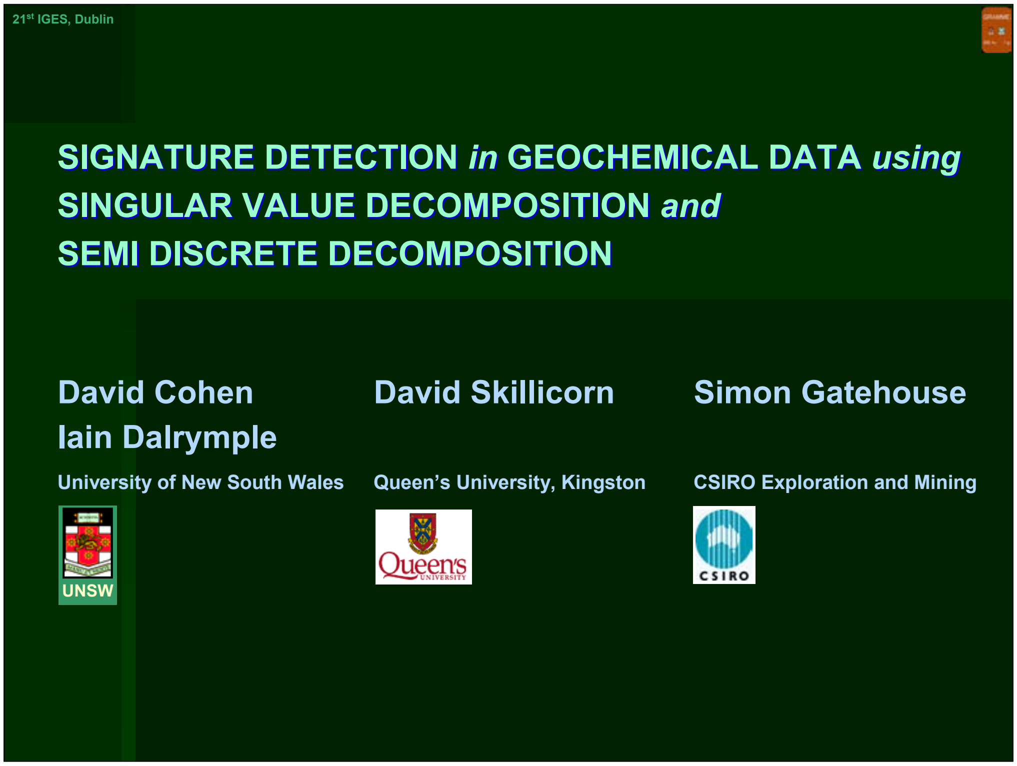
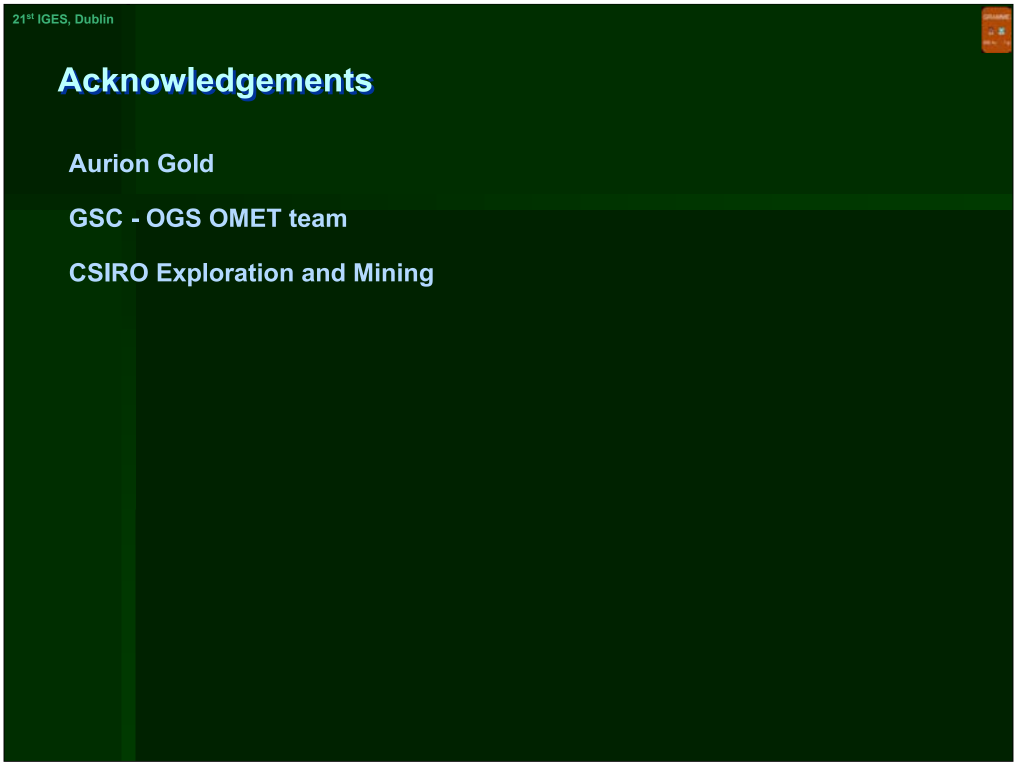


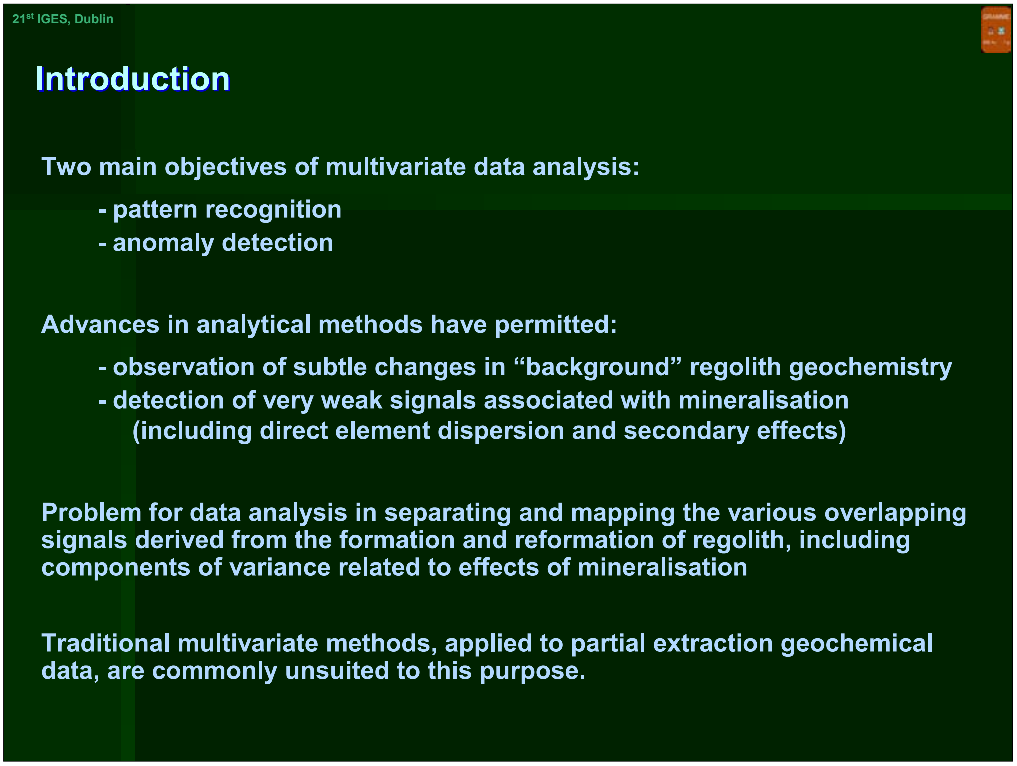

We present the results of a study into the application of a relatively new multivariate data analysis technique -

**Semi Discrete Decomposition**

- to surface regolith geochemical data in areas where mineralisation is overlain by deep, transported cover.




We acknowledge the support of a number of organisations for supply of cash, chemistry and critiques.




The two main objectives of multivariate data analysis in geochemical exploration are:

(a) **pattern recognition** and

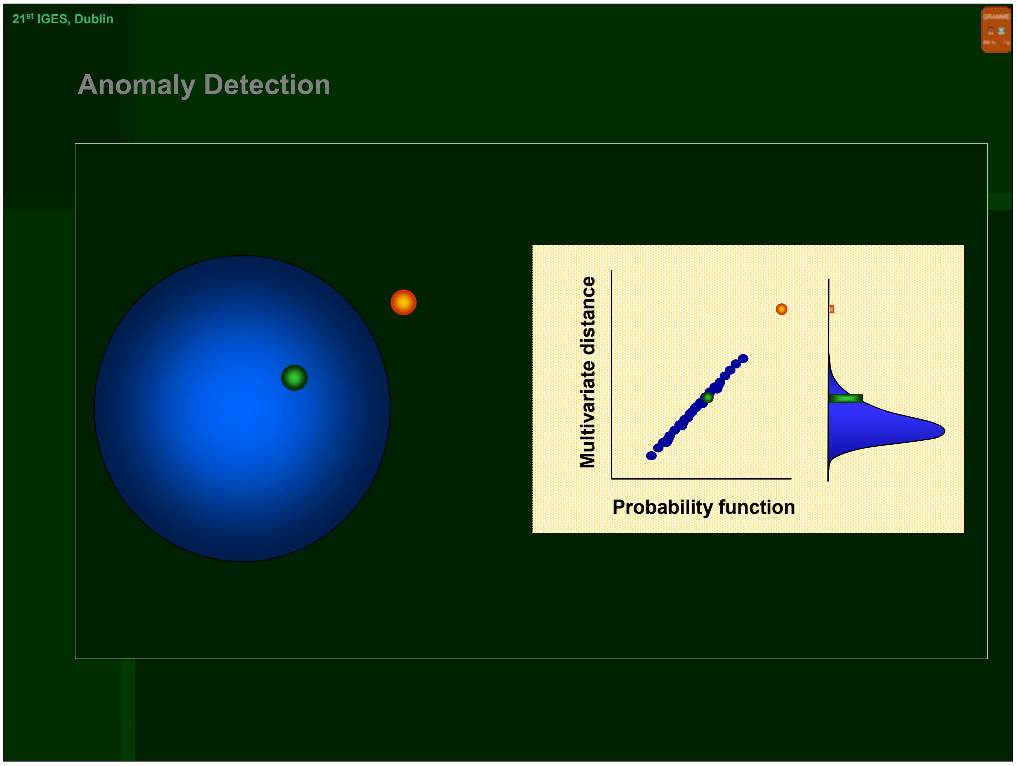
(b) **anomaly detection**,

although these objectives are closely linked.

As highlighted by a number of papers at this conference, advances in analytical techniques now offer the possibility of measuring the true extent of variations in “background” geochemistry of the regolith as well as the hope of detecting weak and transitory dispersion haloes through deep cover.

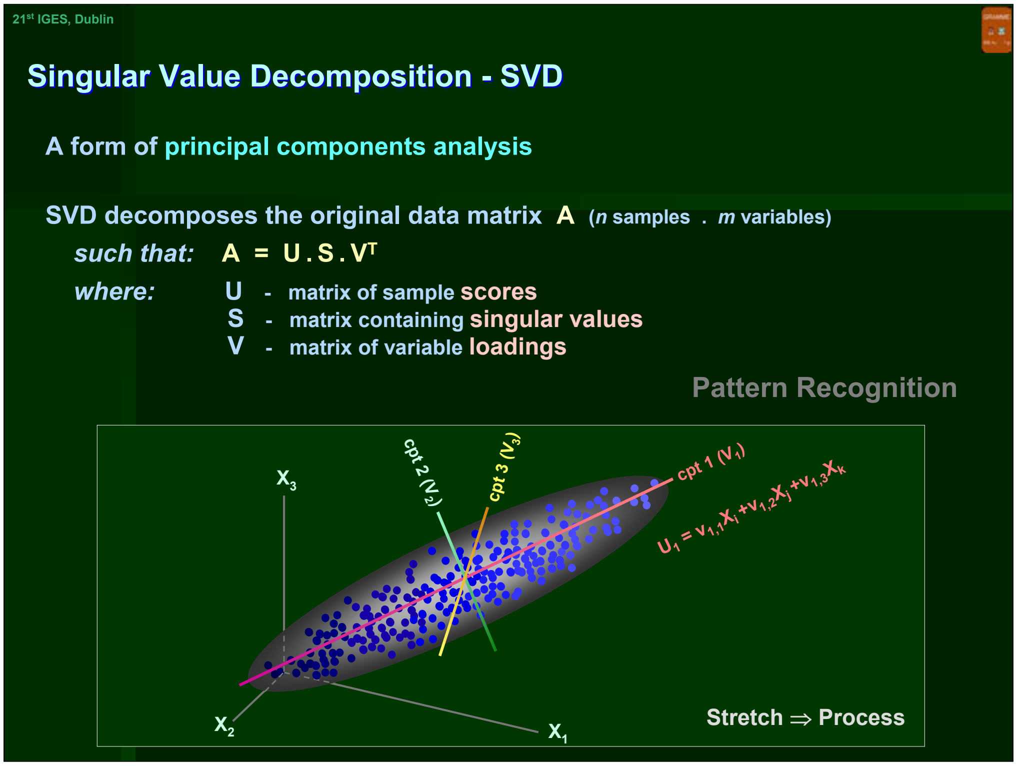
Severe problems arise, however, in separating signals or patterns related to the effects of mineralisation from the multitude of other processes involved in the formation and reformation of the regolith.

Traditional approaches to multivariate data processing have tended to focus on parametric methods related to either population modelling or the geometry of variance in data, which may not be suited to selective extraction data.




**“Anomalies”** are typically defined as points or clusters lying outside the main population (or high values).

But, such approaches progressively fail to detect anomalies as the distance between the main population and anomalous values decrease.



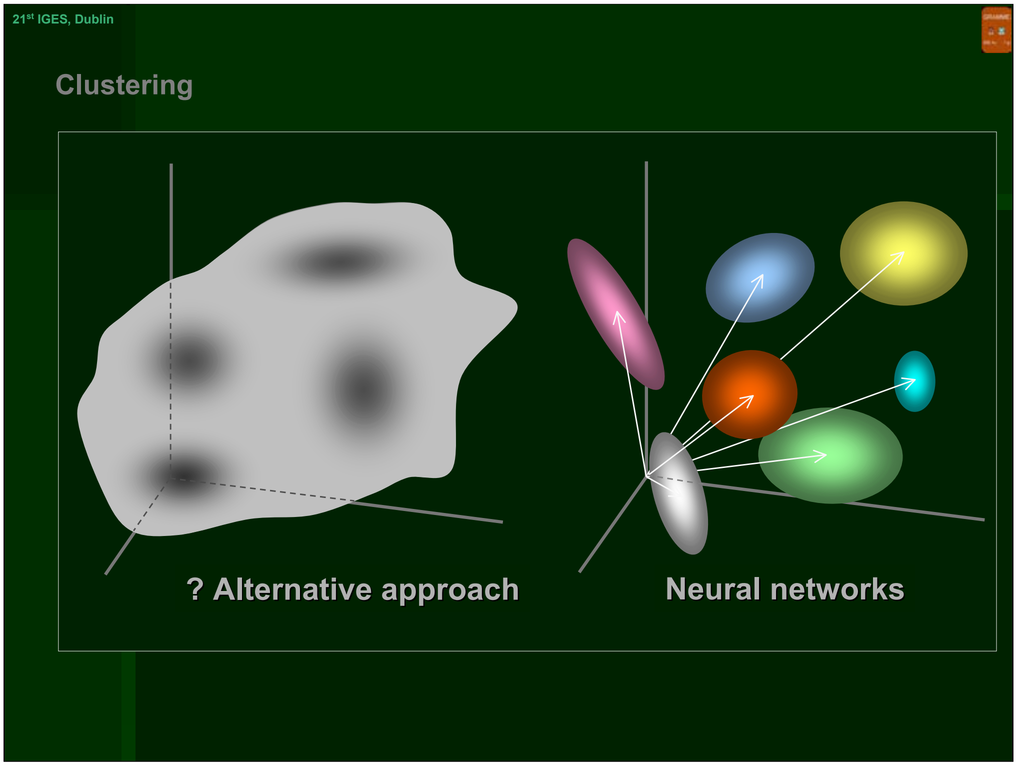

**"Patterns"** are often defined in terms of the observed stretching of data in space, for which principal components and factor analysis are commonly used in geochemical data processing.

One variant on principal components is **singular value decomposition** where the original data matrix (A) is **decomposed** to a matrix of new orthogonal vectors (V), representing linear combinations of the original variables, and a matrix of scores (U) for observations in the new orientation.

The dimensions can be reduced by substituting a lower value for m

The vectors commonly represent factors such as the influence of regolith composition on trace element distributions.

**But,** many of these methods display significant bias if the data are not symmetrically distributed or there are multiple data clusters.

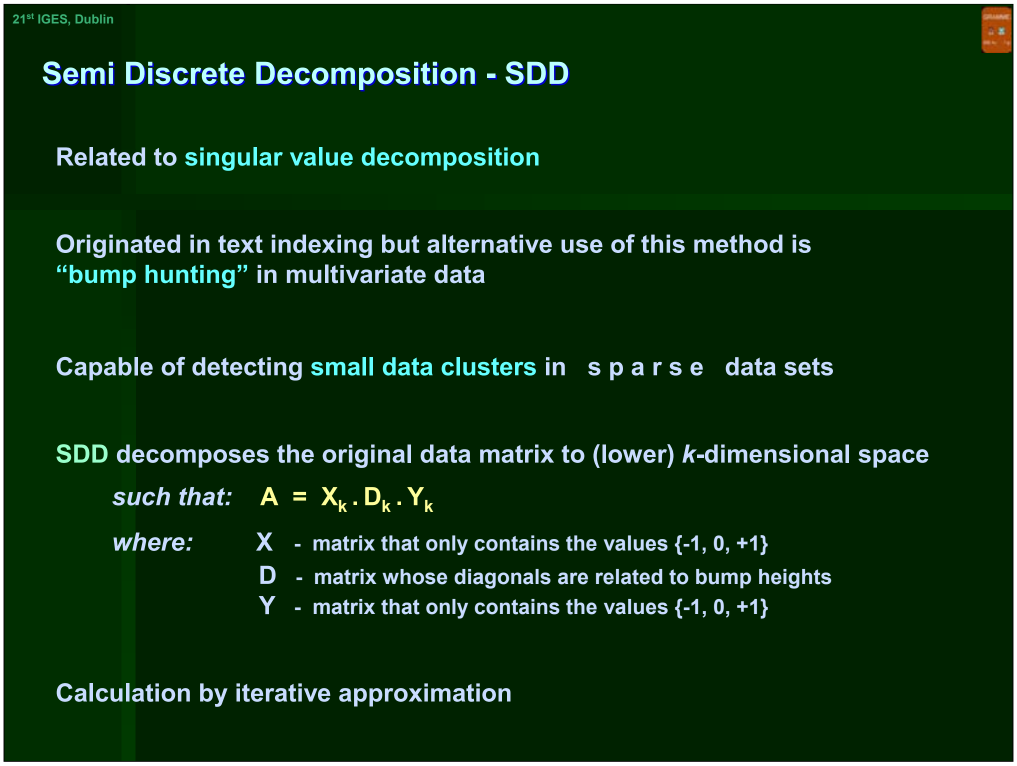



**“Clustering”** is commonly used to define both patterns in the data and also attempt to isolate anomalies.

Approaches include **k-means clustering**, in which a pre-determined number of boundaries are generated between clusters in the data, and more recently use of methods such as **neural networks** in both supervised and unsupervised forms.

Such methods cannot easily detect clusters or anomalies defined by subtle variations in data densities within a cloud of points, especially if such clusters are not geometrically regular shapes.

We were interested in trying an **alternative approach** whereby clusters could be simply defined as regions of increased density of data points, but where there are no underlying assumptions as to the mathematical form or geometry of the data.




One approach we have investigated is the combination of singular value decomposition and a method termed **"semi discrete decomposition"** or SDD.

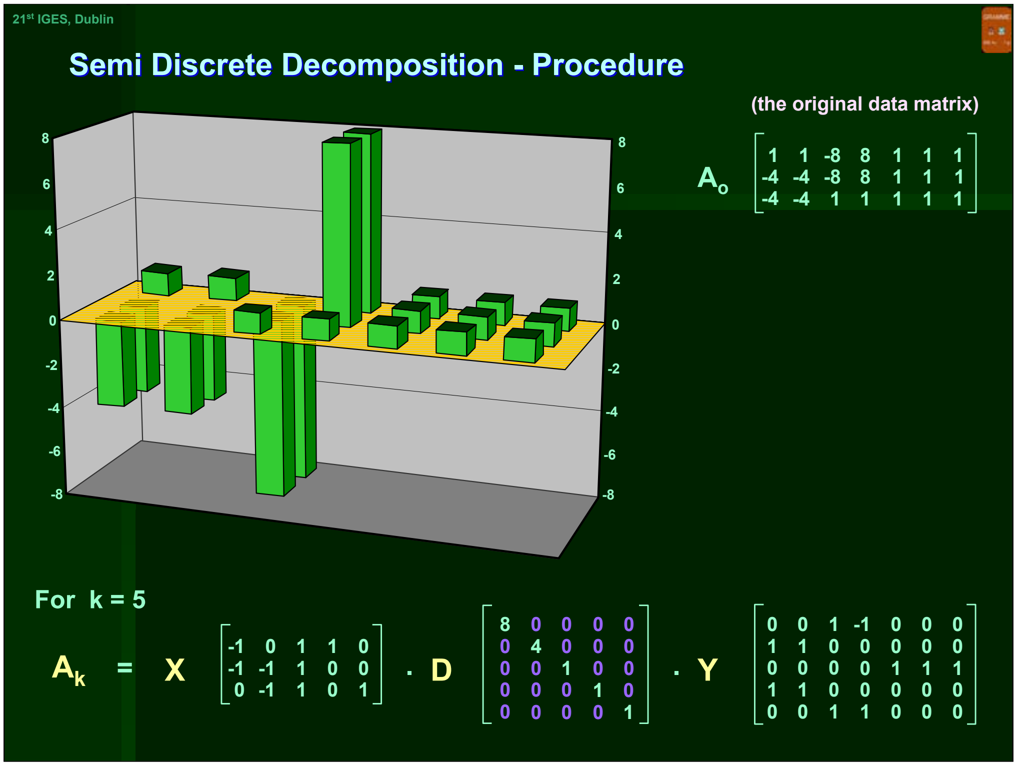
SDD originated in **text indexing** but is effectively a "**bump hunting**" or anomaly detection method.

It has recently been applied to the problem of detecting terrorists from eclectic data sources and used to demonstrate that true wine connoisseurs appreciate both red and white wines (from Australia).

Experience to data suggests its strength is in the ability to detect small clusters in sparse data matrices.

Like SVD, SDD attempts to reconstruct the original data matrix (A) into lower dimensional space (k) such that A = X.D.Y except that in this case X and Y are only permitted values of {-1,0,+1} and D represents the heights of the "bumps".

The solution to the equation is arrived at by iterative approximation.  Any resemblance to voodoo is denied.



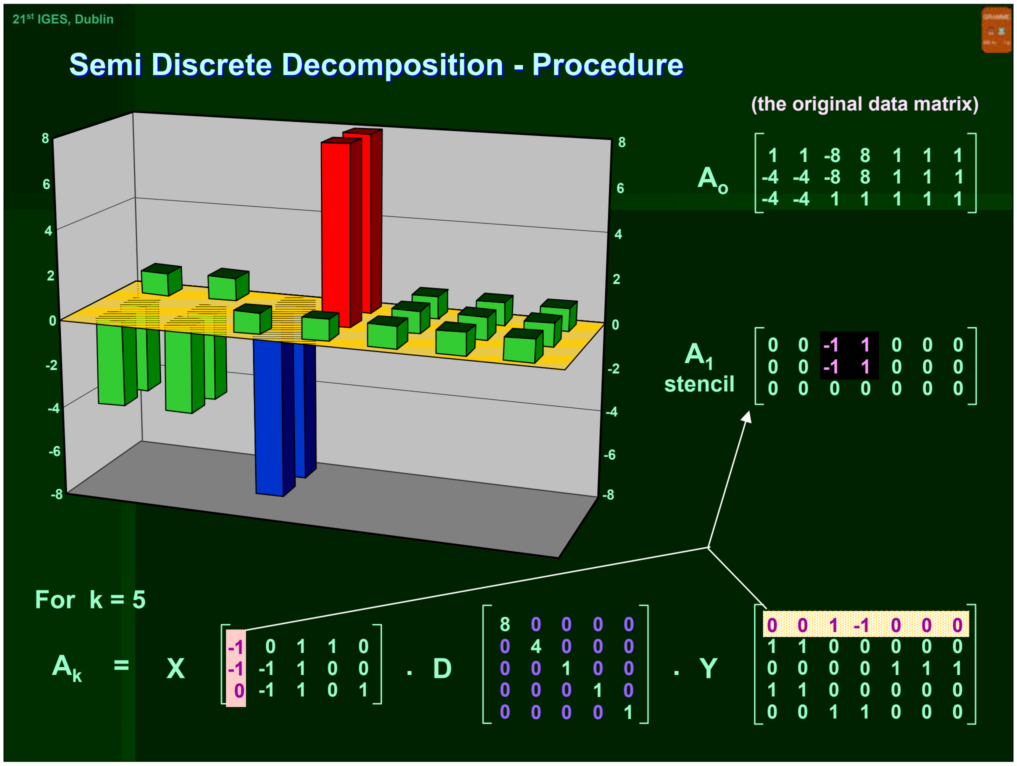

Let me provide a very simplified example of SDD operations.

We have a set of integer observations, laid out in both matrix and city-block forms.

Setting k=5, the matrix solution is provided.

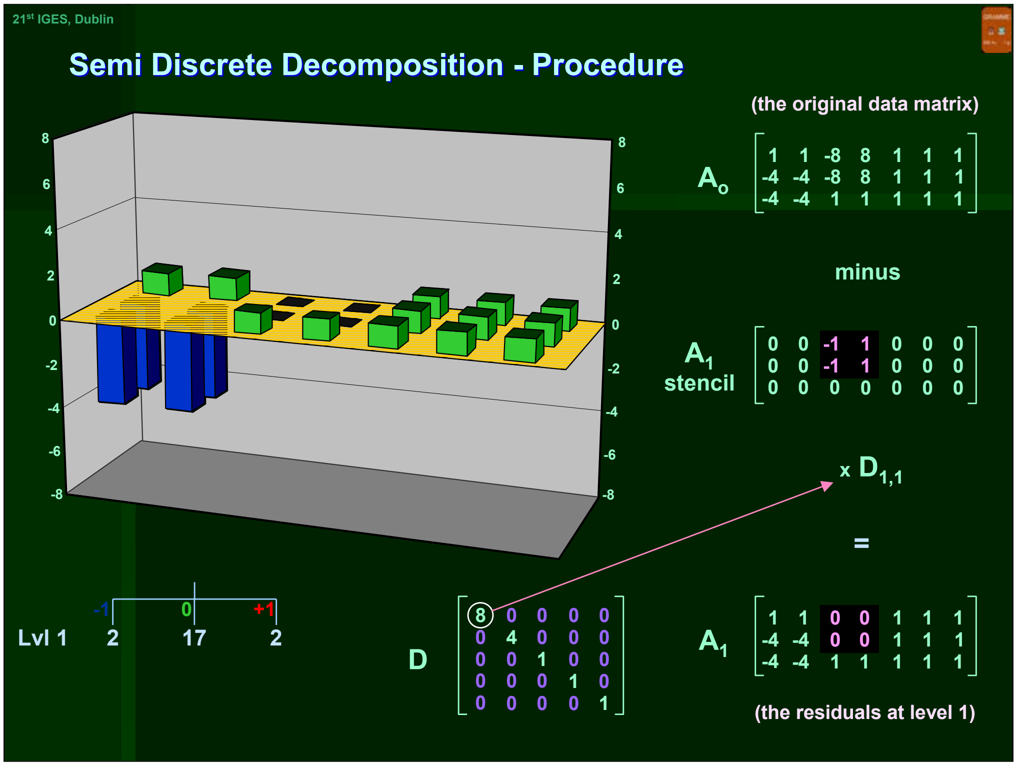
The diagonals of D represent the heights of the various levels in the data. The “8” and “4” are obvious, but not the series of 1s.

In real data, the bumps would contain a range of values and the heights would be the average for the cluster




The first level of clustering is given by the product of the first column of X and first row of Y

The stencil has allocated all observations into one of three categories [-1,0,+1] depending on whether the observations are below, equal or above the main "plane" of observations.

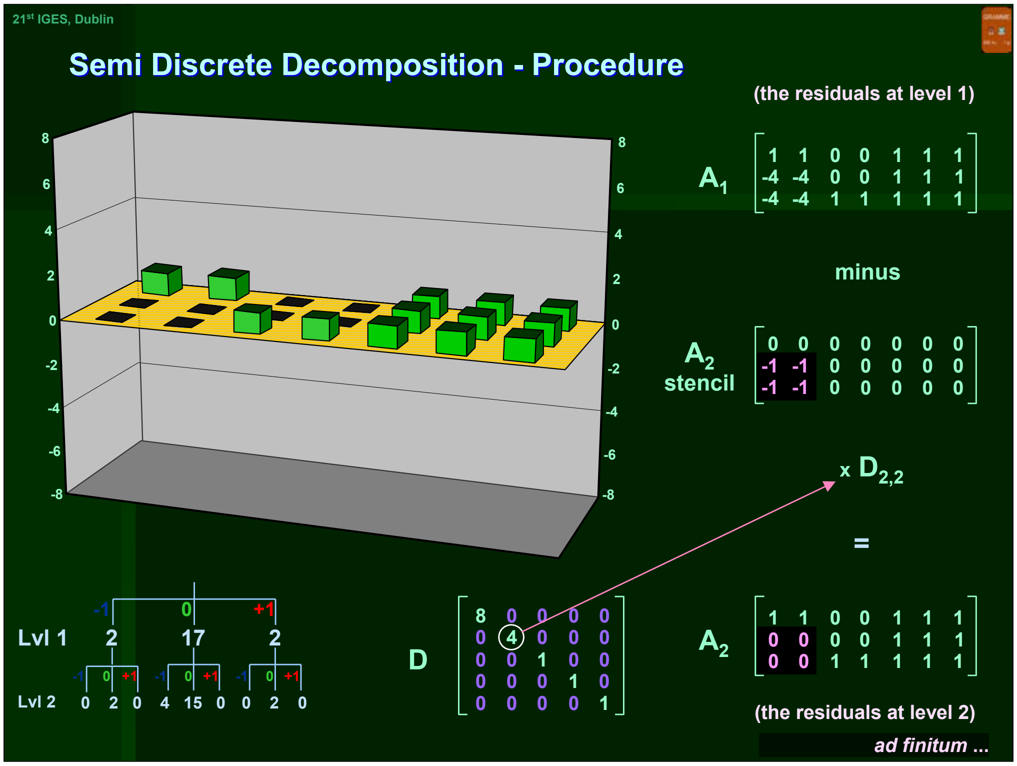



When the first level stencil is multiplied by the first element of D and this is subtracted from the original matrix of values, the heights of the two positive and two negative towers are sent to zero.

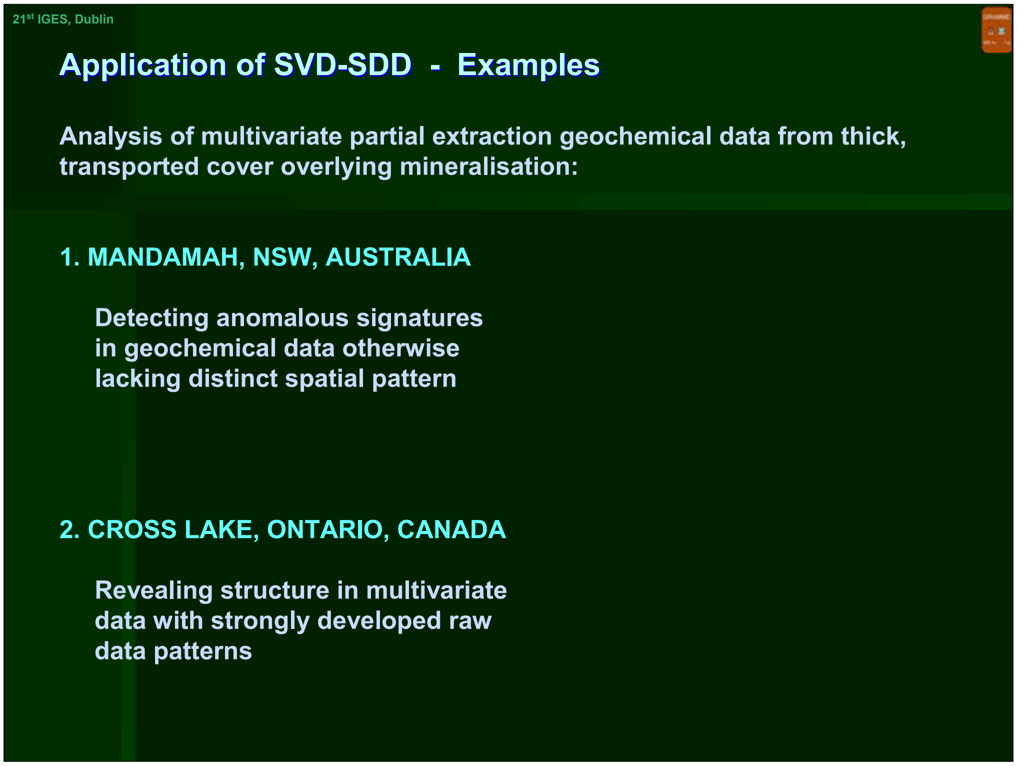
The remaining values left unaffected (in real data there would be some positive and negative residuals residuals).

At the first SDD level there are 2 x “-1”, 17 x “0” and 2 x “1”

The next level identifies the block of four negative values indicated



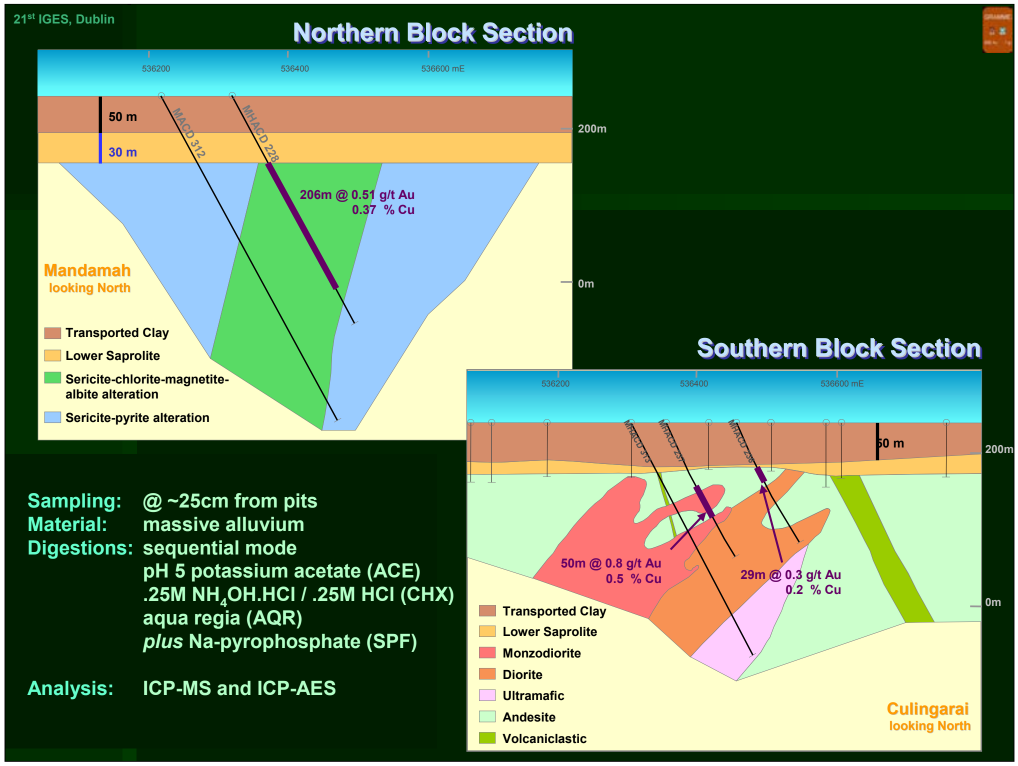

And this proceeds from level….




To demonstrate the capabilities of the SVD-SDD method at anomaly detection and pattern recognition, we have examined two selective extractions datasets from areas of thick, transported cover overlying base-metal mineralisation.

**Mandamah**

**Cross Lake**




At Mandamah and extensive drilling program revealed two pods of sub-economic porphyry-style mineralisation within a volcanic sequence.

The area is overlain by over 50 m of slightly indurated Tertiary silts, that in turn overly a variably thick saprolite.

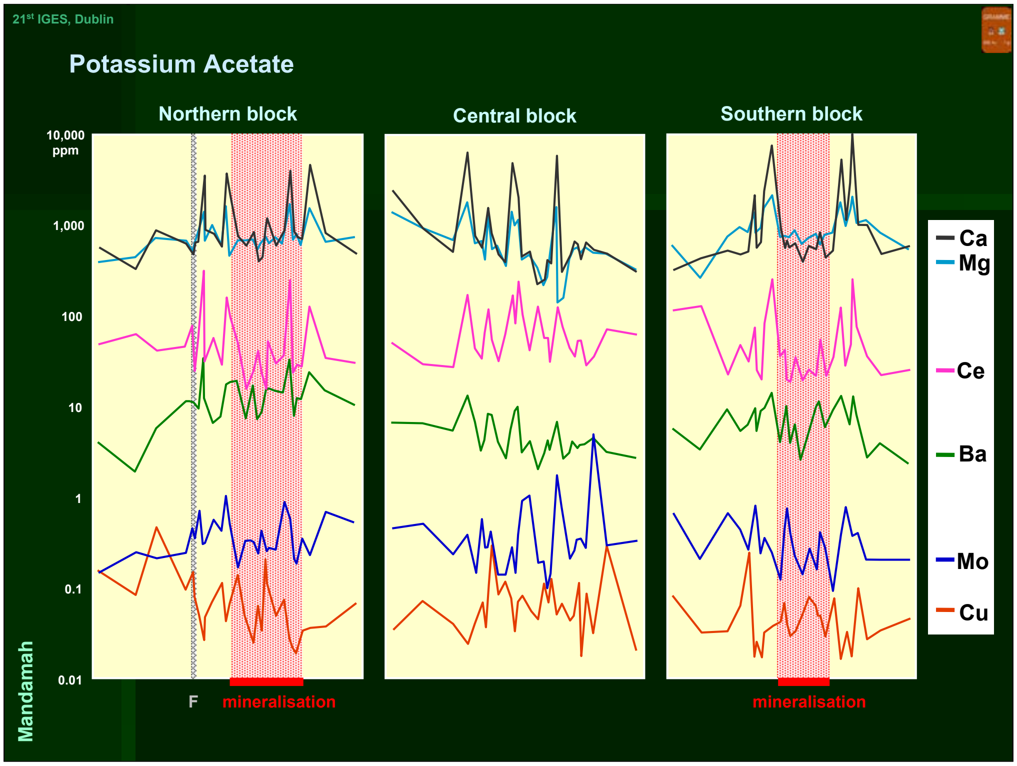
There was little variation in the characteristics of the silts, apart from the amount of pedogenic carbonate (0 to ~2 %) and some slight variation in the clay:quartz abundance.

Such regolith style is common in the LFB, especially to the north of NSW where mineral exploration has proven very difficult.

The montmorillonite content of soils in this agricultural region rendered the digging of a 50 cm deep hole, a 15 minute task with a 20 kg crow bar.

Sampling was conducted on three blocks, the north and south containing mineralisation and a central block not shown to have mineralisation

The upper sample from around 25cm depth was subjected to three digestions in sequence (ACE, CHX and AQR), with a second split subjected to SPF extraction.

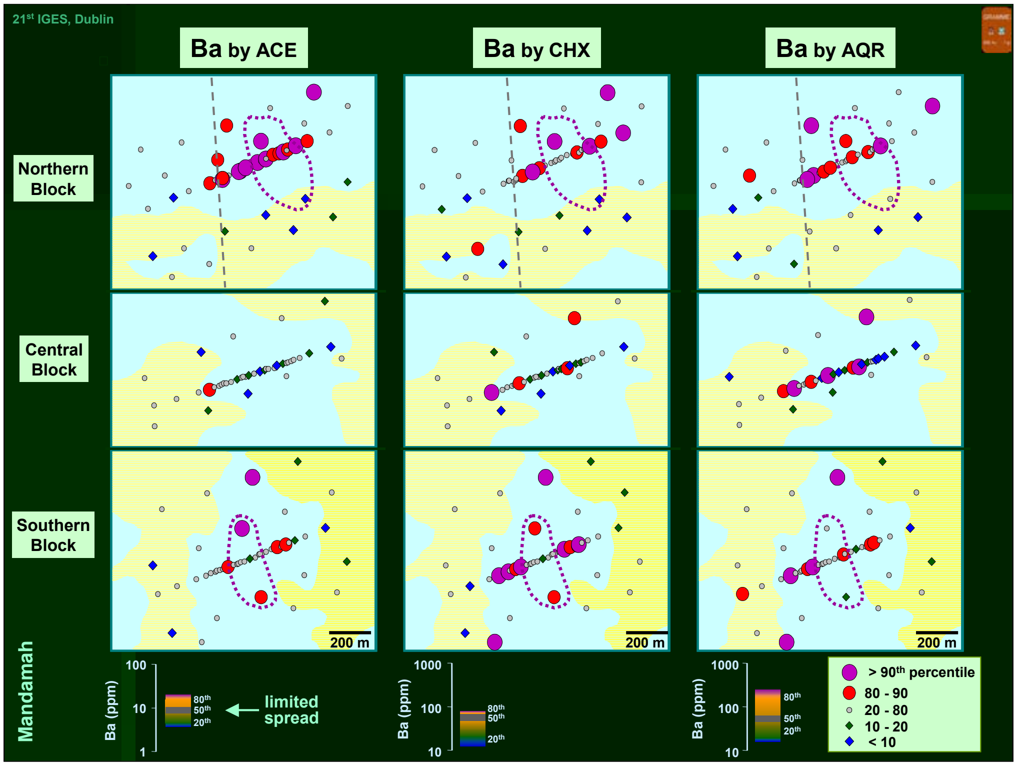



Looking at ACE for the three main sampling lines in profile, the feature that stands out is the rabbit-ears effect for Ca, Mg and the REE above mineralisation.

Similar patterns, however, exist in the central block (except for the REE).

Barium contents were also weakly elevated above mineralisation.

The remaining elements, including Mo, Au, Cu and pH measurements on soil slurries, displayed no obvious relationships to mineralisation.

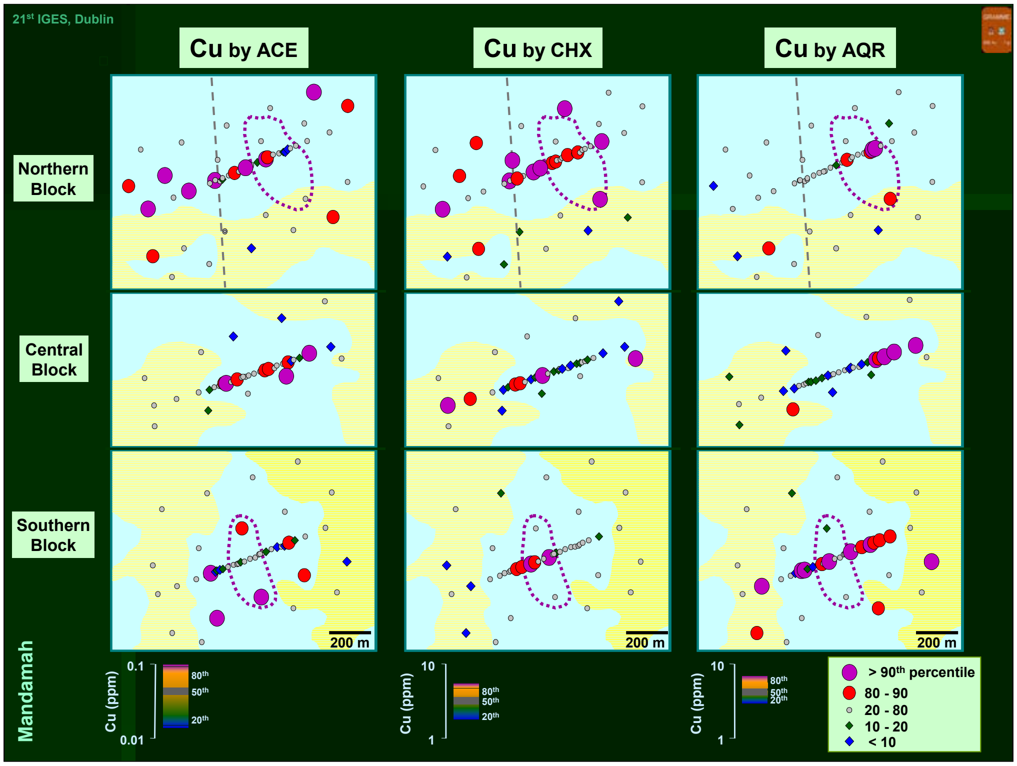



Looking at some variables in plan view, only Ba (and Ca+Mg) appear to display much of a relationship to mineralisation.

In ACE, Ba shows a weak accumulation above mineralisation in the northern block, elevated values at the edge of mineralisation in the southern block and only one elevated value in the central block.

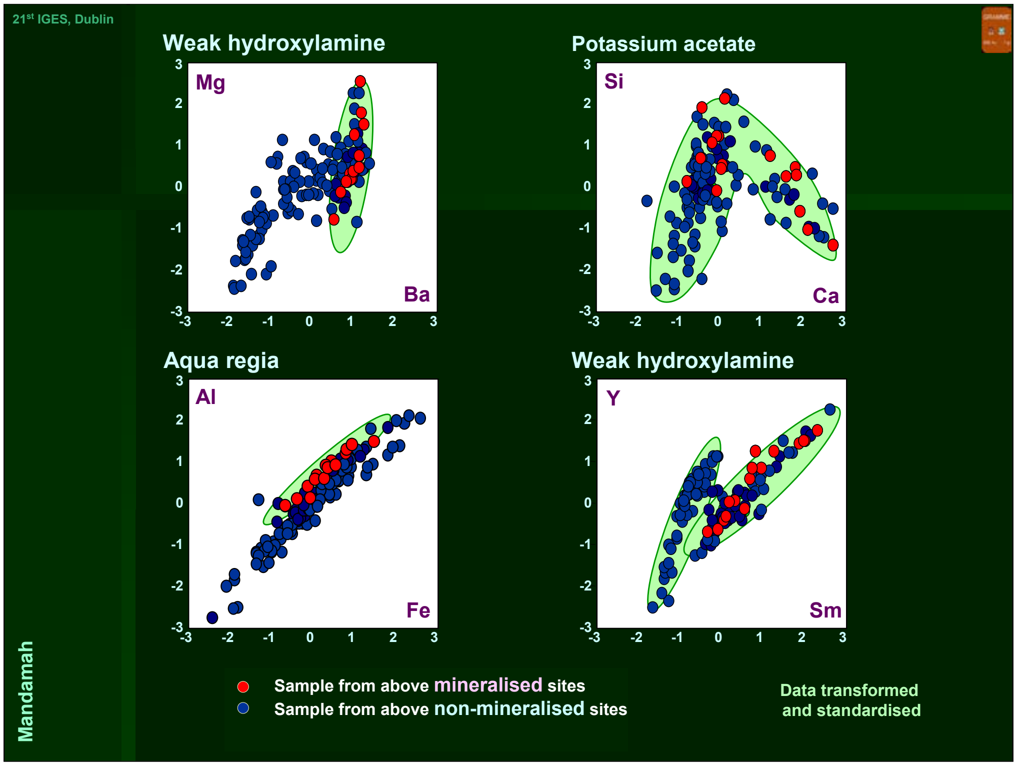
Both CHX and AQR extracted further amounts of Ba, with a donut around mineralisation but a series of elevated values in the central block.

It should be noted that there was little evidence of Ba in the primary mineralisation.




Copper shows very little relationship to mineralisation or regolith type, though one can use the indicated locations of mineralisation to generate the illusion (delusion) of a pattern.

Other elements display similar lack of convincing geochemical response to mineralisation in any of the digestions, when viewing the raw data.




XY-plots display some interesting behaviour.  The data has been coded according to location of mineralisation and regolith age.

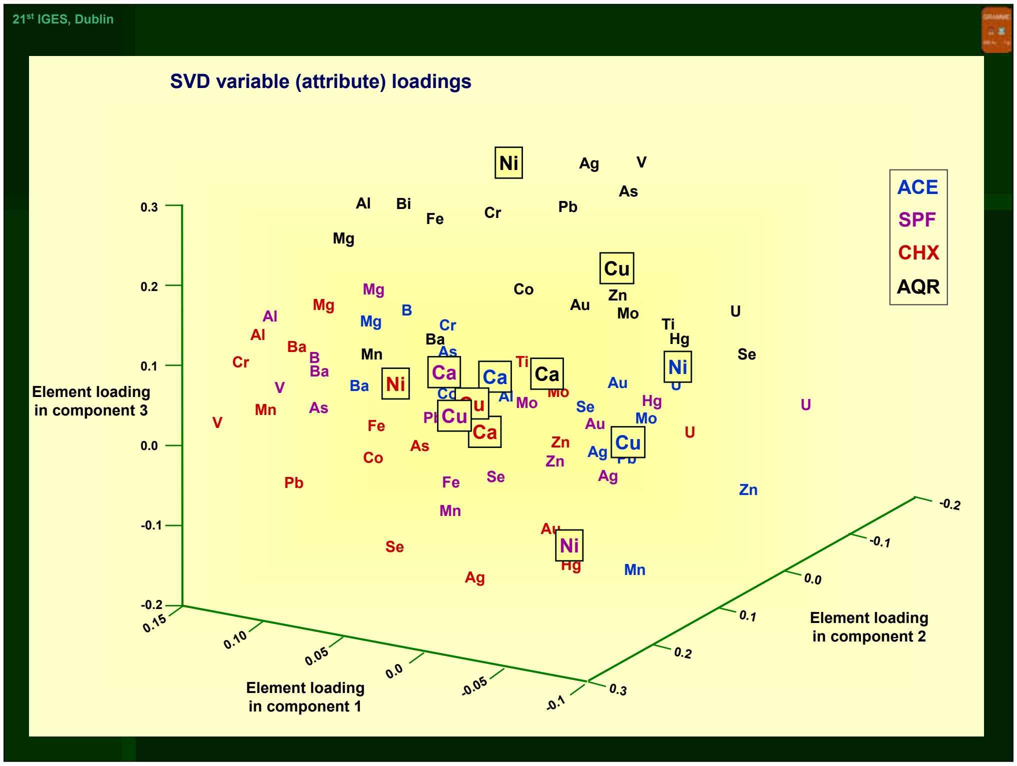
Ba generally displays higher values over mineralisation.

Si versus Ca by ACE displays two opposing trends, depending on whether the Ca is dominantly in carbonates or clays.

For Al versus Fe by AQR, the samples above mineralisation appear as a selvage on the upper side of the main correlation trend, suggesting the effect of mineralisation is to add a fixed amount of some elements to the samples.

REE display the most unusual patterns, including Y versus Sm by CHX where two trends appear and samples from mineralisation occur along only one trend.

Traditional multivariate methods for clustering multivariate data are not designed to cope with such complex patterns, however their existence suggests some opportunity for non-conventional techniques capable of teasing out different clusters relating to the presence of mineralisation or regolith variations.



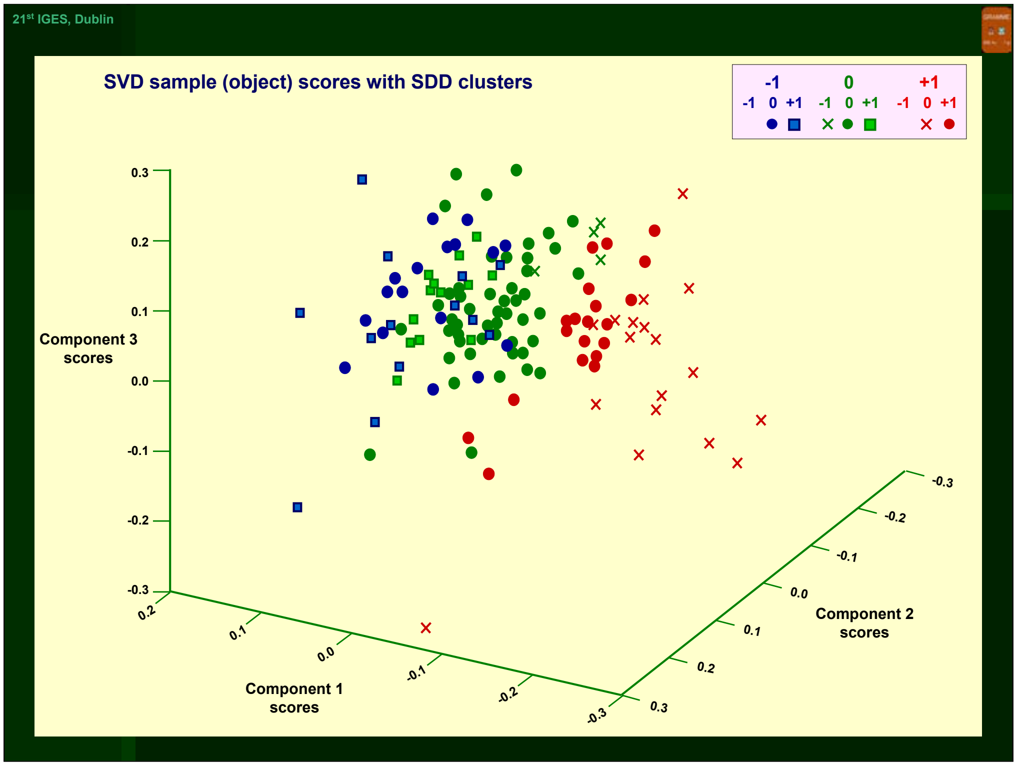

Looking at variable loadings under SVD, following simultaneous processing of all the data.

ACE and SPF plot closely indicating these digestions are attacking similar amounts and probably sources of trace metals.

CHX plots towards the rim of the ACE-SPF region, whereas AQR plots in a distinct region indicating it is attacking a different source and pattern of metals.

The Ca loadings plot very close to each other, suggesting a common mineralogical source of the element, with each successive leach extracting just a bit more of the total Ca.

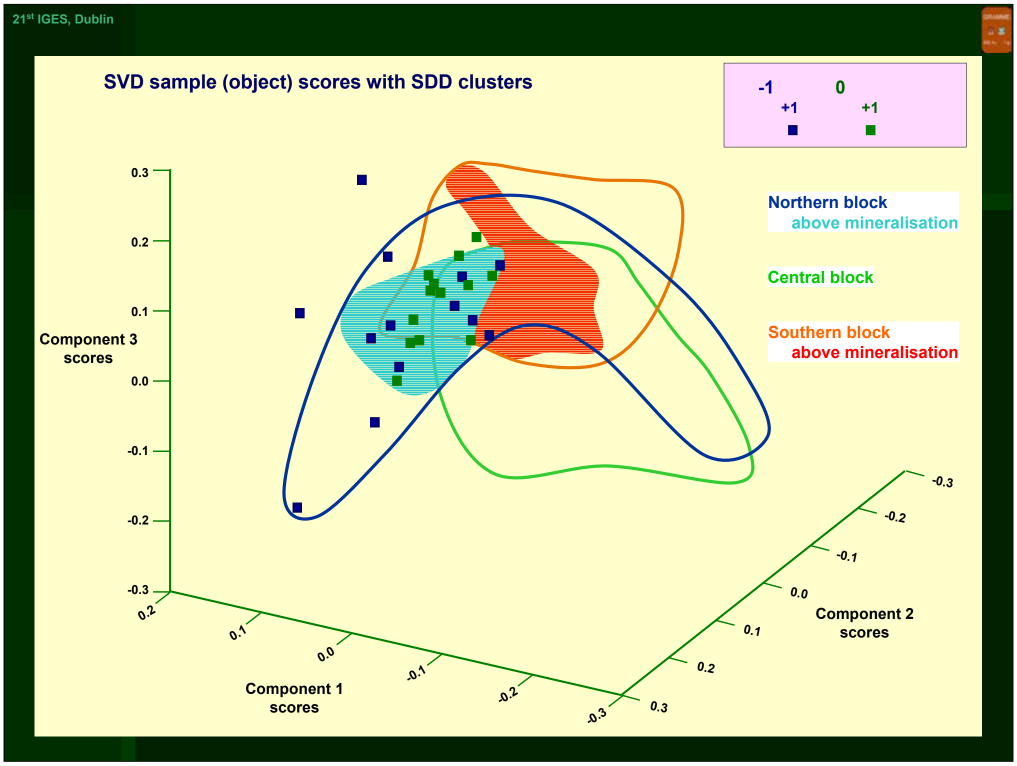
Cu displays more scatter and Ni high scatter, demonstrating a lack of correlation between Ni in the leaches and a distinct mineralogical associations and probably temporal origins for the Ni within the regolith.




Looking at the SVD-SDD structure for the complete Mandamah dataset, we commence with the plot of the first three SVD "component scores", coded according to SDD cluster for levels 1 (by colour) and 2 (by symbol).

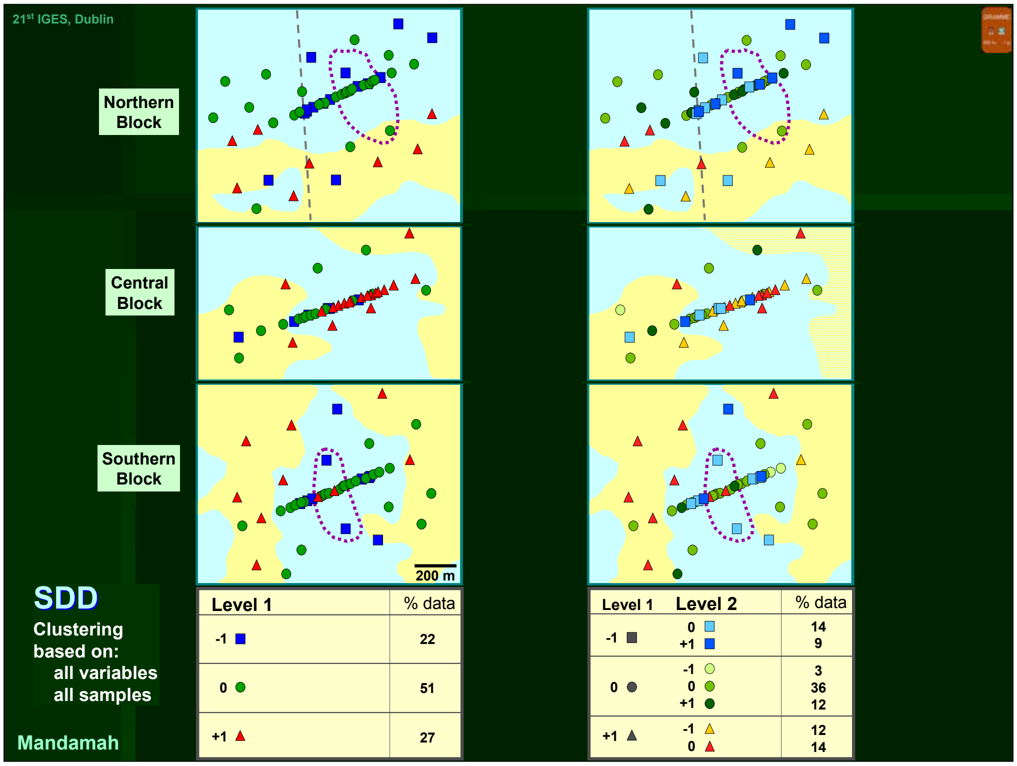
It is somewhat difficult to represent the data clouds in 2D.

At the first SDD level, there is strong separation between the clusters (in SVD component score space), with further spatially distinct at level 2.



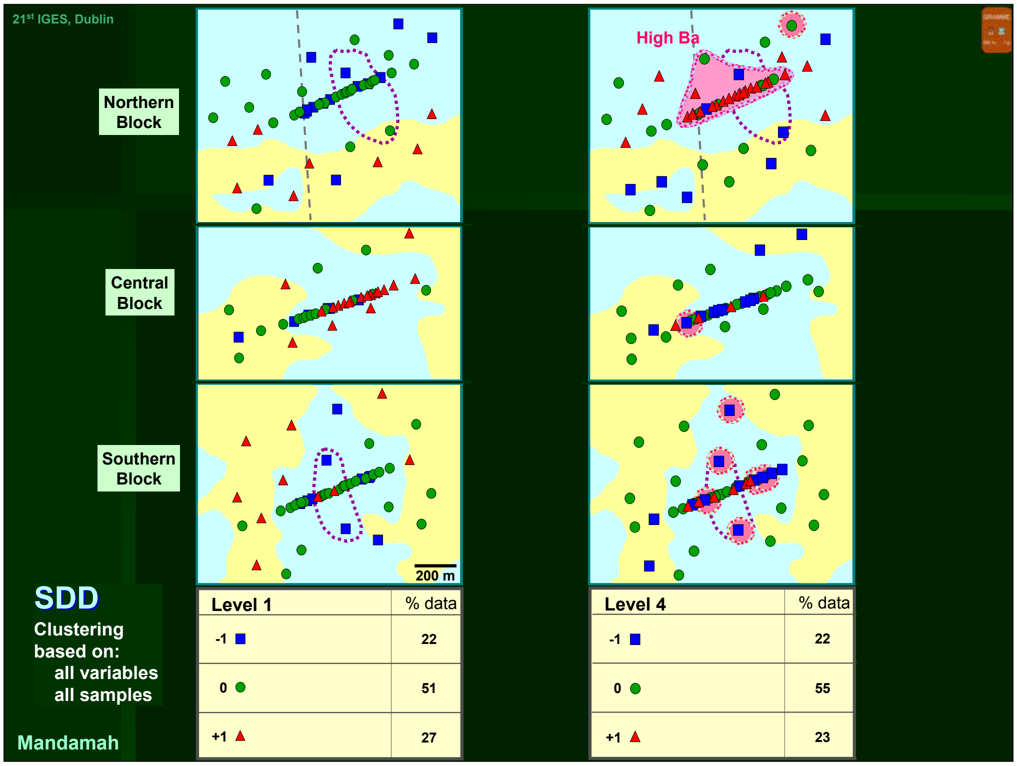

When the various groups of samples from the different blocks are overlain, there is initially no obvious relationship with the SDD clusters.

The closest relationship appears to be the {-1,+1} and {0,+1} clusters with the northern block mineralisation.



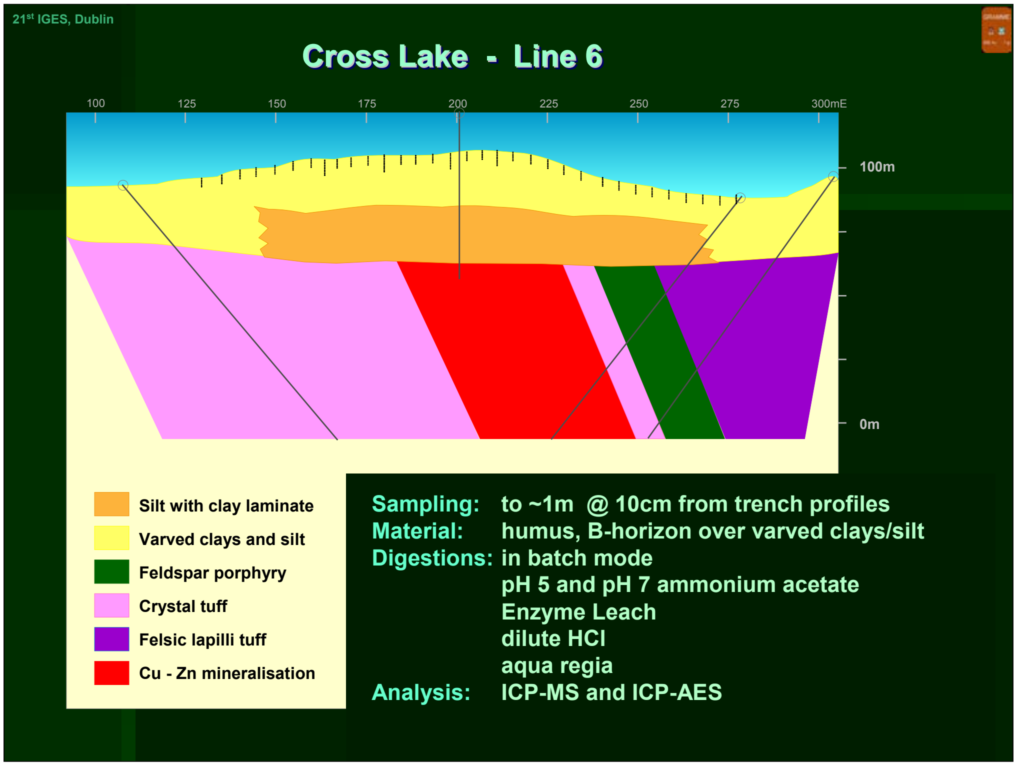

The Level 1 clusters do not deliver much information about the location of mineralisation, but the {+1} cluster appears mostly restricted to the younger regolith in the northern and southern blocks and a group in the middle of the central block.

If Level 1 is further subdivided to Level 2, there is no further light shed on the structure of the data, although the {0,+1} cluster is mainly restricted to the northern block.

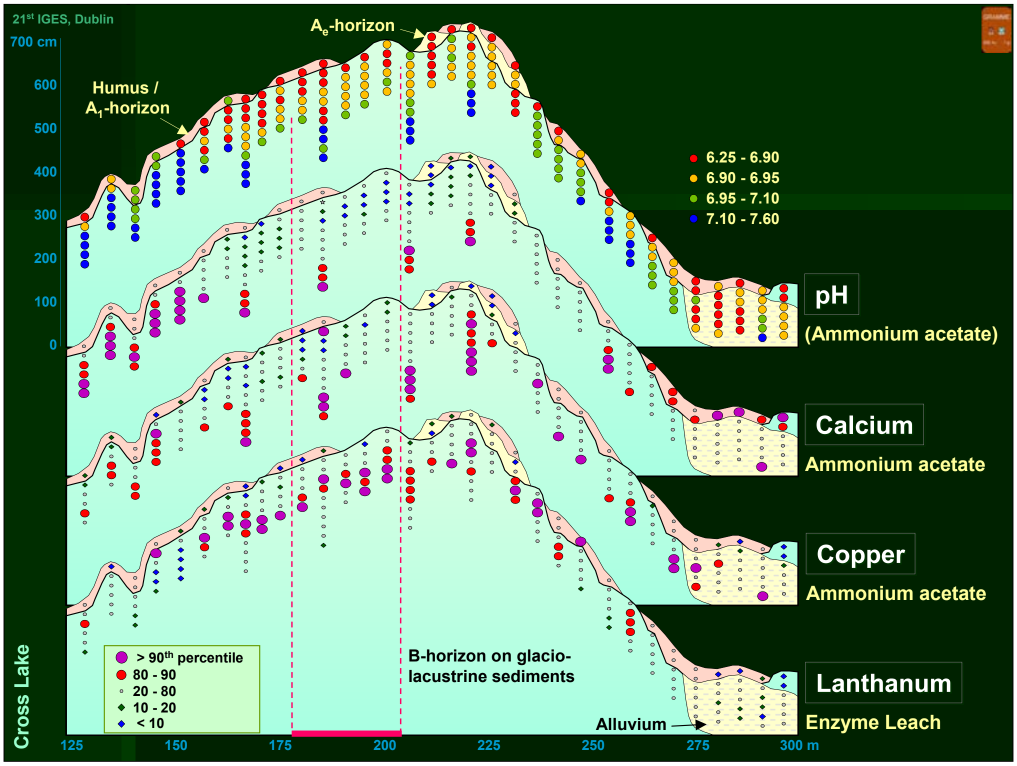



The "anomaly detection" capabilities become more apparent when looking at Level 4, in terms of its deviation from the previous level. Here we see that the $\{+1_{lvl4}\}$ cluster is mainly observed across mineralisation on the northern block (although some points do occur outside this area)

This zone is not completely coincident with the very weak Ba anomalies that appeared, at first pass, to best define mineralisation. Some of the slightly elevate Ba values are not within the level 4 {+1} cluster, and the converse.




As described by Stew Hamilton and Gwendy Hall, Cross Lake contains a zone of sub-economic Cu-Zn mineralisation within a felsic volcanic package, that has been overlain by around 30 m of glacio-lacustrine clays and silts, and some more recent alluvium. There is a thin humus layer. The watertable is close to surface. The region above mineralisation has strong redox and pH anomalies.



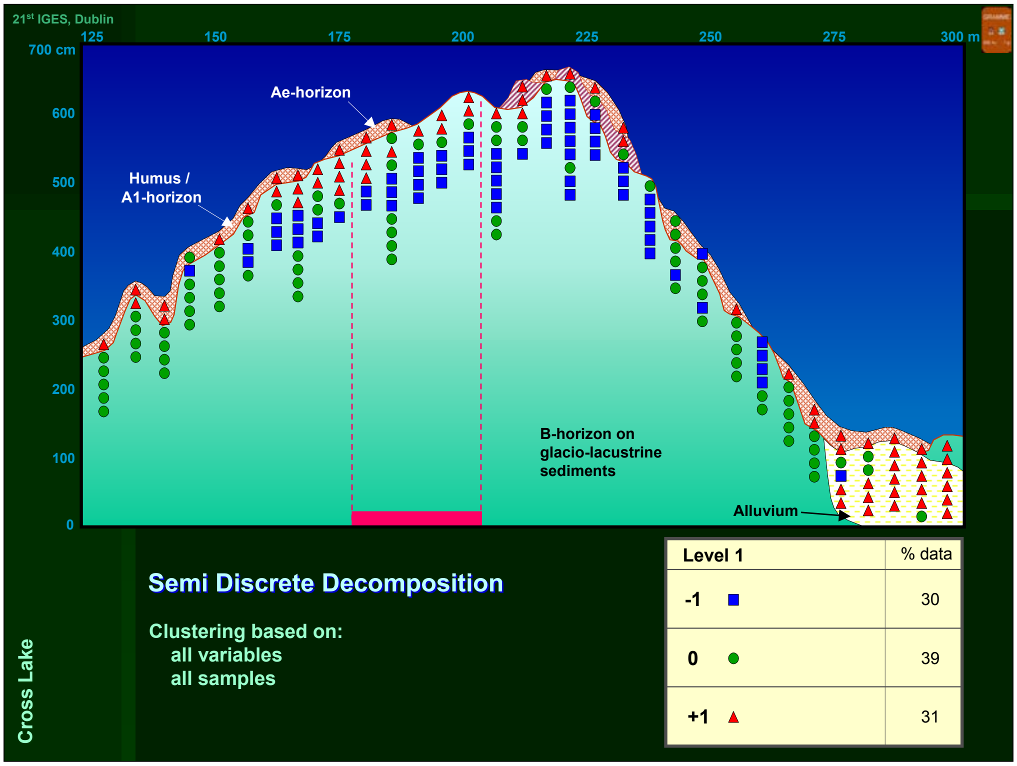

The vertical projection from mineralisation is characterised by a low pH zone in the upper 40 to 50 cm, although this also represents the northern side of a low-relief ridge (note the extreme vertical exaggeration). Low pH values are also present in the alluvium. The boundary also appears to coincide with the water table, although this would vary during the season.

The low pH zone is characterised by a near-absence of Ca in AAC-5 and but an accumulation to the sides of the cell (a simple redistribution of Ca?) and in the upper part of the alluvium.

AAC and ENZ Cu display some relative enrichment at depths below 30 cm across the entire site, with a zone of relative depletion on the left flank of the ridge.

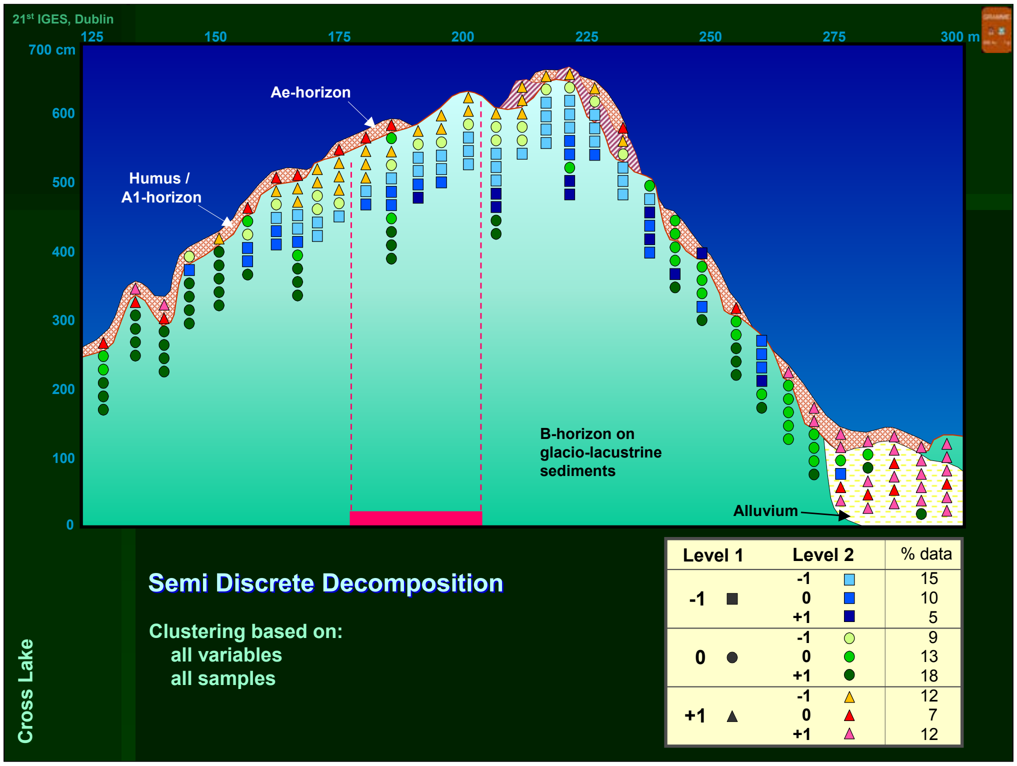
Lanthanum is strongly enriched around the 30 to 50 cm depth across the ridge.

While there is both strong vertical and horizontal (or surface-parallel) structure in the data, there is also a degree of noise.



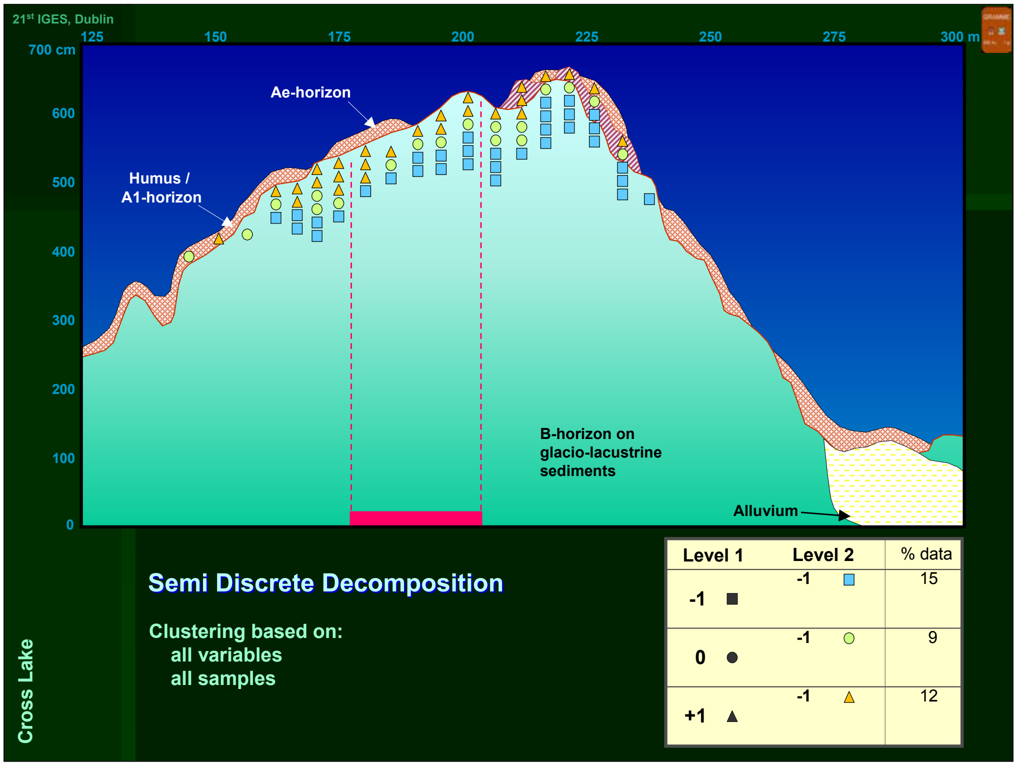

Looking at just the SDD for Cross Lake, we have combined all variables (apart from those with severe truncation) and all samples.

At the first level, the data clusters appear strongly influenced by sampling depth. Samples in the {-1} group form a lens over and immediately to the sides of mineralisation, surrounded by the {0} group. The {+1} group forms the upper layer, but not necessarily coincident with the humus layer, and extend into the alluvium.



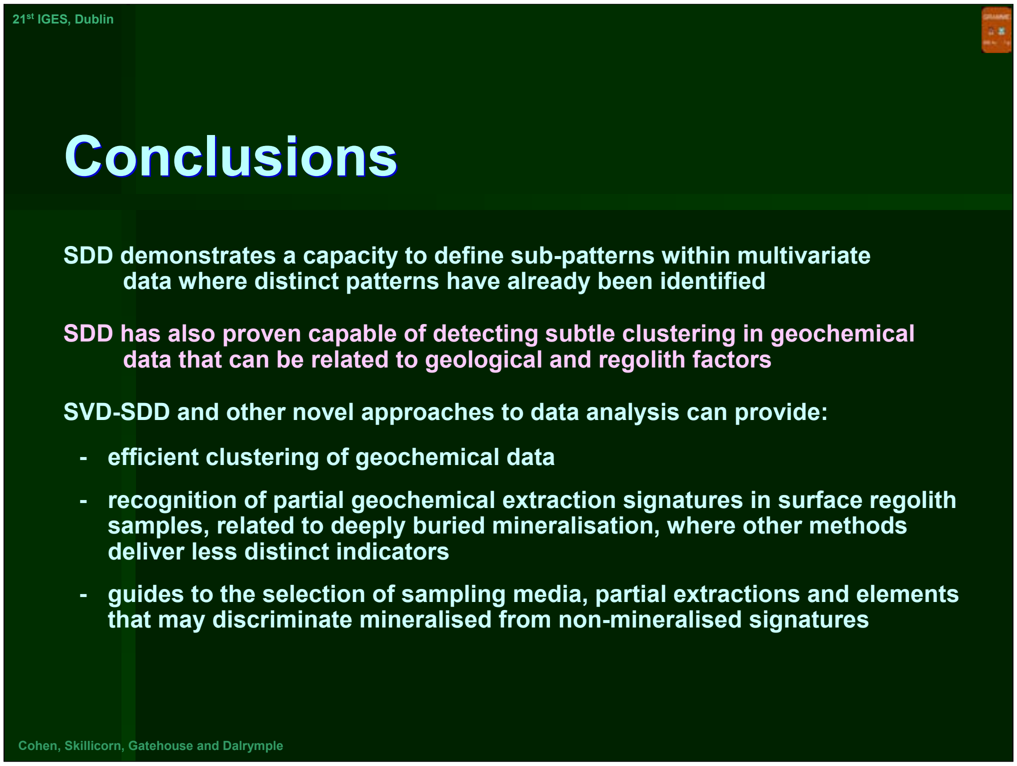

At the second level there is still very strong spatial continuity of the clusters. Both the {0} and {-1} clusters for level 1 have been subdivided into sample depth related zones. These appear to follow either topography, or more likely the position of the watertable.

The {+1} clusters of Level 1 are now divided into three sub-clusters; {+1,0} is mainly found in humus and at 40-50 cm depth in the alluvium; {+1,+1} is wholly contained within the alluvium and {+1,-1} occurs as a tight cluster of samples between 10 and 30 cm depth above the projected mineralisation.




If we just examine observations the are classified as {-1} at the second level, these form a vertical stack centred over the mineralisation and up to 25 m either side.

Note that there are no occurrences of these groups outside this very restricted zone.




[Read conclusions]

But as an aside, there is some frightening similarities between element distributions at Cross Lake and Mandamah.  We may even be convinced that reduced chimneys exist and can be preserved long after the onset of aridity.