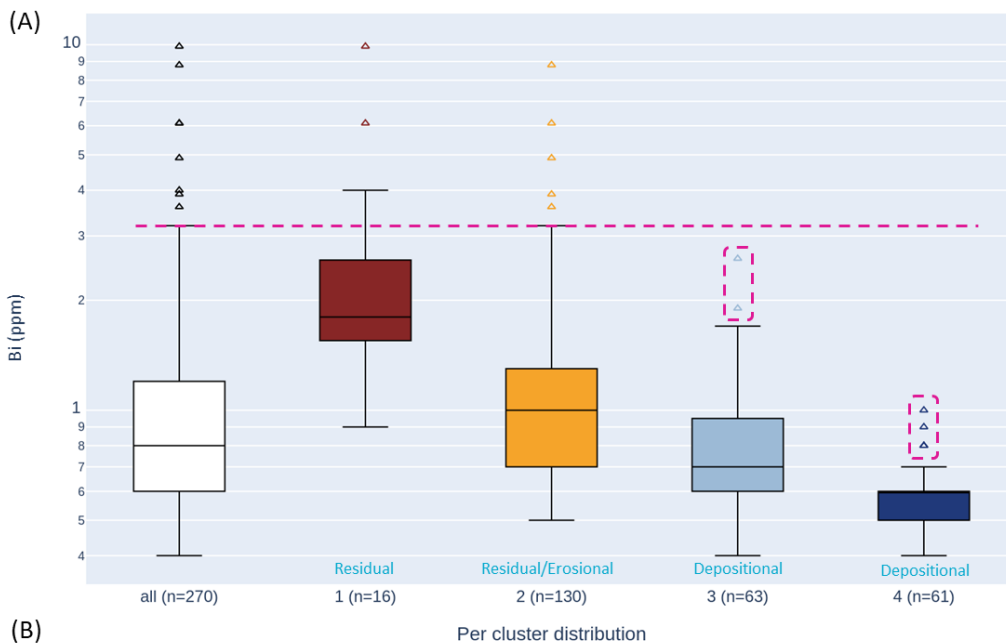
You don't have to give your landscape types a name, but you need to know (a) which output to use (4, 8 or 12 landscape clusters), and (b) whether you are looking at depositional, residual or erosional landscapes. If you are unsure how to pick the most appropriate number of clusters, a How-to guide with an example of how we derive proxy regolith types using spatial data and/or the DSO is available [here](#). The explorer tends to know their

area well, and it is likely that they will easily associate certain landforms with a distinct colour of the machine learning derived proxy regolith types.

**TIP:** The *readme.txt* file within your *Data Package* contains a recommendation on which model (kmeans4, agg8 or agg12) is the most appropriate for your soil survey.

## How do I read the boxplots?

The below example of Bi over the Wagga Tank project area (read the full report [here](#)) shows a typical output for a landscape model with four landscape clusters (kmeans4). The sample population has been broken down into four coloured boxes that each represent a landscape cluster. Traditional outliers calculated from all data as a single or whole sample population are always presented for comparison in a white box on the left-hand side. Outliers are indicated as triangles to highlight potential anomalies within different landscape settings. For most of these landscape types, outliers (triangles) below the dashed line would have been considered unremarkable (background concentration values) if evaluated as part of the whole data set. However, there are outliers from samples in landscape sub populations in depositional settings (grey-blue and dark blue boxes). The dark brown and orange populations on the other hand are from residual and/or erosional landscape settings and are therefore, as expected, well represented by outliers in the overall sample population (compare white box/triangles with brown and orange triangles above the dashed line). Outliers in the light and dark blue, depositional landscape settings form potential targets in cover.



A note on **small soil surveys**: We consider the ideal minimum number of samples for statistically relevant analyses to be 50. As a precaution, outliers in clusters with less than 10 samples will not display on maps or in shapefiles. In the above example, the sample population for landscape cluster 1 (dark brown; n=16) is too small to be considered for meaningful interpretation.

## How do I compare my outliers by landscape type to “traditional” outliers?

While the workflow was designed to identify outliers in transported cover, the workflow accommodates identification of anomalies in other settings by also generating data, maps and shapefiles for the whole sample population. Hence, original, elevated elemental signals are not lost. This should be considered when interpreting data where prospects are located in outcrop, sub-crop or shallow residual soils. You can look at outliers by landscape type on the maps supplied (see above) and compare them to the outliers of the whole data population (white box in box plots) on maps which are also supplied as PNG files (under *3 Outliers Independent of Landscape Type → Maps*). Shapefiles for both, outliers by landscape type and “traditional” outliers, are available. The shapefiles have 3 values: -1 (for negative outliers – below the box), 0 (part of the general background) and 1 (positive outliers) in case you would like to interrogate further or look at abnormally low concentrations of elements, too. The easiest way to get a quick overview of your outliers is by toggling items on and off in the Digital Sample Observer (DSO) – a html file that contains all of the generated data. You can find this file for each model in your Data Package under *2 Digital Sample Observer*. An overview video on how to use this file can be found [here](#).

### Outliers calculated on the whole survey population versus on sub-populations by landscape clusters

The example below is from a large survey area (approximately 15 km x 20 km) with approximately 5000 samples. The screenshots from the DSO below show that, when outliers for zinc are calculated on the whole dataset, only two values display as outliers (A). When the same data is separated into 12 landscape clusters and outliers are calculated for each, more outliers are identified (B). When zooming in on the north-eastern outlier (Inset A), the potential target is essentially extended south and with more anomalous samples providing more confidence in the target area (Inset B).

