



# How to use UltraFine+® Next Gen Analytics - Pick a Landscape Model

## What are landscape models?

The UltraFine+® Next Gen Analytics workflow clusters publicly available spatial data to produce models of proxy regolith types, or landscape clusters, that represent the physical regolith types within your project area. The standard output of the UltraFine+® Next Gen Analytics provides each site with three different models (kmeans4, agg8 and agg12) that were developed to fit a range of different project sites. Ideally, “kmeans4” (a four-cluster k-means clustering) is used for small areas (< 20 km<sup>2</sup>), “agg8” (an eight-cluster agglomerative hierarchical clustering) for larger areas (> 20 km<sup>2</sup>), and “agg12” (a twelve-cluster agglomerative hierarchical clustering) for larger areas with dense sampling.

The appropriate number of clusters for a given project area depends on area size, the resolution of the input layers, the complexity of the respective landscape, and the needs of the data user. The UltraFine+® Next Gen workflow was developed to generate landscape context for geochemical samples. Hence, the number of samples across a given area must be considered to avoid small clusters with few or no samples. The resulting number of landscape clusters are, therefore, a balanced approach to represent the major landscape types while also enabling meaningful interpretation of geochemical data.

## How are the landscape types derived?

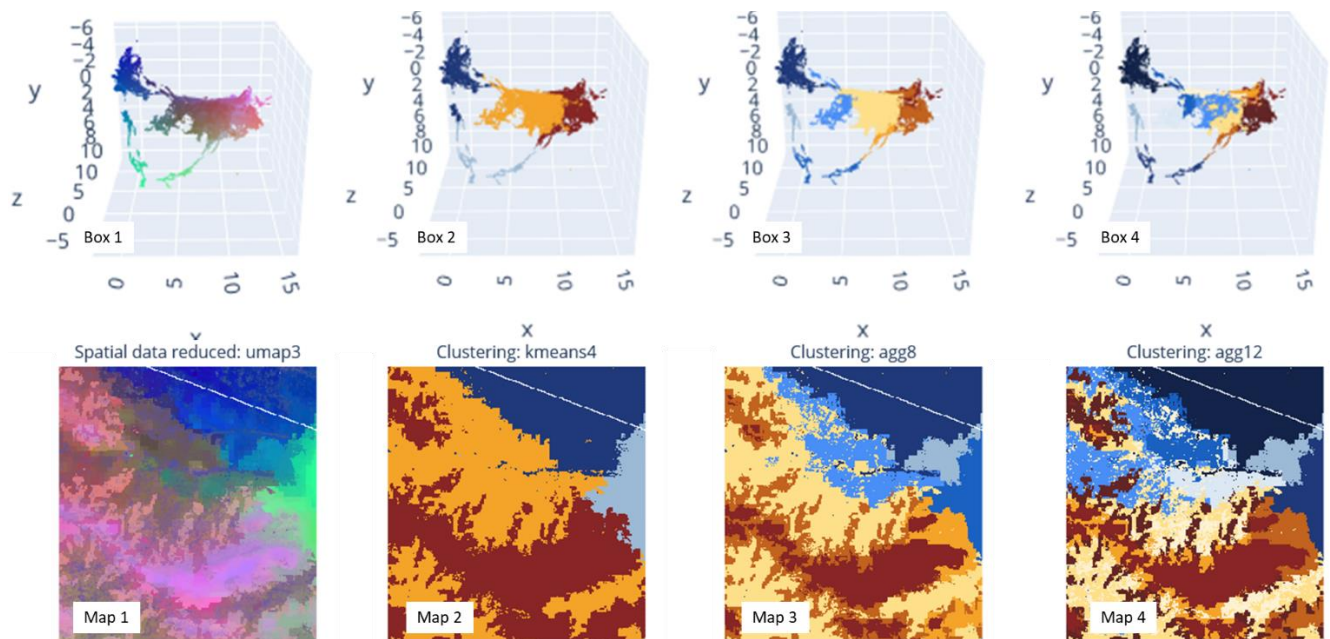
The workflow looks for clusters of data with similar spatial properties and gives each cluster a unique colour. The machine learning does not take any information on the geographical location or UltraFine+® data/soil properties of these points into account. The spatial data layers that are used to generate the proxy landscapes are selected for their relationship to regolith landforms and can provide information on landscape position, depth of transported cover, and parent and/or regolith materials. To minimise the introduction of human interpretation, we do not include surface geology or regolith maps.

SPATIAL INPUT LAYER	LANDSCAPE INFORMATION	DESCRIPTION
DEM	General landscape position	Copernicus GLO-30 Digital Elevation Model
MrVBF	Estimate of depth of transported cover	Multi-resolution Valley Bottom Flatness index (refer to Gallant and Dowling 2002)
Radiometric data	Indication of parent material	Radiometric data, K pct, U ppm, Th ppm
Regolith ratio	Regolith material	A product derived from ratios of bands from Sentinel-2 satellite data (adapted from Gozzard 2005)

We extract point values from these data layers (usually at a resolution of 30 m) for each project site and use an algorithm (called Uniform Manifold Approximation and Projection or UMAP) to reduce each sample point to 3 dimensions and apply an RGB (red, green, blue) colour scheme. This is what we refer to as dimensionality reduction and it makes it easier for humans to visualise (see “cluster boxes” in the figure below) and prepares the data for clustering. The closer the data points are to each other, the more similar they are. If most of your points are close together, this indicates that there is not much landscape variability between those points. If you have points that form distinct clusters, especially offset from other clouds of data points, it indicates that you have distinct landscape types – the more distinct clusters you can see, the more different landscape types exist in your area. Multiple clustering algorithms (“kmeans” and “agglomerative”) are applied to the data with varying numbers of clusters (4, 8, and 12) to separate out different landscape types. This results in three standard outputs for each site: kmeans4, agg8, and agg 12.

After the data is clustered, each point is reassigned its coordinates and is projected on a map coloured by the landscape cluster it belongs to. This is the basis of the clustering model. Ideally the landscape clusters will resemble the UMAP RGB closely and this is best viewed in the Overview DSO (Digital Sample Observer) file

([ASX\\_CompanyCode\\_Site\\_overview\\_DSO.html](#)). We test this for each site and write a recommendation in your *readme.txt* file.



Top row: Example of cluster boxes showing how the samples are associated in 3D space using an RGB colour scheme (Box 1) and coloured by landscape type (Box 2 with four landscape types, Box 3 with eight landscape types, Box 4 with twelve landscape types). Bottom row: Points from the cluster boxes projected in real-world spatial context creating proxy regolith maps with an RGB colour scheme applied (Map1) and with four (Map 2), eight (Map 3) and twelve (Map 4) landscape types over the same project area. The white diagonal line is a masked-out road and is not part of the clustering schemes.

## Where do I find the landscape models?

You will find your main files for landscape review in your *Data Package* under *2 Digital Sample Observer* → [ASX\\_Company\\_overview\\_DSO.html](#). The *Overview DSO* is the best place to compare the three clustering algorithms before a more thorough assessment based on the preferred clustering output.

## How do I know which landscape model to use?

The Overview Digital Sample Observer is the best way to quickly determine which landscape model to use. It will show you the UMAP (3-dimensional representation of the spatial attributes of your area) along with the three landscape models. It also shows the input layers (MrVBF, radiometrics, DEM and regolith ratios) along with a true colour satellite image. The available surface geology and regolith geology maps are included for comparison. These layers and your own knowledge of the site will help you determine how varied the landscape is. Generally, if you have lots of distinct clusters in the

cluster boxes, you will want more landscape types in your model. This will usually be reflected in changes in the input layers and, if a good regolith map exists, this should also give you a rough indication.

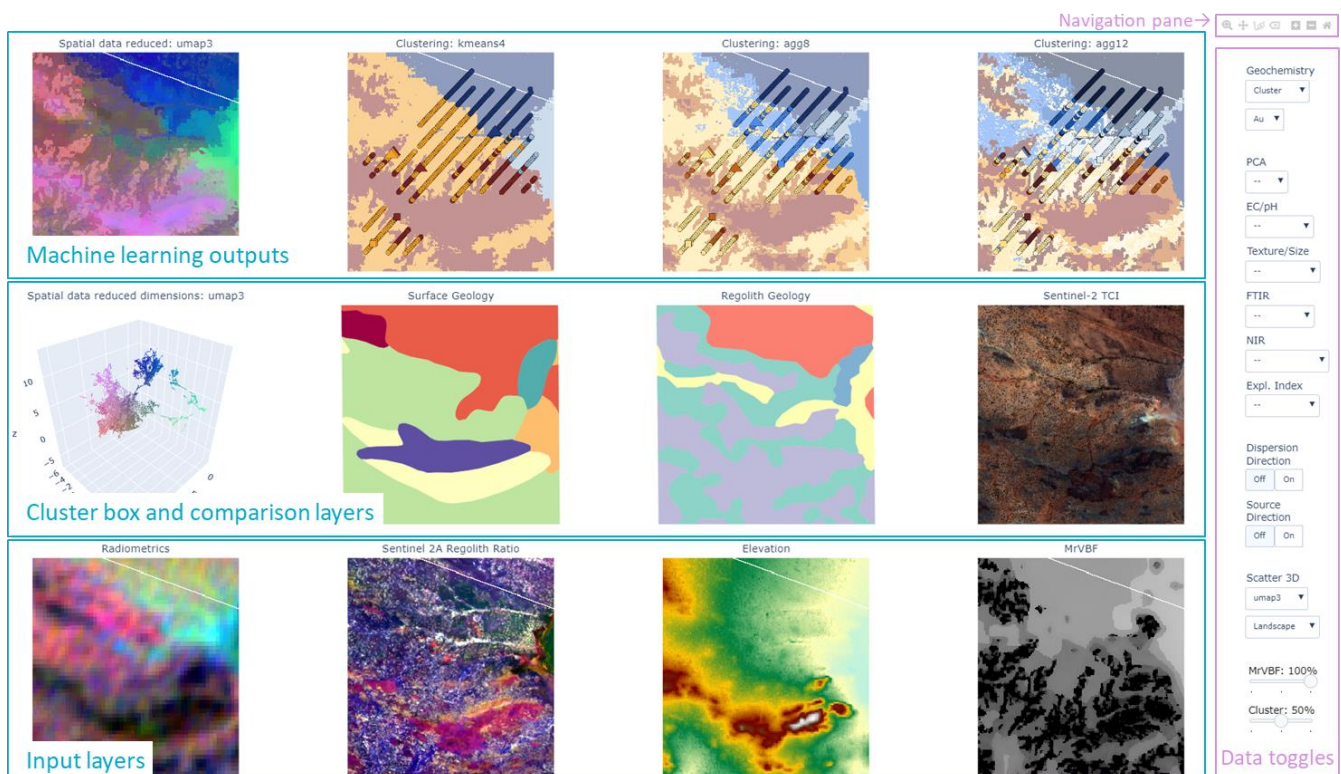
**SHORTCUT:** The *readme.txt* file within your *Data Package* contains a recommendation on which model (kmeans4, agg8 or agg12) is the most appropriate for your soil survey. This will save you a lot of time, especially when you are not familiar with machine learning and/or the different input layers!

**ATTENTION:** We often recommend a landscape model with less landscapes than is ideal for your project area. This is because the main aim of the Next Gen Analytics outputs is to provide context for your geochemistry. Hence, we need to consider the number of samples across your project site to avoid small clusters with few (<50) or no samples which would render the statistics for your project site meaningless. Picking the right number of landscape clusters is a balanced approach to represent the major landscape types while also enabling meaningful interpretation of geochemical data. You can of course still use the most detailed landscape model to plan future/follow-up exploration activities.

## How do I use the Overview DSO?

To decide which landscape model (how many clusters) to choose, use your Overview DSO.html tool to compare the landscape cluster maps:

- **UMAP**  
Your preferred cluster map should match this closely
- **Cluster box**  
Are there multiple separate clusters of data points and have each been coloured differently? The more distinct clusters, the more landscape types are needed to represent your project area. Use the “Scatter 3D” data toggle on the right-hand side to display the clusters in the box in 4, 8 or 12 colours
- **Available surface and regolith maps, satellite imagery**  
If good maps are available for your site, they should provide you with some context to have confidence in the appropriate amount of clusters
- **Input layers (radiometrics, regolith ratio, DEM, MrVBF)**  
The highest variability at any given sample point will have the most “pull” on the models. Clusters should reflect one or more of these features
- **Your knowledge/additional data**  
You know your project area best - utilise any other high-resolution data and your on-the-ground observations
- **The spread and density of your samples**  
Toggle on your samples with the “Geochemistry” toggle on the right-hand side and ensure that most of your landscape clusters contain at least 30 to 50 samples. If you have a small survey with only 200 samples spread across 12 landscape clusters, the interpretation of outliers by landscape type will be statistically meaningless. Choose a model that has 30 to 50 samples per cluster as a minimum. An easy way to establish the numbers of samples in each cluster is by opening the boxplots in your *Data Package: 3 Landscape Models* → *kmeans4/agg8/agg12* → *2 Outliers by Landscape Type* → *Boxplots*



Review the different clustering maps with the supplementary data provided in the DSO, such as the cluster box, surface and regolith geology and input layers. The DSO allows you to zoom in on clearly identifiable features and look for similarities that will explain why these data are grouped. The recommendation for this example site is 12 landscape clusters to ideally represent the regolith types within the area. However, due to limited samples in some landscape types, the geochemistry (and all other workflow outputs) is best reviewed in the context of 8 landscape clusters.

## References

1. Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*. 39, 1347, doi:10.1029/2002WR001426
2. 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.