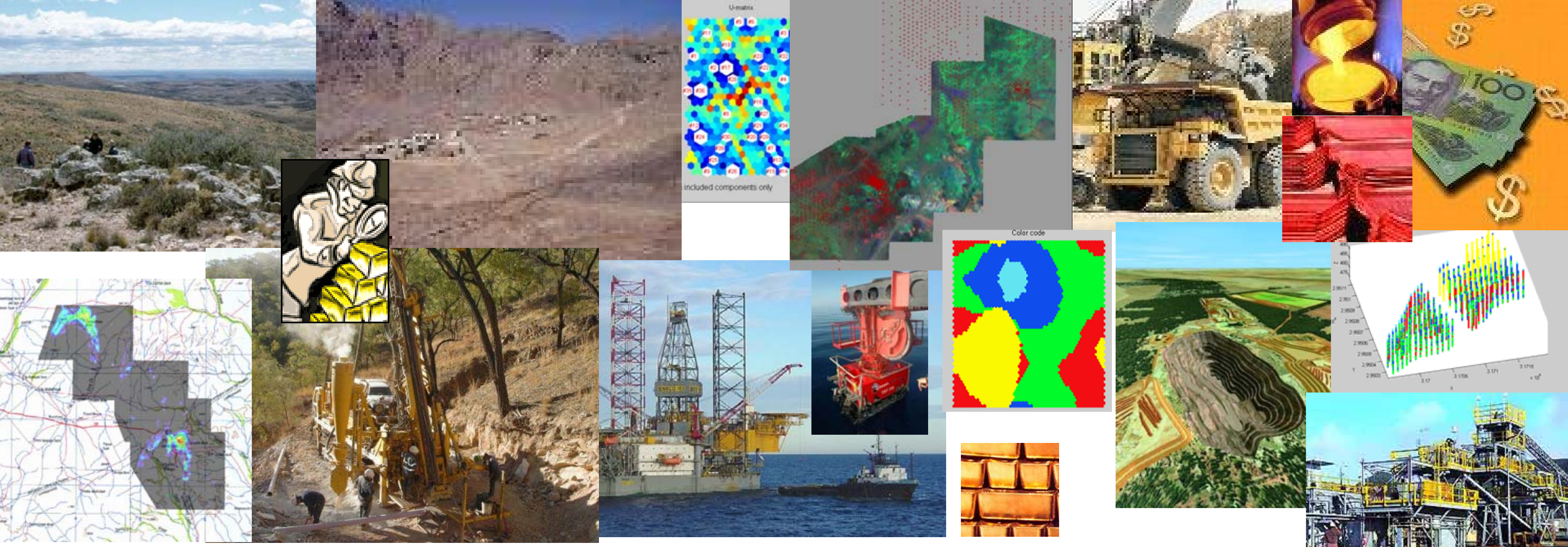
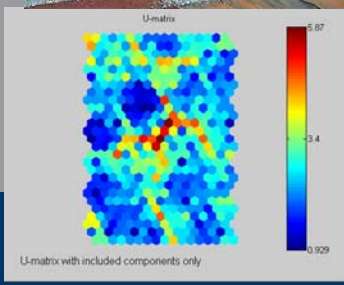


# Data Mining Geoscientific Data Sets Using Self Organizing Maps




Stephen Fraser[1] & Bruce Dickson[2]

[1] CSIRO Exploration & Mining

[2] Dickson Research Pty Ltd

# Background - Geo-Informatics

**Geoscientists gather data faster than it can be interpreted.**

**Data bases and GIS enable data storage and display; but do not resolve the issue:**

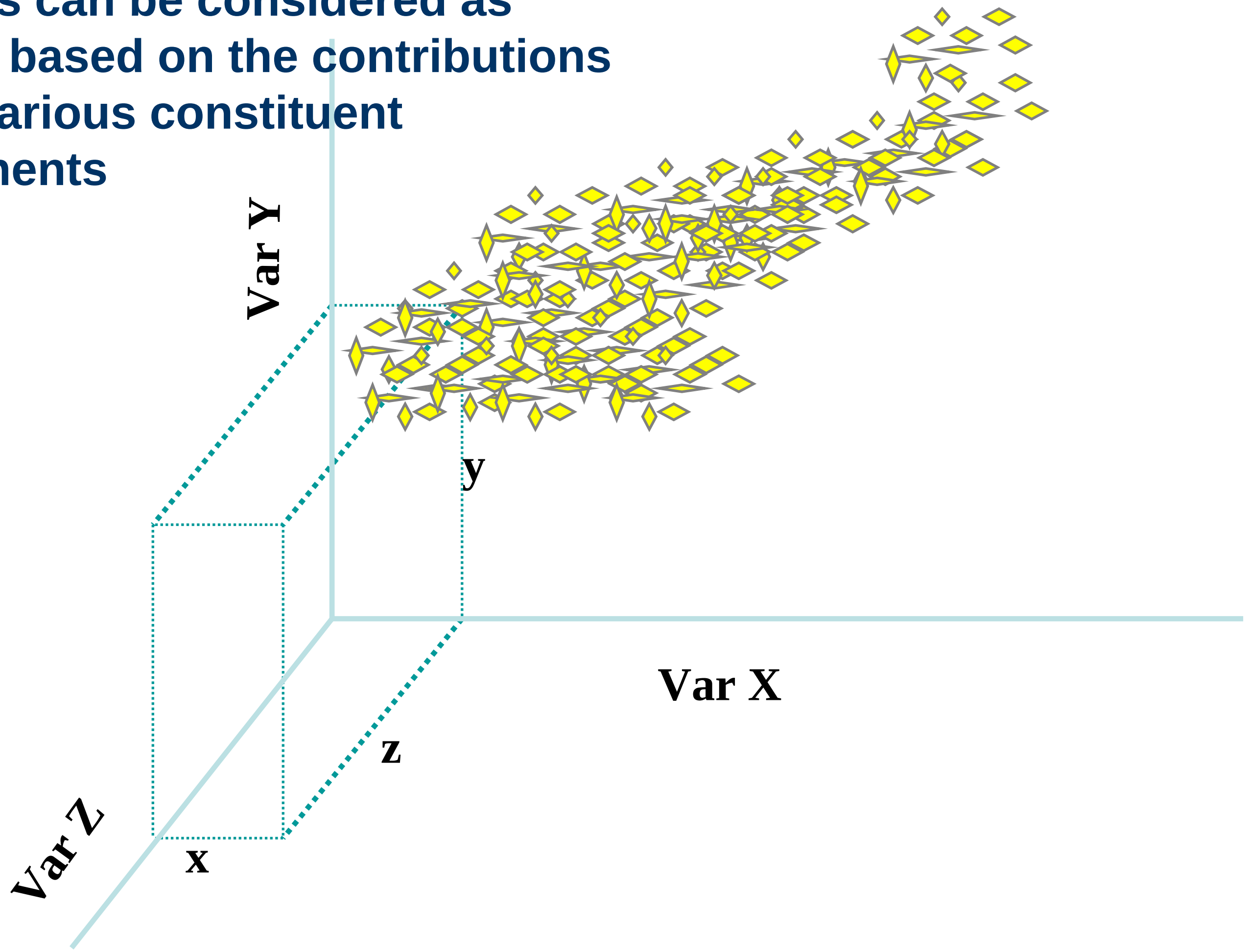
***"How do we intelligently analyze and interpret the volumes of data we collect?"***

**Computational Tools are needed to assist in the process that turns data into information, and information into knowledge.**

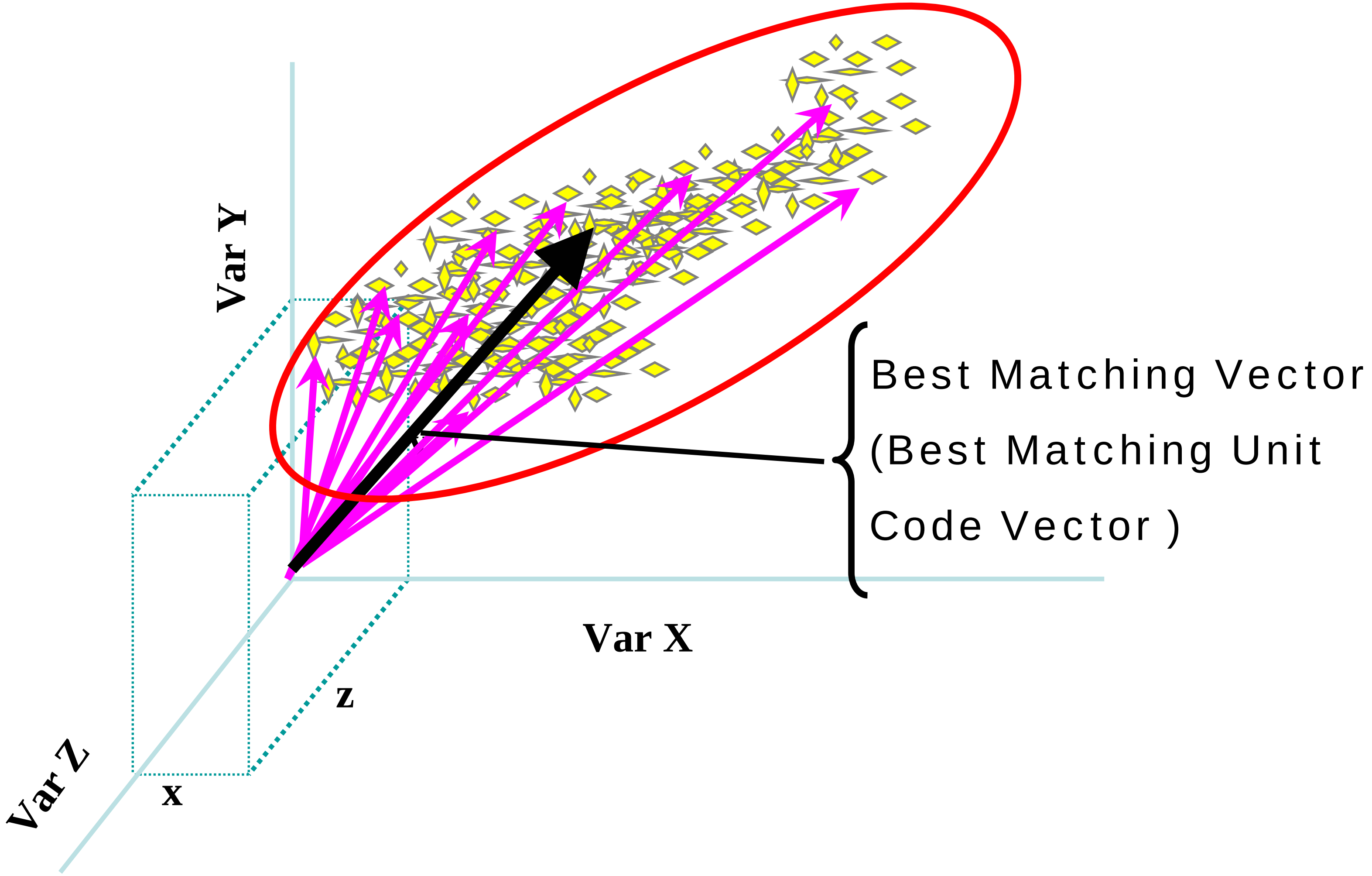
# SOM – Introduction #1

**Consider a grouping of similar/related samples in n-D space**

**Samples can be considered as vectors based on the contributions of the various constituent components**



# SOM – Introduction #1

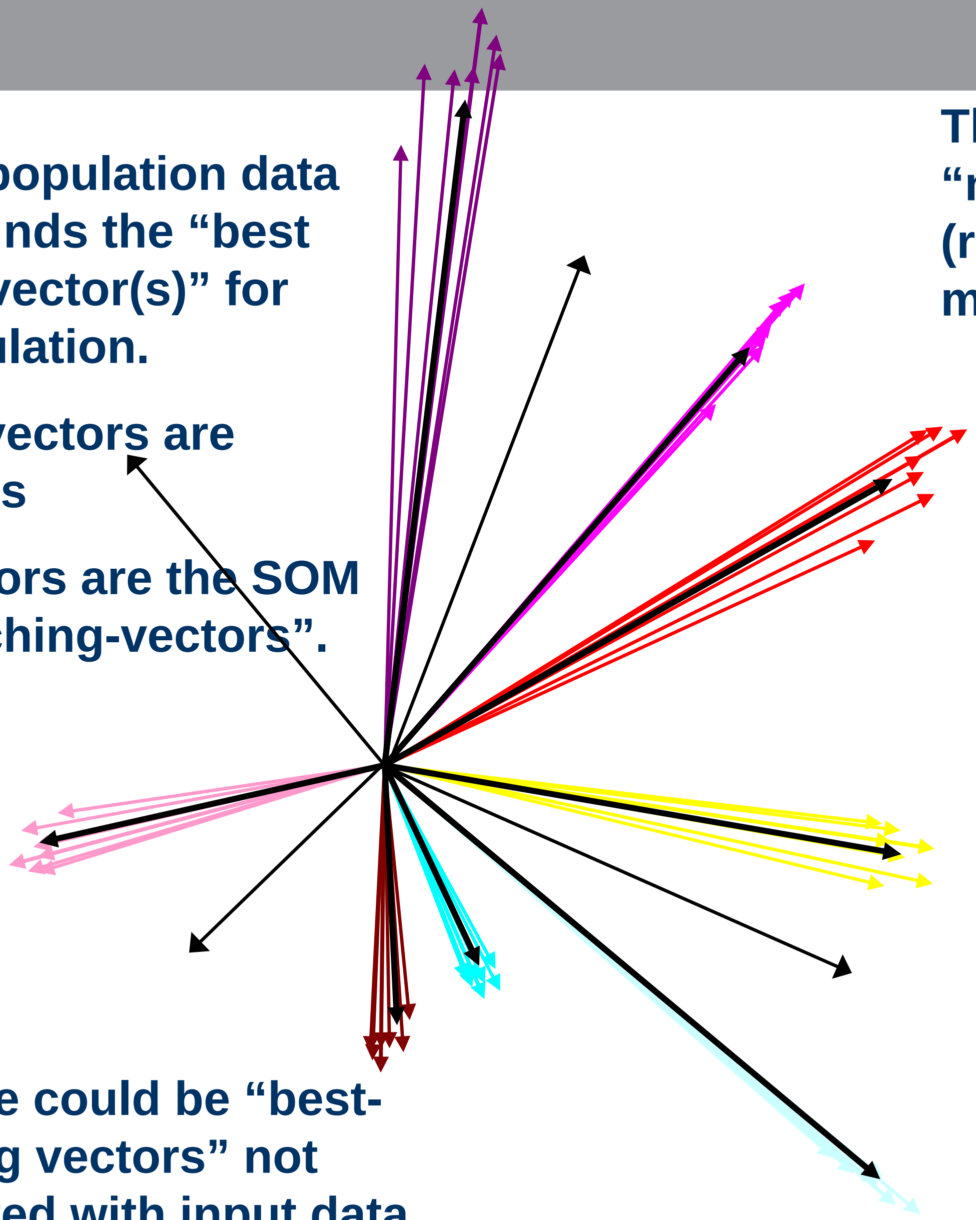


# SOM – Introduction #2

**In a multi-population data set, SOM finds the "best matching-vector(s)" for each  population.**

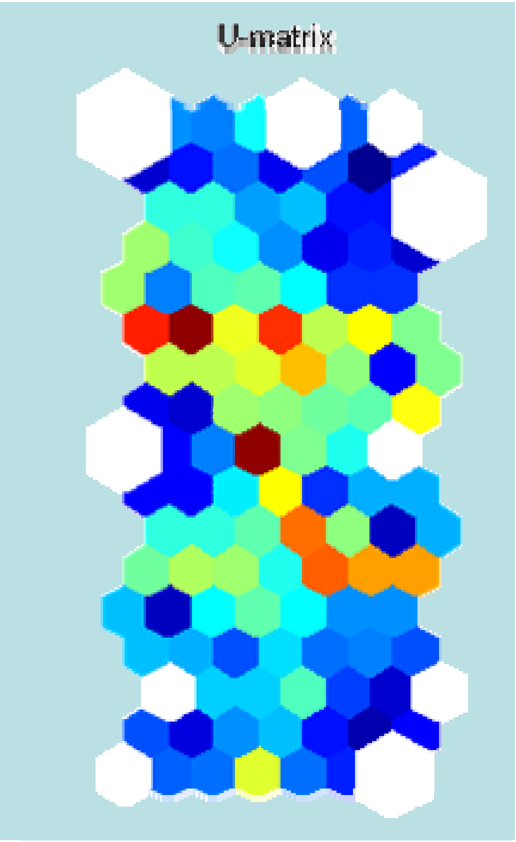
**Coloured vectors are populations**

**Black vectors are the SOM "best matching-vectors".**



**But there could be "best-matching vectors" not associated with input data**

**Then displays them as a "map", so that topology (relationships) is maintained**



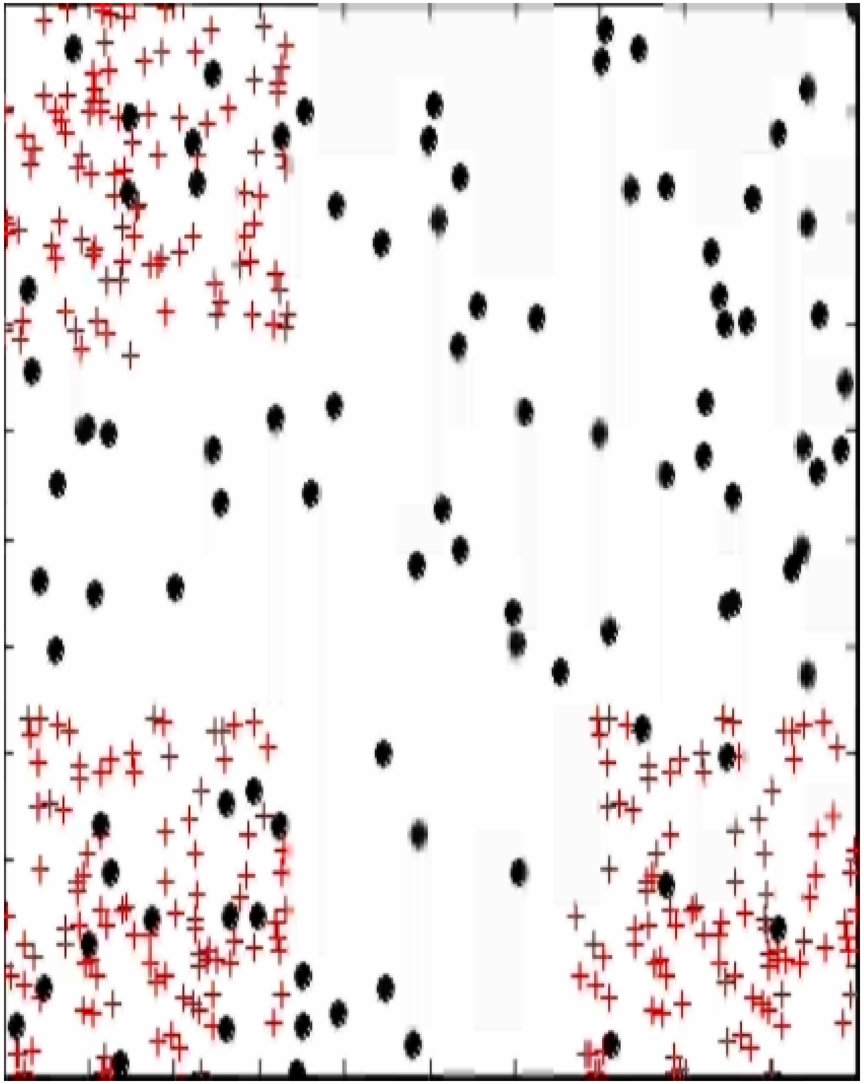
# SOM Training Visualization in 2D – Step 1 Initialize the SOM

**Red crosses represent data points in 2 dimensions**

**A SOM of 12x8 has been chosen**

**Begin by “randomizing” the SOM to cover the data space**

**Black circles are  the random SOM “seed” vectors**

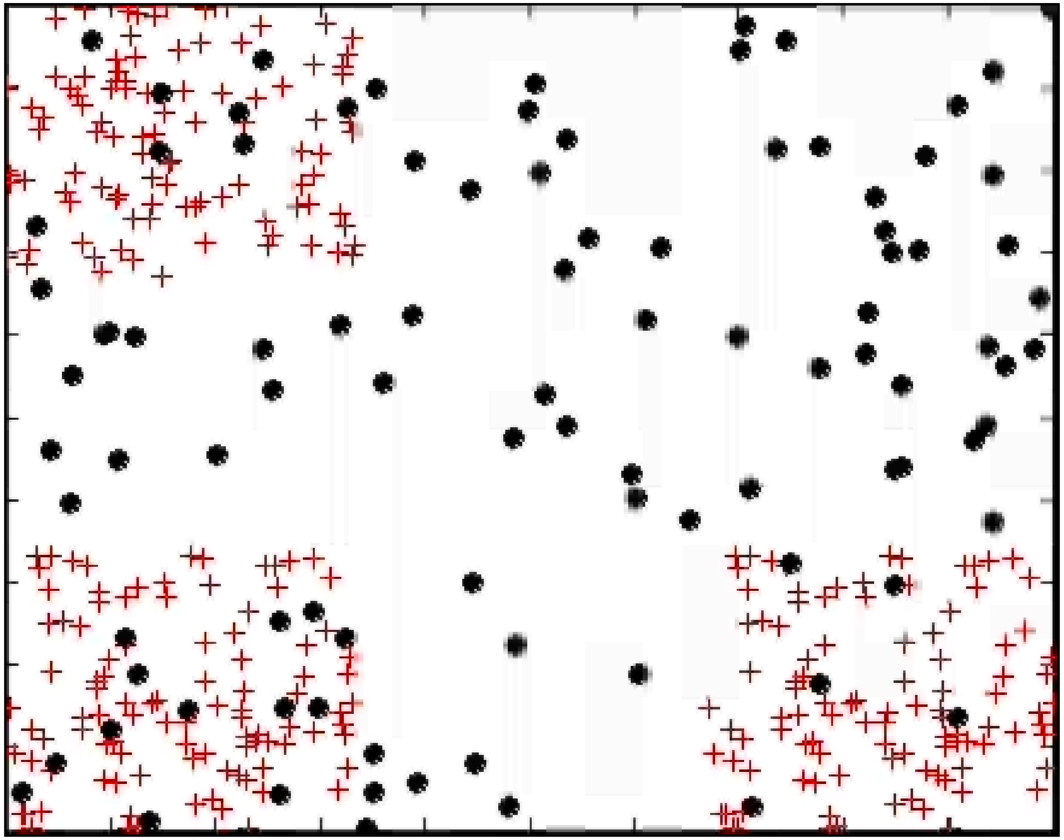


# An Example Visualization in 2D – Step 2 Train the SOM

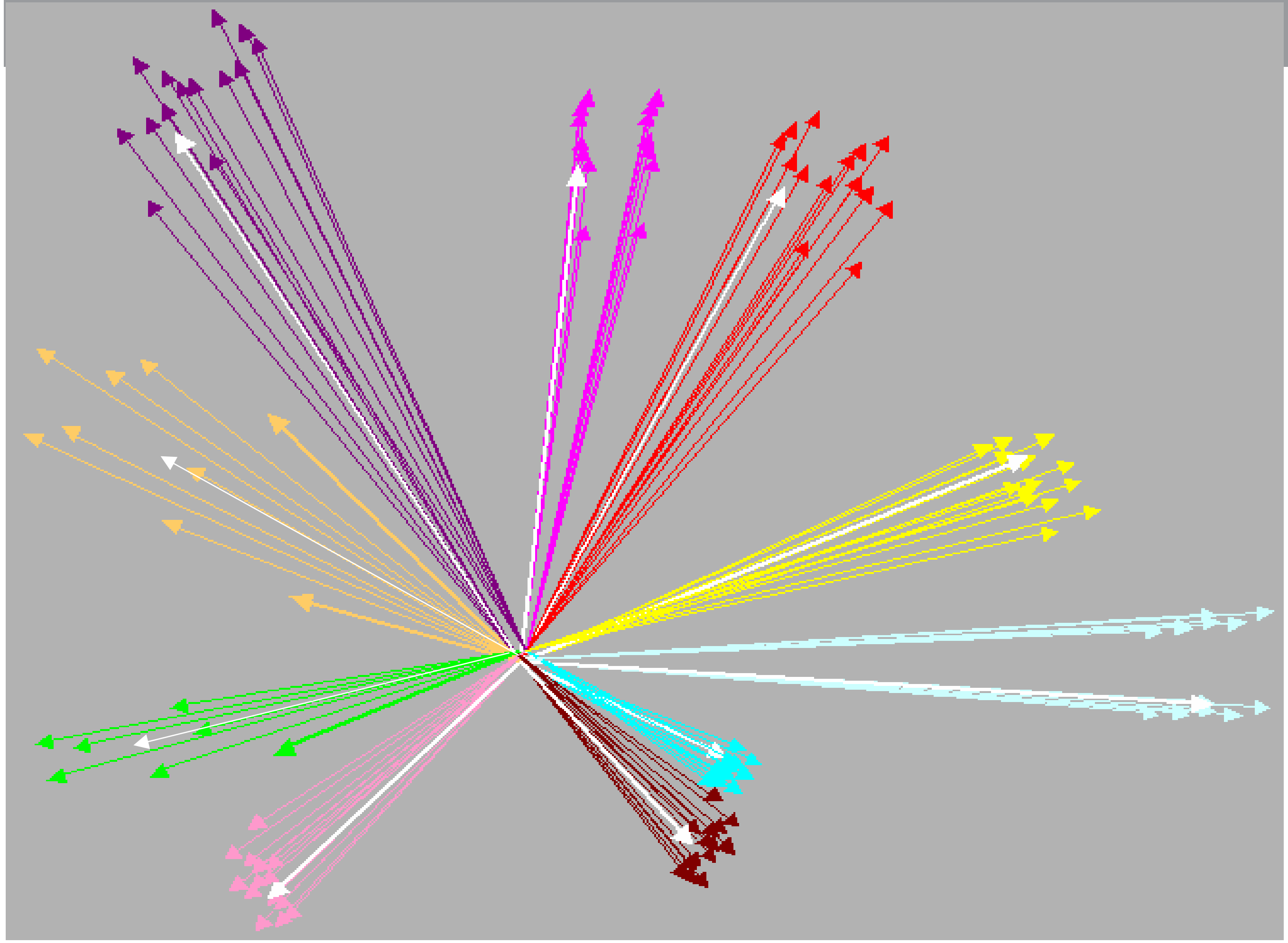
**The training is based on two principles:**

**Competitive learning: the "seed" prototype vector most similar to a data vector is modified so that it it is even more similar to it. This way the map learns the position of the data cloud.**

**Cooperative learning: not only is the most similar "seed" prototype vector modified, but also its neighbours on the map are moved towards the data vector. This way the map self-organizes.**

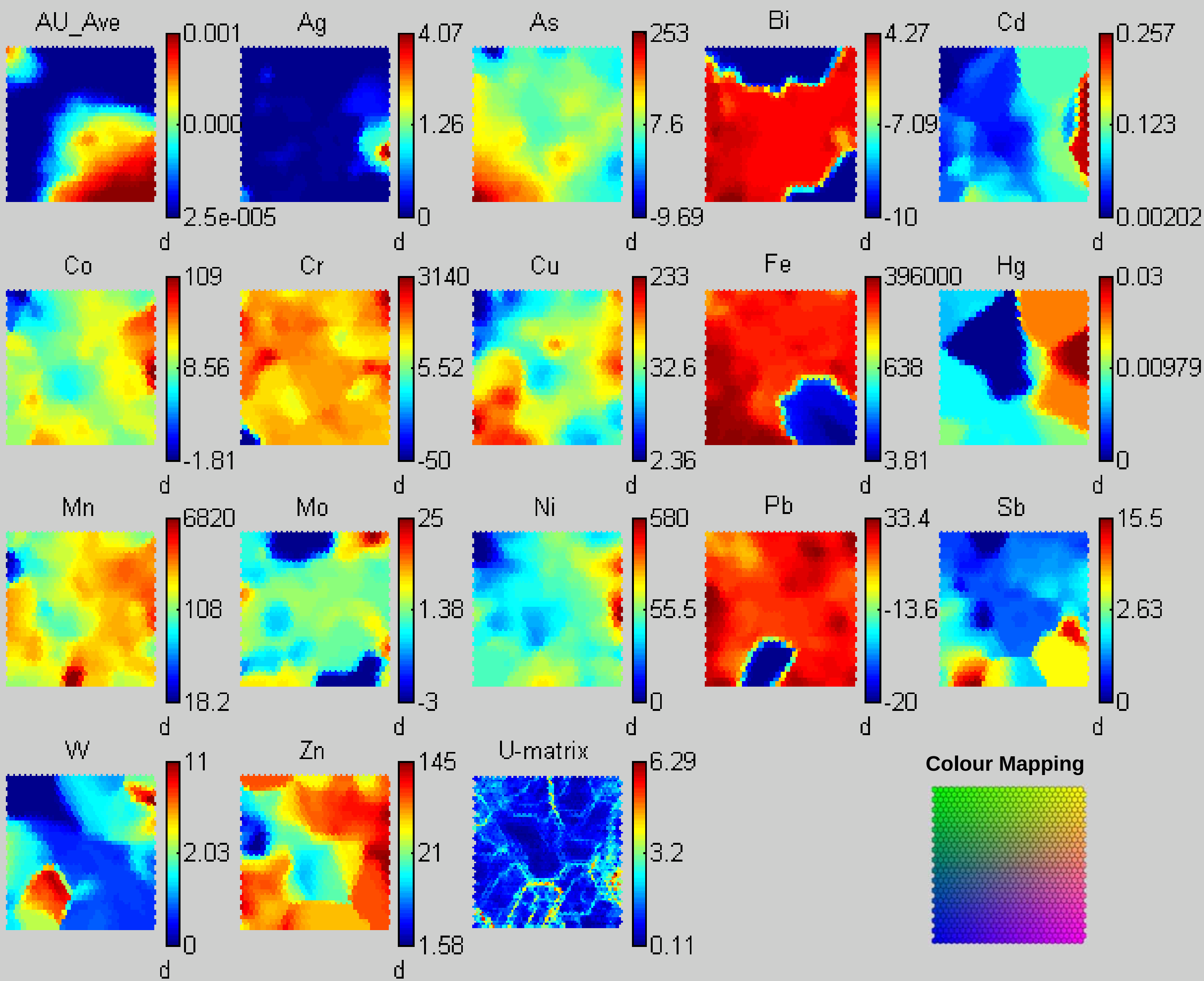


# COMPONENT PLOTS

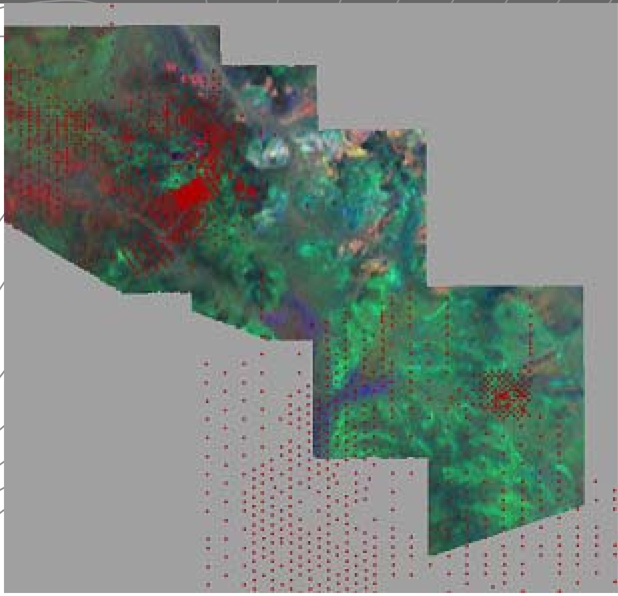


**Each "Best Matching Vector" can be described by its variables: {c1,c2,c3,…cn ,d1, d2, d3…dn etc }**

# Example of Component Plots and U-Matrix and Colour Map



# Geochemistry #1

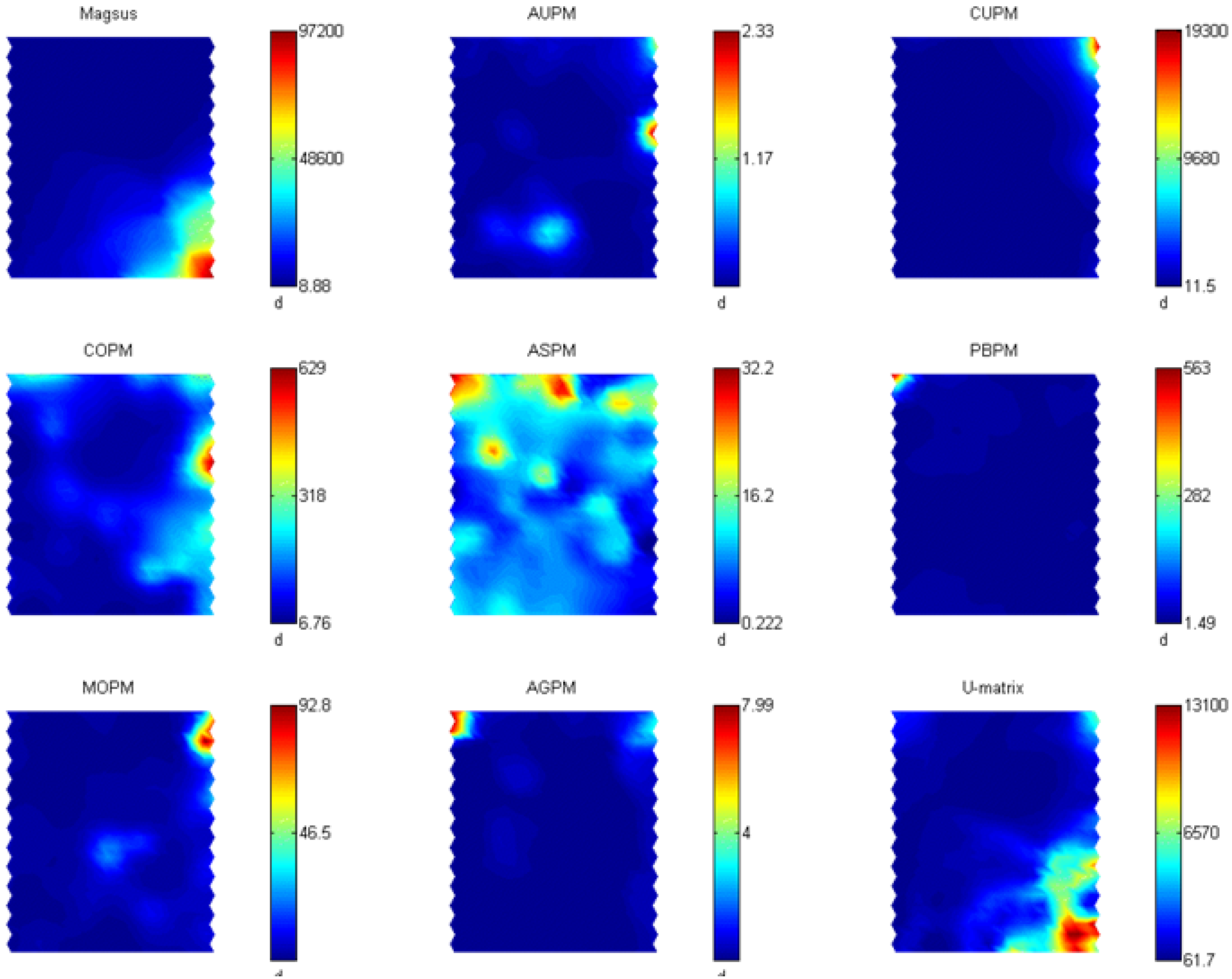


~ 15,000 RAB, RC, Air Core drill holes:

~ 40,000 located (XYZ) geochemical samples with up to 13 elements assayed:

~ 60% of data base is “empty”

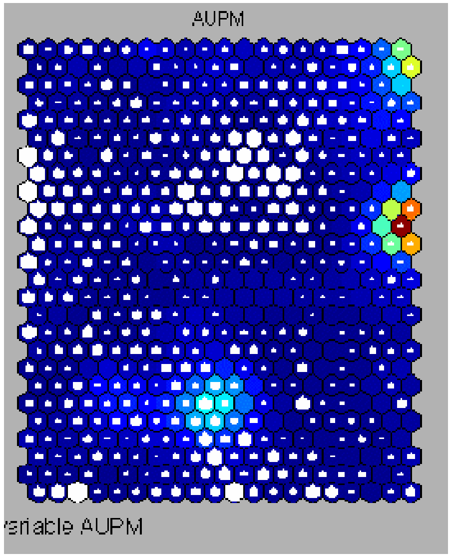
# Geochemistry #1: Component Plots

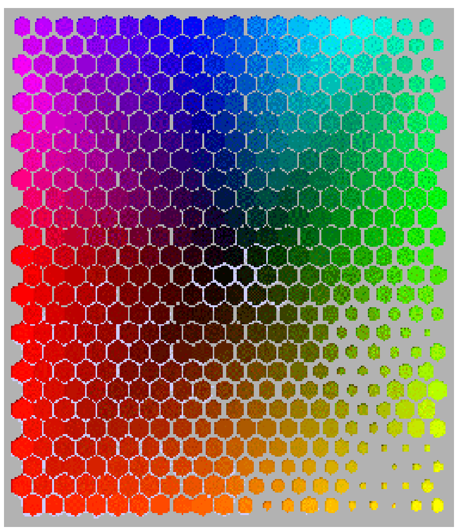


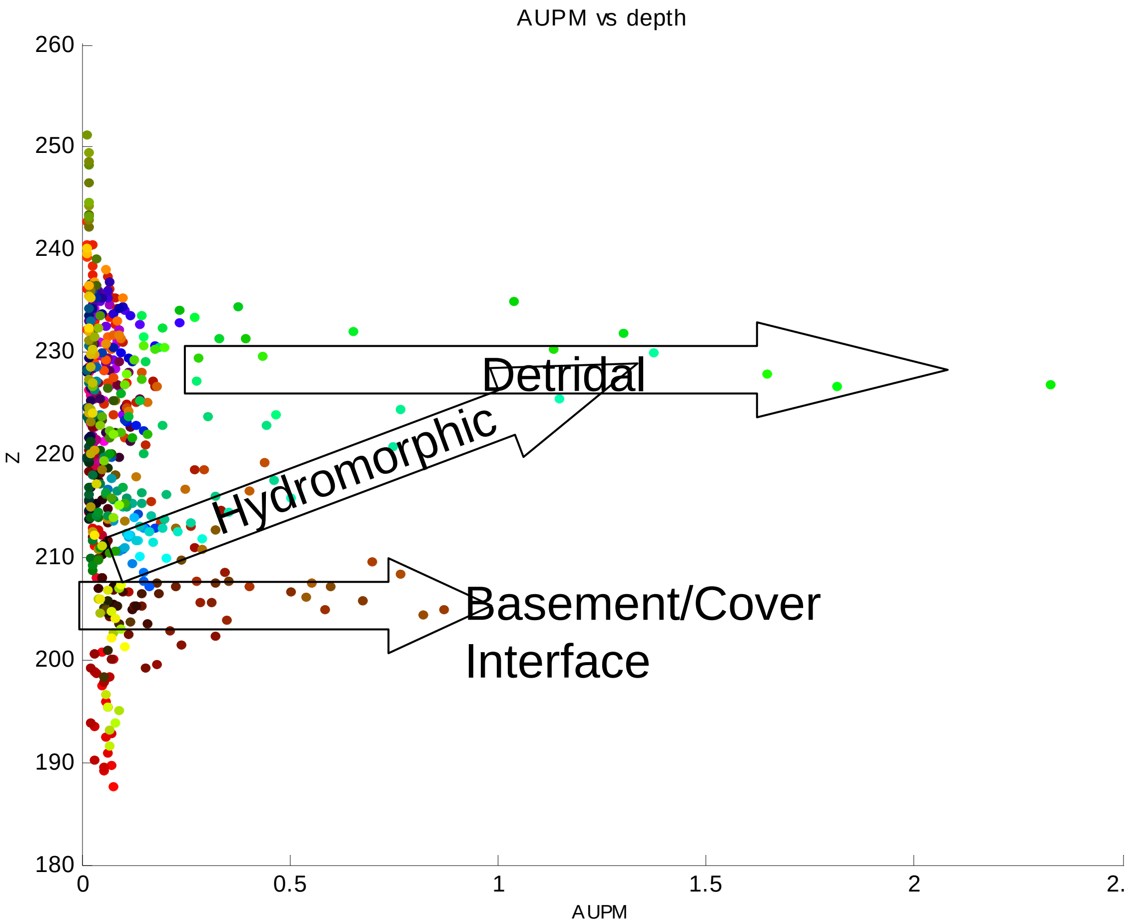

## Component Plots

# Geochemistry #1

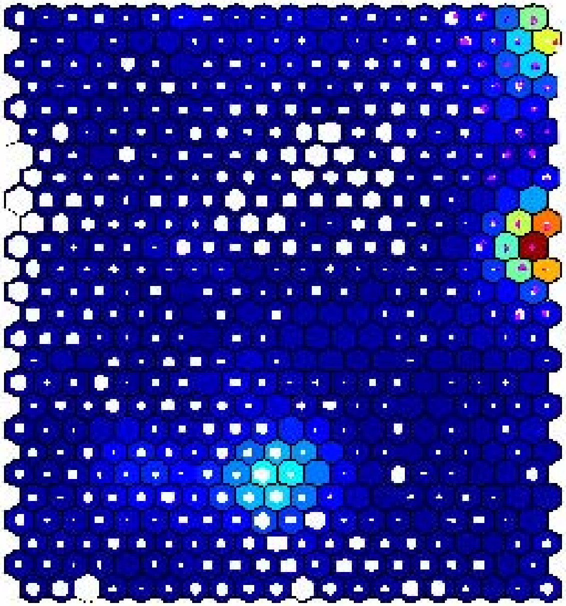
# Depth Plot



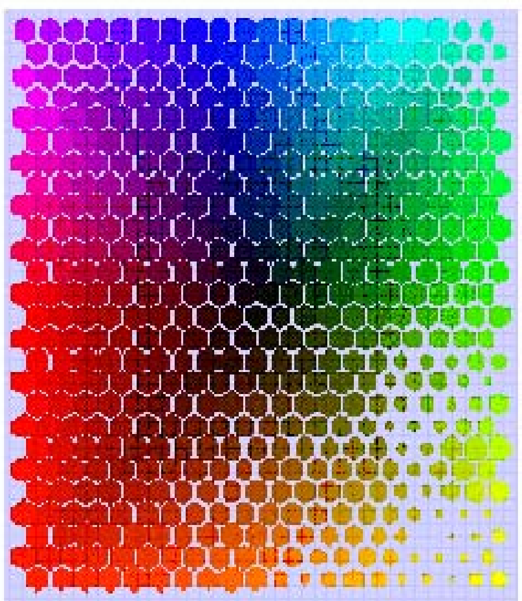



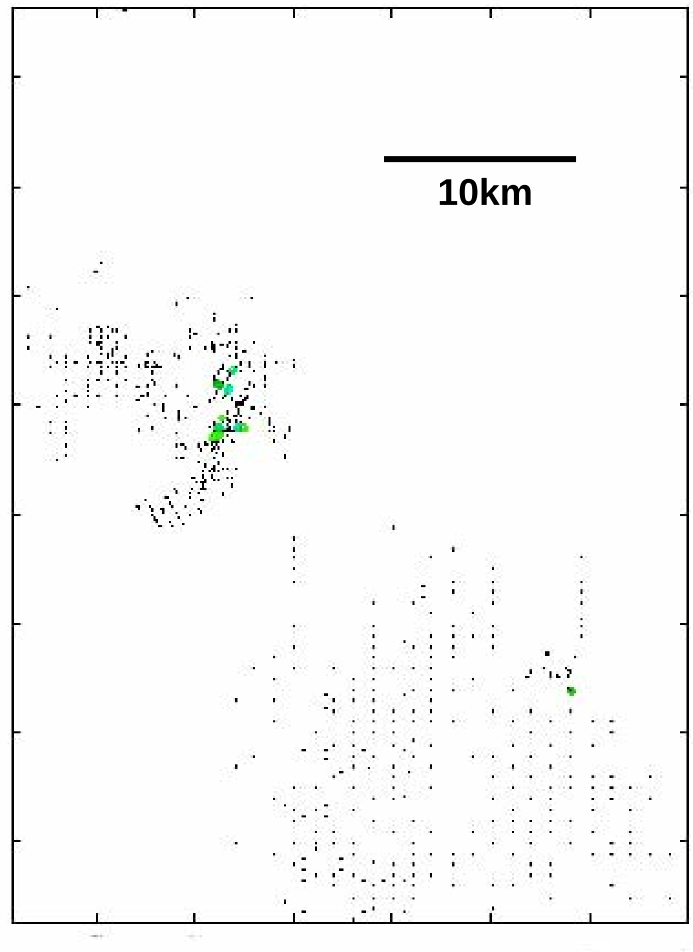



**Au Vs Depth**



variable AUPM






**Plan View**

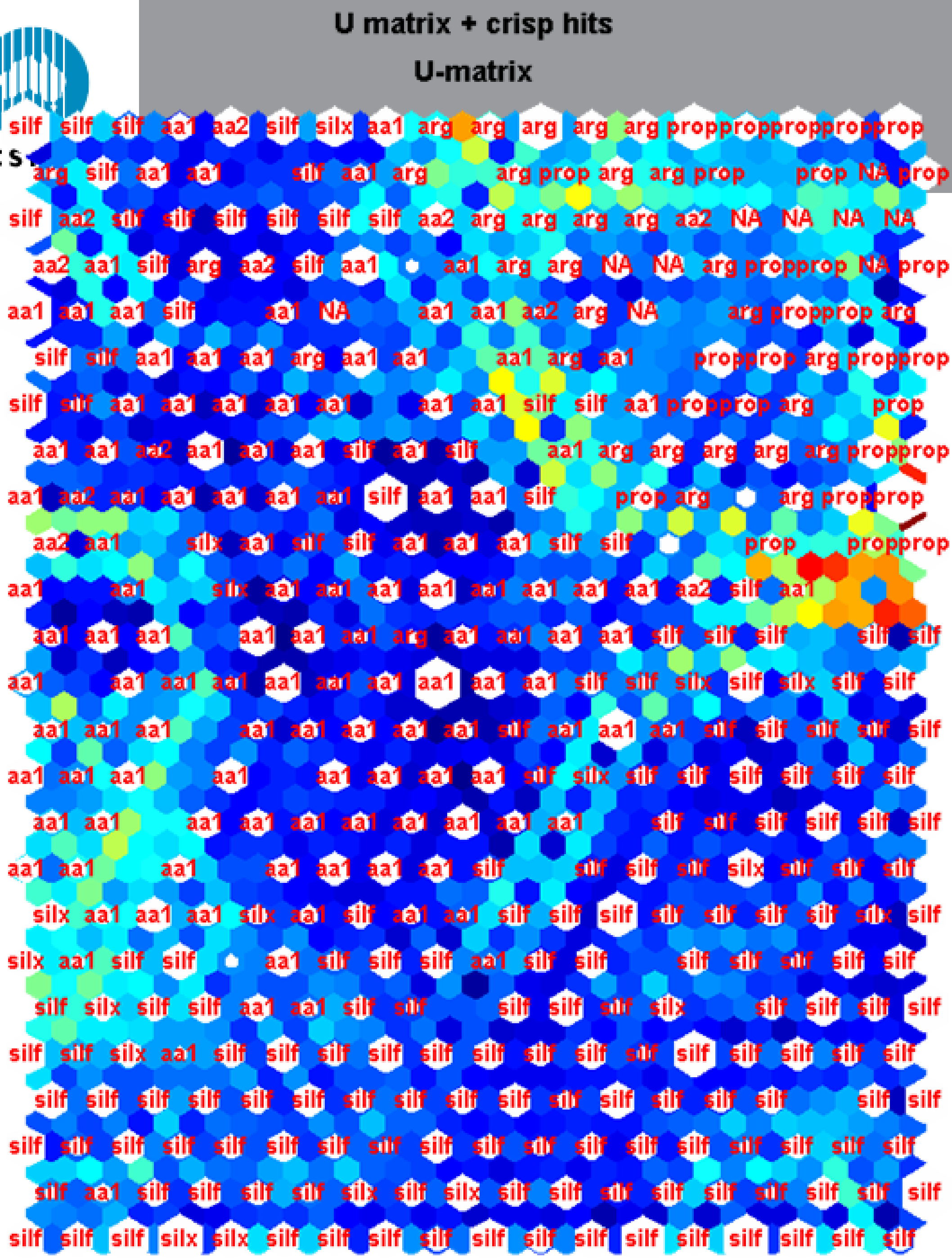
# Geochemistry #2

## High sulphidation Au deposit

**Geochemistry with “alteration” labels (not included in the processing, but carried through as labels)**

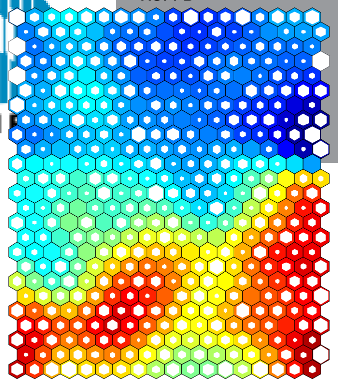
**~2500 RC Chip & Core samples (2m composites)**

**~ 20 Elements & Alteration Label (propylitic – silica flooding)**

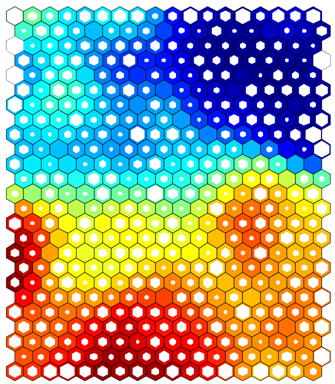



Can relate Alteration Mineralogy to Geochemistry

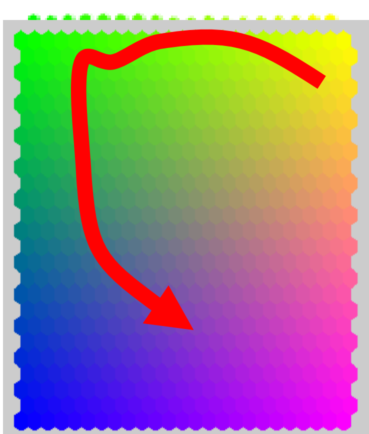
# Area “C”
# Au vs Ag Scatter Plot of BMU - Nodes

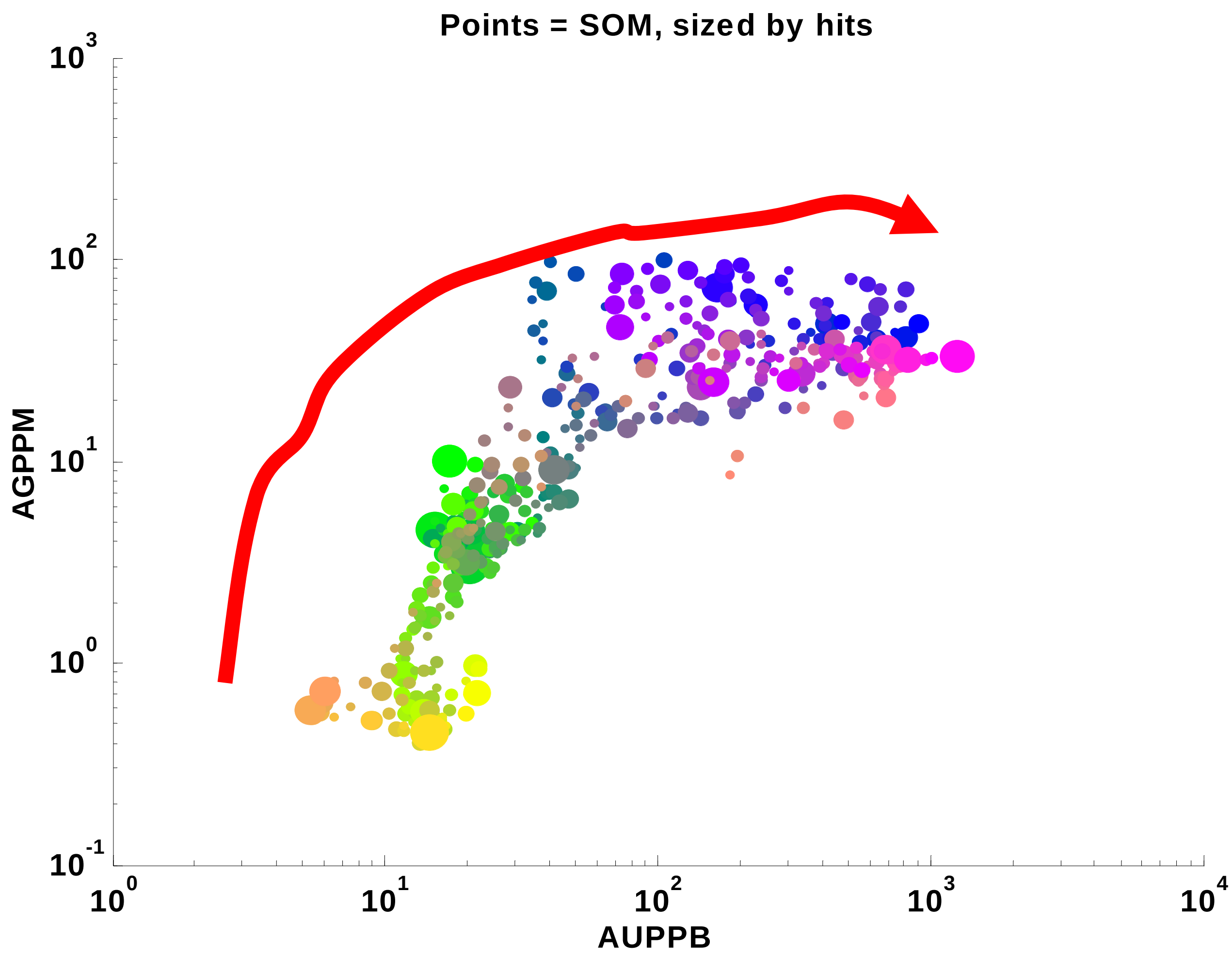



variable AUPPB



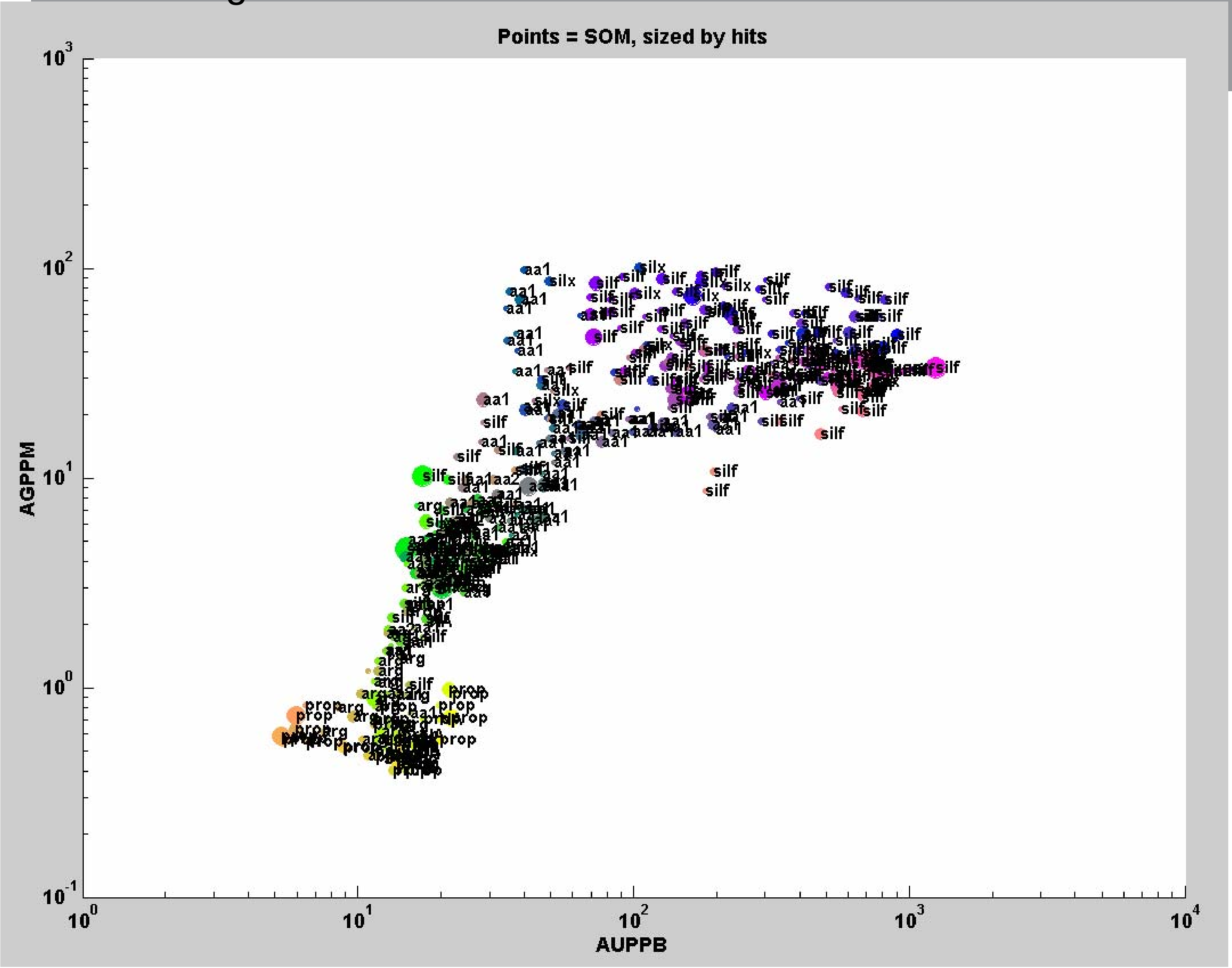

variable AGPPM



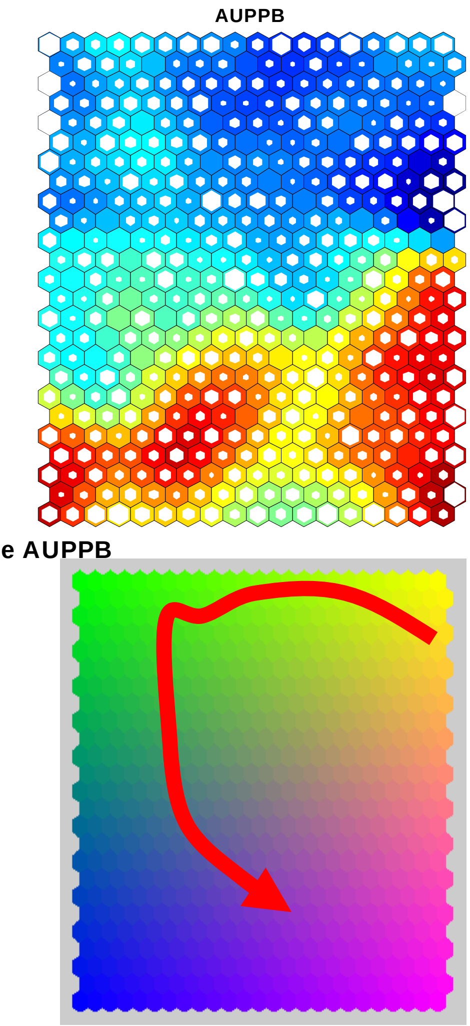


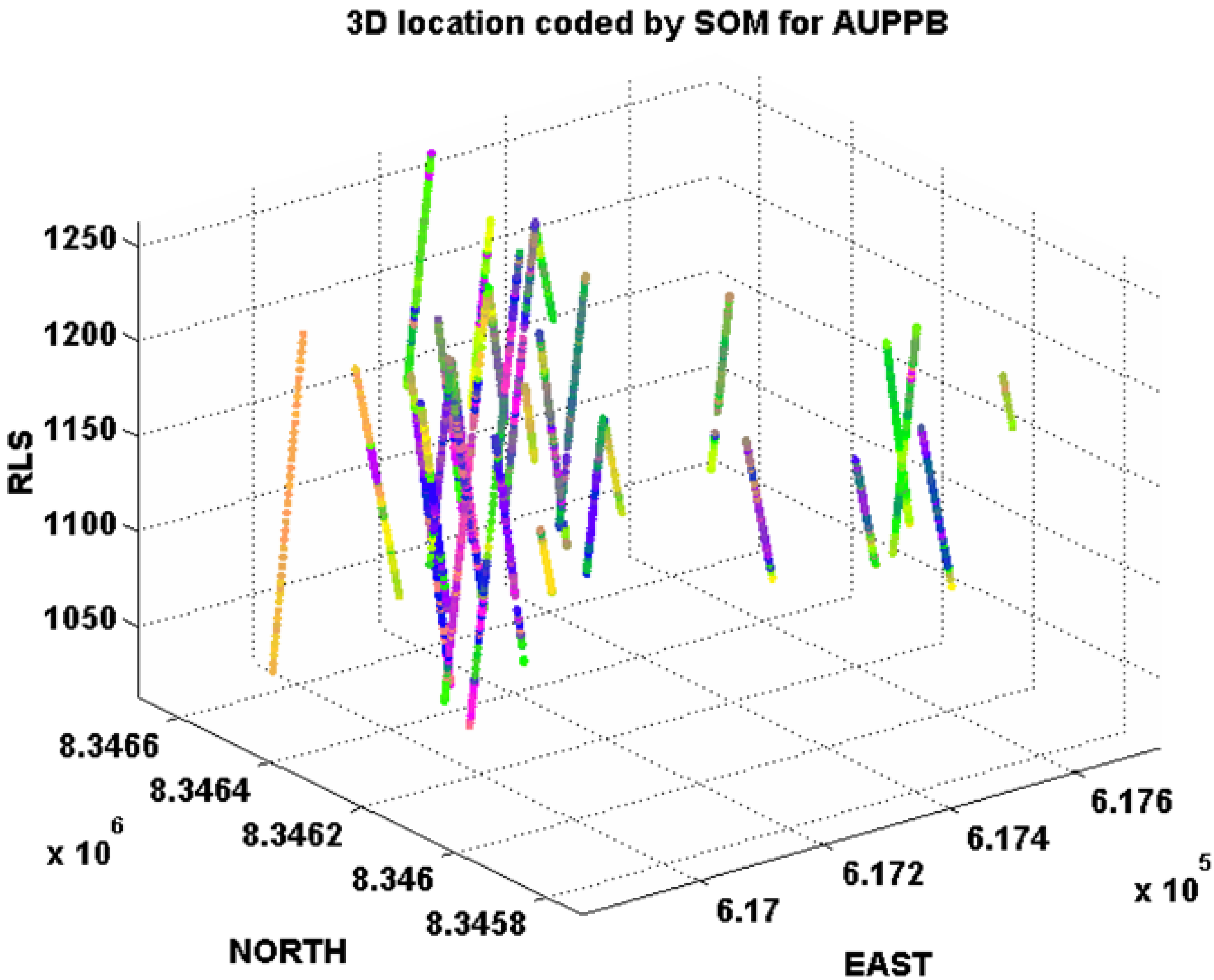

**SOM can assist in showing the “process” of mineralization (i.e., vectors to ore!!)**

# Au vs Ag Scatter Plot of BMU – Nodes with "Alt" labels



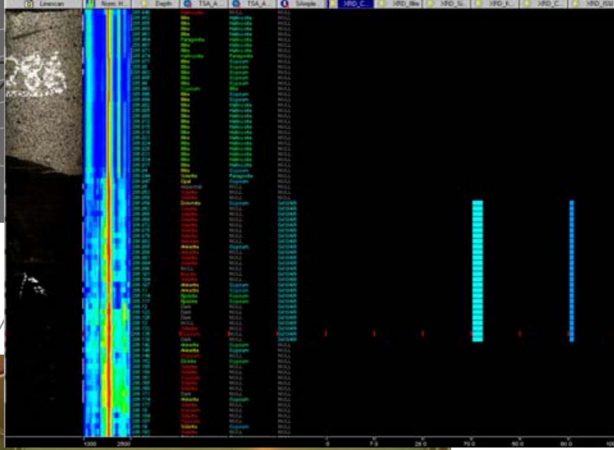
# All Au Samples – Colour-coded by SOM Colour - LUT

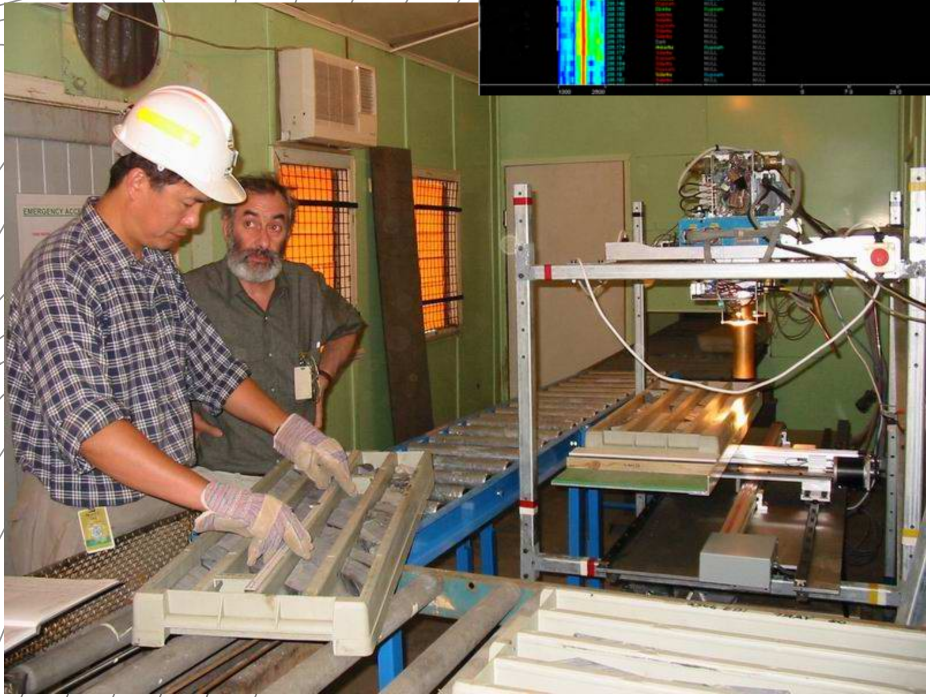


# HyLogger Data

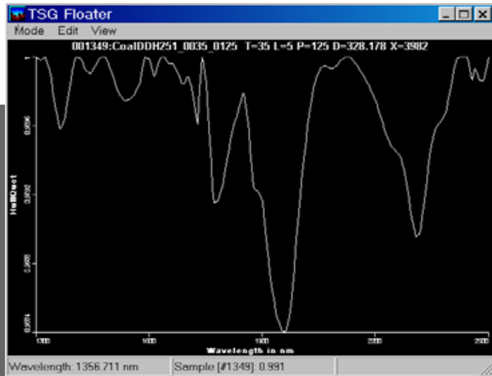
## Identification of Geologically-significant “Intervals” of mineral spectra.



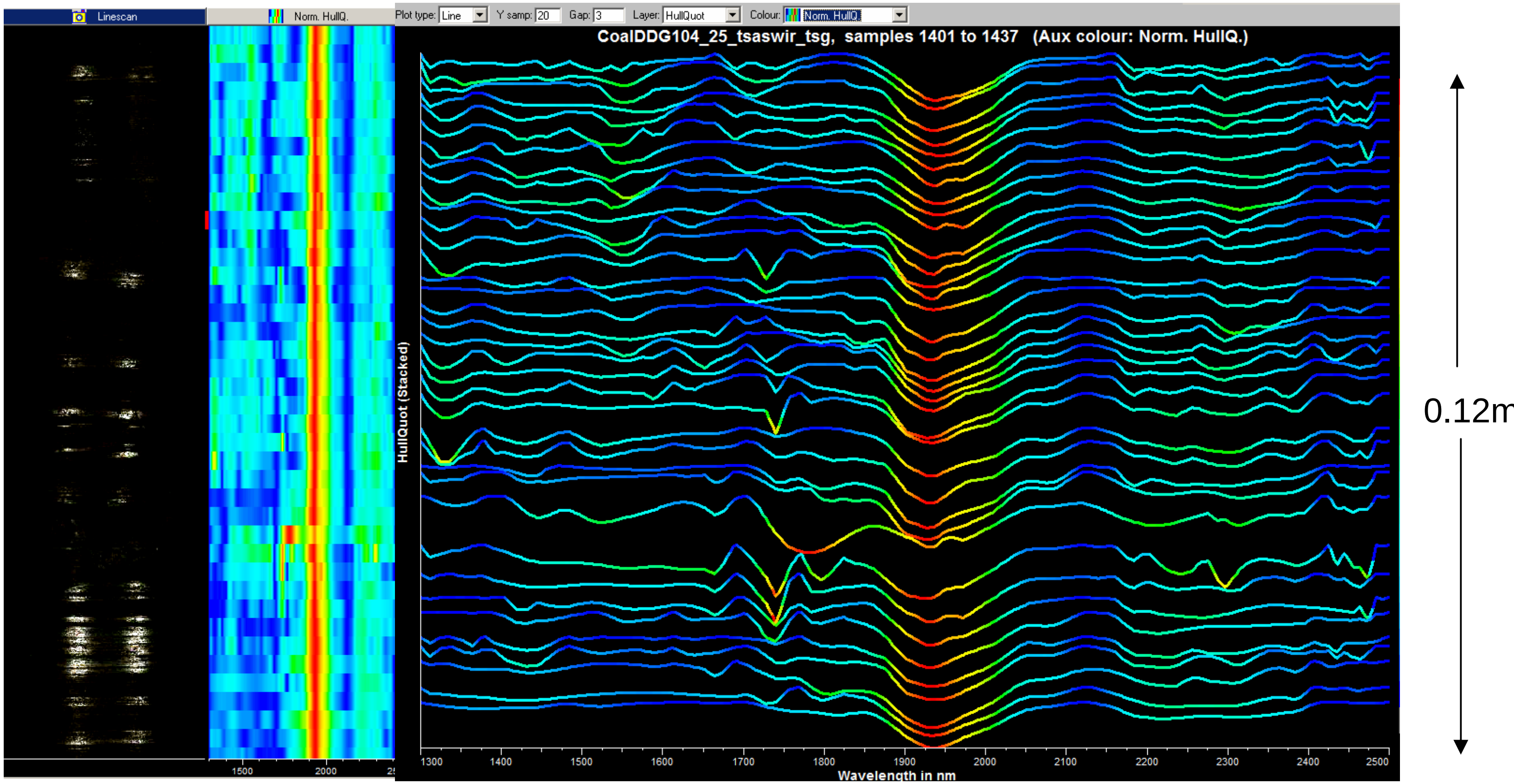




ACARP Project **C13014:**



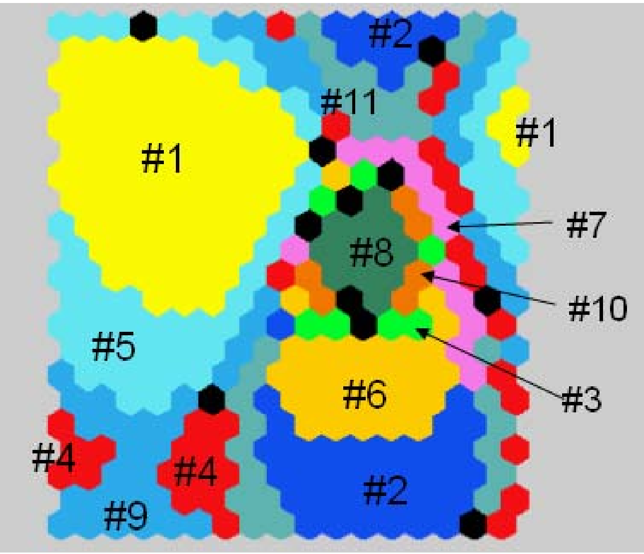
# HyLogger – TSG Spectral Output




~ 40,000 spectra each with ~ 520 variables (1300-2500nm) channels input to SOM

A measurement every 0.003 m !

# SOM-derived “Class-Spectra”




#1 – Siderite: (H20 0.4)

#2 – Illite 64%: Disordered Kaolinite 36% (H2O :0.5)

#3 - Problematic

#4 - Illite 100% (H2O: 0.55)

#5 -Siderite 67% Illite 33% (H2O: 0.38)

#6 – Illite 51% Siderite 49% (H2O: 0.40)

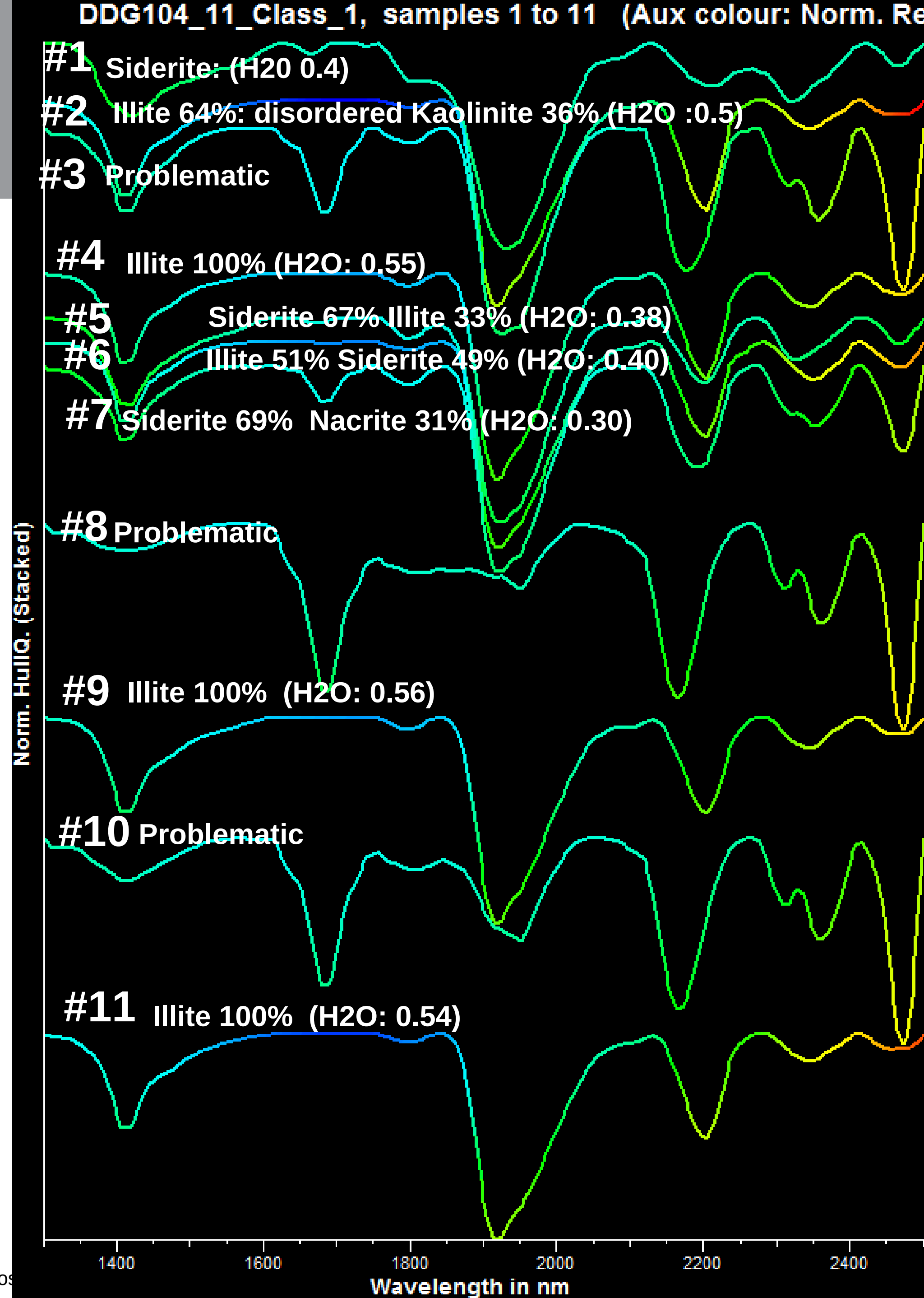
#7 – Siderite 69% Nacrite 31% (H2O: 0.30)

#8 - Problematic

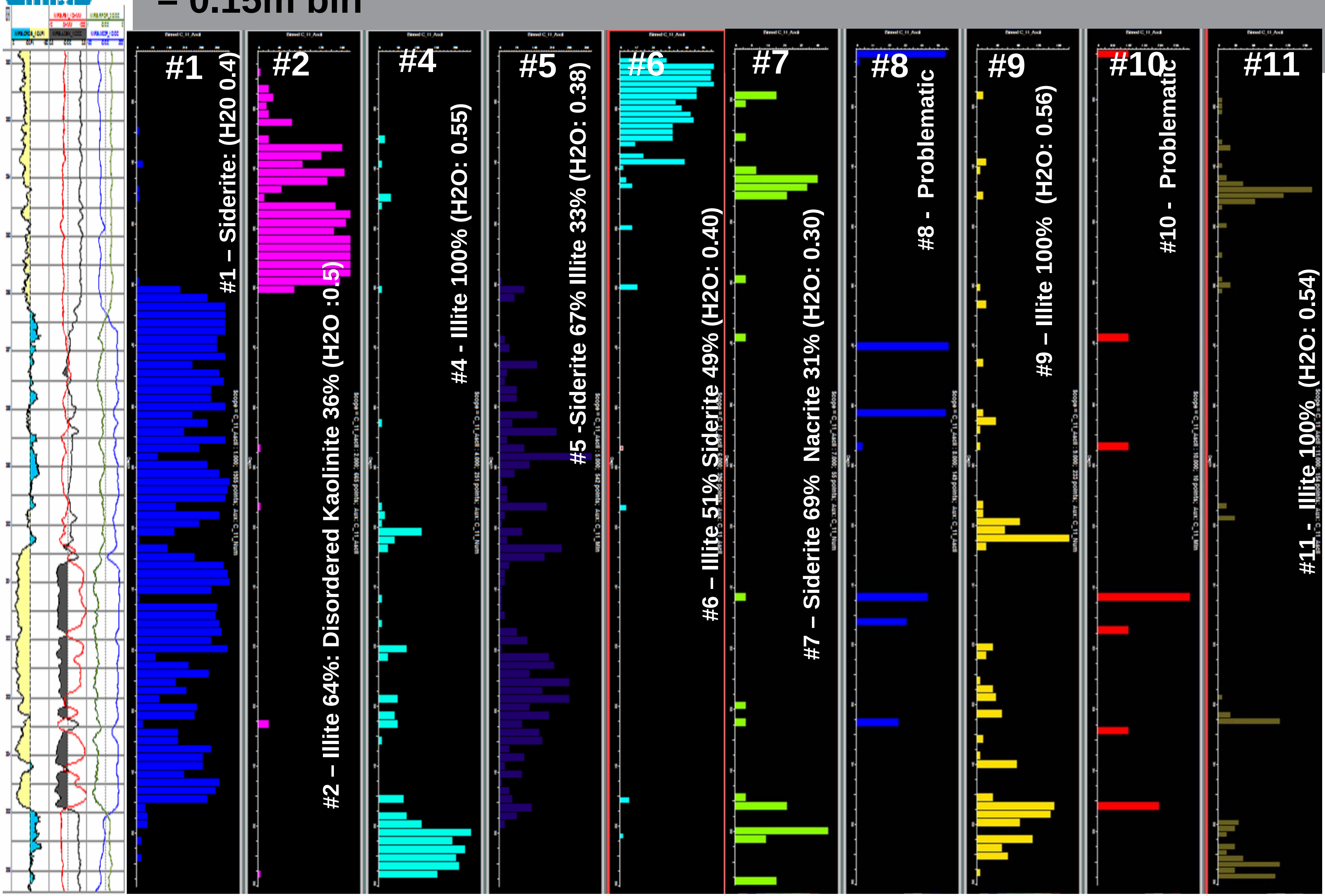
#9 – Illite 100% (H2O: 0.56)

#10 - Problematic

#11 - Illite 100% (H2O: 0.54)



# “Binned” CSOM-Derived “Spectral Classes” vs Depth – 0.15m bin



# SOMe Specific Conclusions

**Geochemistry #1:**

- Able to identify geological process in the Mesozoic and basement related to Au distribution;
- Able to identify “anomalous” older samples despite incomplete assay suite.

**Geochemistry #2:**

Confirmed “alteration” relates to mineralization

Identified the process of mineralization; “vectors-to-ore”.

**HyLogger Data:**

Simplified a complex data set into ‘packages’ related to geology and sedimentary facies.

# SOMe General Conclusions

SOM is an unsupervised, data-driven, exploratory data analysis tool;

Non-traditional Non-Statistical approach to data analysis;

Ideal for “sparse” geological data;

Opens the door to “Integrated Analysis and Interpretation of Disparate Data Types”;

The spatial coherence and juxtaposition of SOM “clusters” is important;

Provides “simplification” of complex data sets;

Scatterplots of SOM nodes highlight geological “process”;

A SOM “framework” once computed can be used to process or predict responses for new data.

Thank you for your time and interest

# THE END

# QUESTIONS ?