Cluster Analysis of XRPD Data for Ore Evaluation

Accuracy in Powder Diffraction IV

NIST, Gaithersburg MD, USA

April 22-25, 2013

Uwe König

Thomas Degen

PANalytical B.V.

Almelo, The Netherlands

Outline

1. XRPD in mining Industries & iron making process
2. Iron ore mining and traditional grade control
3. Cluster analysis introduction / basics
4. Cluster analysis example – from measurements to via analysis results to automation
5. Conclusions / Benefits

XRPD in mining industries

- XRPD analysis methods:
  - Mineral identification
  - Mineral quantification
  - Sample classification / clustering

- XRPD application areas:

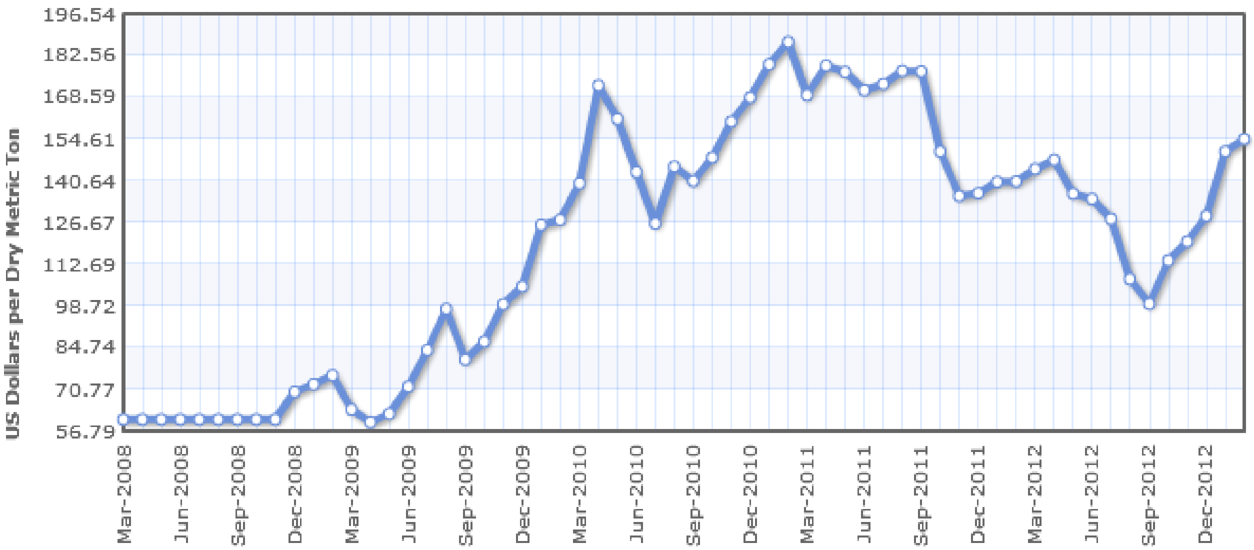
Exploration

Processing

Quality control / ore blending
(import regulations…)

Markets

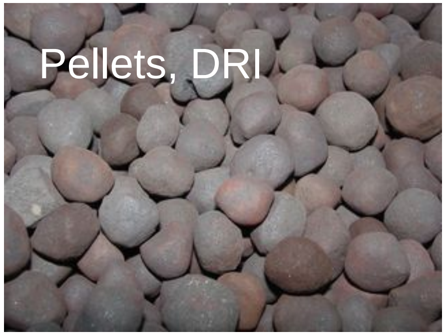
- Iron ore prices remain at high levels
- Need for better grade control due to
  - Decreasing quality of ore
  - Increase lifetime of mills
- Various XRPD applications in the iron making process:
  - Iron ore
  - Iron ore sinter
  - DRI (sponge iron)
  - Steel
  - Slag



Iron making - value chain

Raw materials

Intermediate products



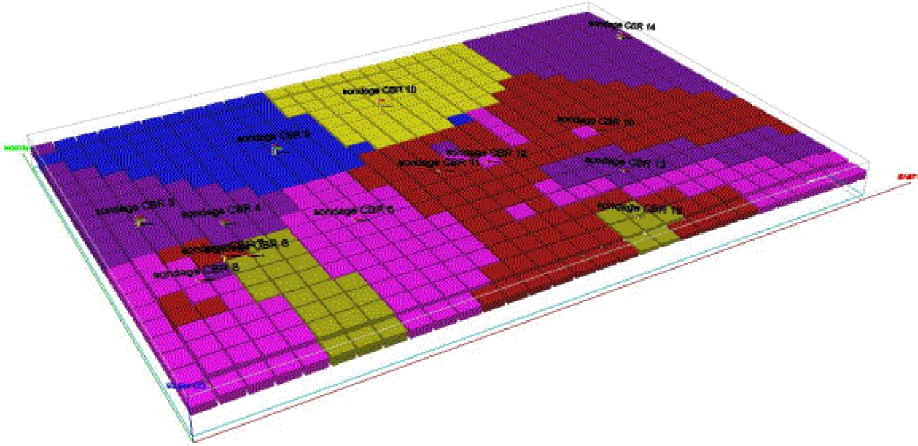
Final product

Waste

Mining

Steel plant

Mining of iron ore: Grade control / Classification



High Grade
Fe > 65%, low Al

BENEFICIATION / SINTER PLANT

Low Grade
Fe < 55%

WASTE

Intermediate
Fe > 55%, **other phases**

$ $ $

BENEFICIATION

Classical prospection of deposits

- Geological prospecting; drilling campaign, 20 x 20 meters grid of grade blocks
- Elemental analysis:
  - Wet chemistry
  - X-ray fluorescence
- Phase composition neglected
- Human factor (geologist, lab chemist)

WRONG DETERMINATION OF ONE GRADE BLOCK (20.000 T) = LOSS OF 100.000 €

Analysis of the Fe and Al content

- Grade is defined mainly by:
  - Fe content (mainly: hematite Fe_2O_3, magnetite Fe_3O_4, goethite FeOOH)
  - Al content (increases viscosity of the slag in blast furnace)
  - Plus other trace elements: P, S, Si
- Goal: **Fe enrichment**, **removal of Al** phases
- Analysis of Al containing phases, especially of Al containing goethite, are useful for the grade differentiation
- Removal of Al from the ore is time-, energy- and cost-intensive depending on the mineral the Al is bound to
  - Kaolinite, $Al_2Si_2O_5(OH)_4$ → Removable
  - Gibbsite, $Al(OH)_3$ → Removable
  - Goethite, (Fe,Al)OOH → NOT Removable !
  - …

Cluster Analysis Basics

Potential (XRPD) Analysis Routes:

- Qualitative Analysis – Search/Match and Phase ID
- Quantitative Analysis – RIR / Rietveld Refinement
- Classification Analysis – Multivariate Statistics / Cluster Analysis => Goal: Detect how many groups are present

From Wikipedia:

“Cluster analysis is the task of grouping a set of objects in such a way that objects in the same group (called **cluster**) are more similar to each other than to those in other groups”

Cluster Analysis Basics

Cluster Analysis is an umbrella term for lots of different multivariate statistics methods that sort items into groups like:

- Hierarchical Clustering
- K-Means Clustering
- DBSCAN / OPTICS
- Fuzzy Clustering
- Principal Component Analysis
- MMDS
- ….

Cluster Analysis Basics

- The implementation can be seen as a multistep process:

1. Data Reduction – The Correlation Matrix
2. Agglomerative Hierarchical Cluster Analysis – The Dendrogram
3. Estimation of the Number of Clusters – KGS Test / Largest step
4. Principal Component Analysis – 3D Score Plot / Eigenvalue Plot

Lots of additional features are available like: Well plate display, thumbnail views, silhouettes, fuzzy clustering, coinciding peaks views and many other graphical options.

PANalytical

Cluster Analysis – Data Reduction

1. *Data Reduction:*

In fact it is possible to directly use the XRPD raw data for the Cluster Analysis. In a typical case with 65 scans and 3600 data points per scan this sums up to:

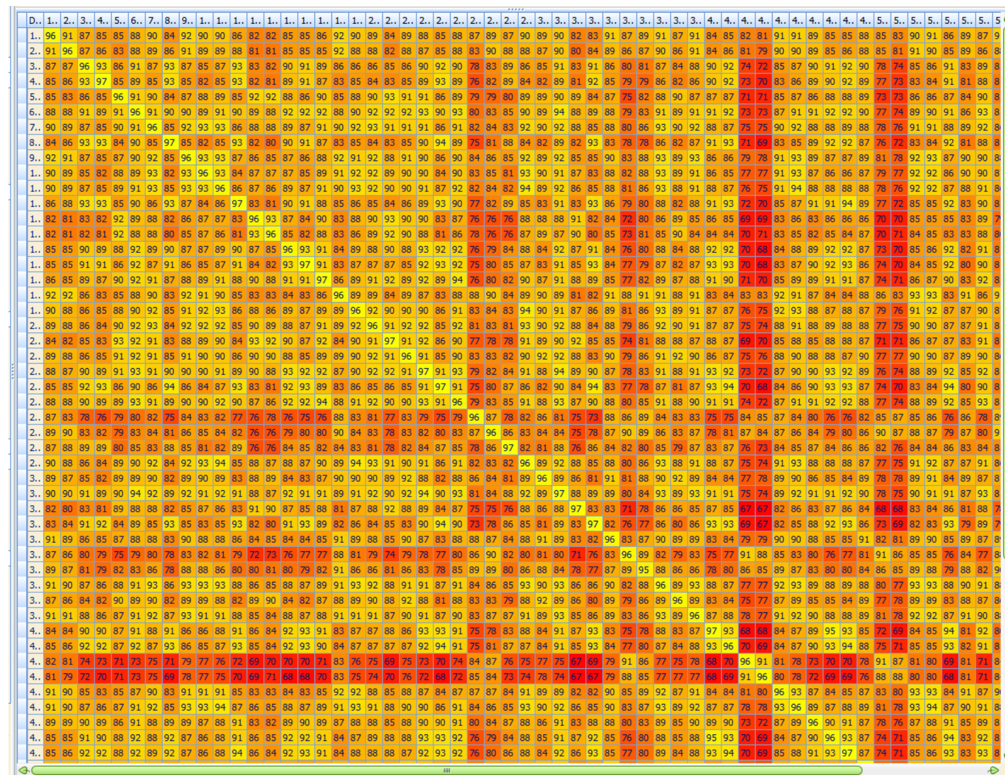
234.000 data points !

To prevent using this huge amount of data we generate the correlation matrix by comparing each scan with each other one. This produces one score for each scan pair. Thus the whole data gets reduced to:

65 * 65 = 4225 scores

Cluster Analysis – Data Reduction

The score is a number between 0 and 100 and it quantifies the similarity of a scan pair. 100 means scans are identical, 0 means they are totally different.



The result is a color coded **n** x **n** matrix where **n** is the number scans.

The score calculation is carried out by the same algorithm that is used for Search/Match.

The algorithm first reduces all scans & peaks to probability curves.
The scores are calculated by a direct comparison of these curves.

Cluster Analysis – Hierarchical Cluster Analysis

2. *Agglomerative Hierarchical Cluster Analysis – The Dendrogram*

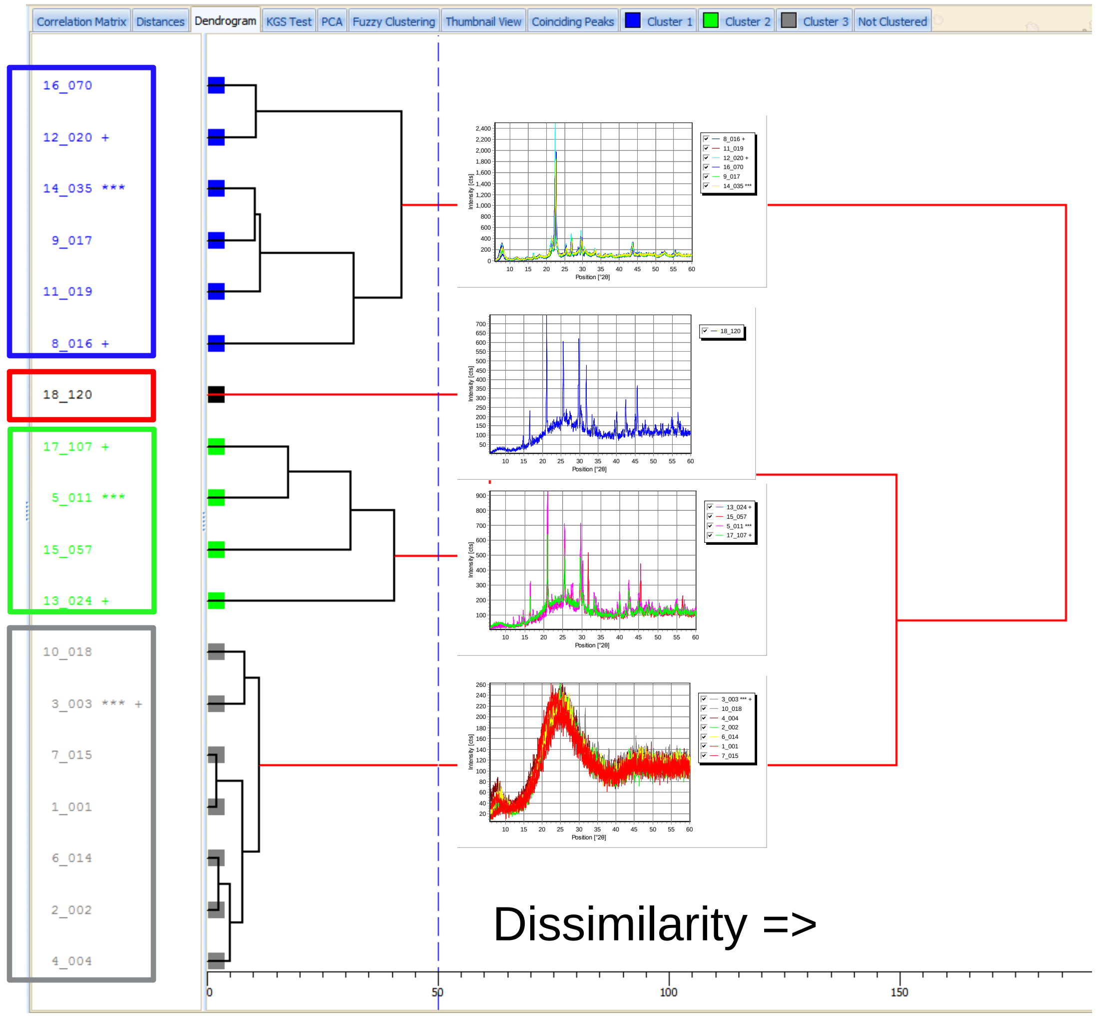
- The correlation matrix generated in step 1 is used as input to a **hierarchical agglomerative cluster analysis**, which puts the patterns into classes defined by their similarity. This method starts with each dataset representing a distinct cluster. At each step of the analysis two clusters with the highest degree of similarity are merged into a single cluster. The process stops with the final step, when only one cluster containing all datasets remains.

- The result of this analysis is usually displayed as a **Dendrogram**. Each pattern starts at the left side as a separate cluster, these **Clusters** amalgamate in a stepwise fashion, linked by vertical tie bars. The horizontal position of the tie bar represents a dissimilarity measure.

Cluster Analysis – Hierarchical Cluster Analysis

The **Dendrogram** is a graphical display of the result of an agglomerative hierarchical cluster analysis, actual cut-off analysis, indicated by a blue stippled line.

The final task is to determine the “right” **cut-off**!



Cluster Analysis – Number of Clusters

3. *Estimation of the number of clusters*

A well known and in principle unsolved problem is to **find the "right" number of clusters**. This means cutting the **Dendrogram** at a given dissimilarity and retaining a meaningful set of clusters, where the scans inside a cluster are closely related while the different clusters are different enough to keep them apart.

HighScore (Plus) offers two automatic ways to solve this:

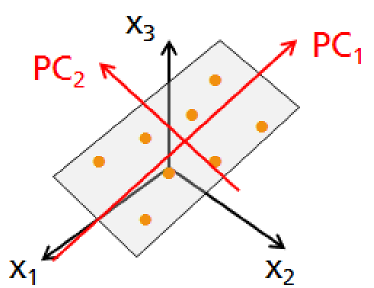
1. Put the cut-off at the position of the largest relative step on the dissimilarity scale between the clusters. This mimics what people are doing by looking at a Dendrogram.

2. The KGS (**K**elley, L.A., **G**ardner, S.P., **S**utcliffe, M.J.) test. This test has been developed to find the representative clusters of a set of protein structures derived from NMR spectra.

Cluster Analysis – Principal Component Analysis

4. *Principal Component Analysis (Optional)*

Principal **C**omponent **A**nalysis (PCA) can be carried out as an independent method to visualize the quality of the clustering. It is used *to* extract the systematic variance in a data matrix. The aim is to create a set of variables that is smaller than the set of original variables but **still explains all the variance** in the matrix.
In mathematical terms **PCA** transforms a number of correlated variables into a smaller number of uncorrelated orthogonal variables, the so-called principal components.



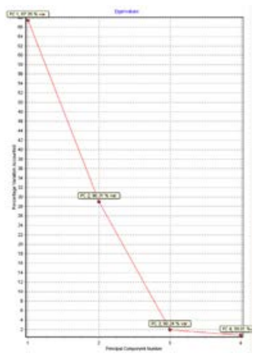
Large data matrix

Scores

- Summarize the observations
- Separates signal from noise
- Allows to observe trends

Loadings

- Summarize the variables

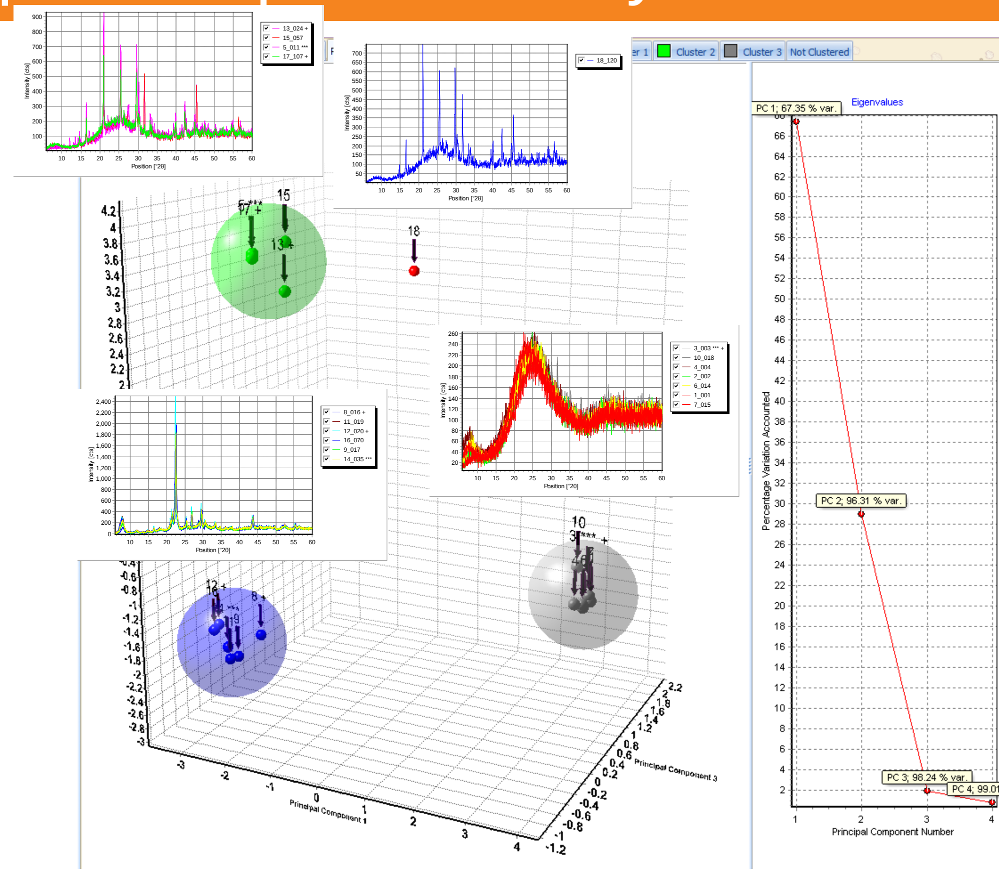


Factors, predictors (X-variables)

Extracted X – scores (PC_X)

Cluster Analysis – Principal Component Analysis

- The resulting eigenvectors are used to generate a **three-dimensional score plot** that helps to indicate which datasets belong to which cluster.
- The three axes (X, Y, Z) correspond to the first three principal components.
- The datasets are displayed as small spheres. The color of the spheres indicates the cluster they belong to.

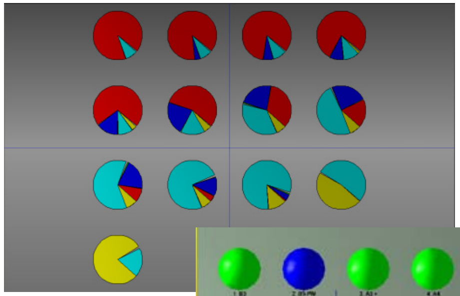


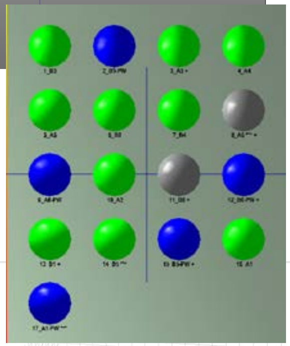
The **PCA score plot** (left) shows the clear separation of the scans into 3 clusters and one outlier; the first 3 PC’s cover 98% (67% + 29% + 2%) of the data variation (right: **Eigenvalue plot**). The plot confirms the result of the hierarchical cluster analysis.

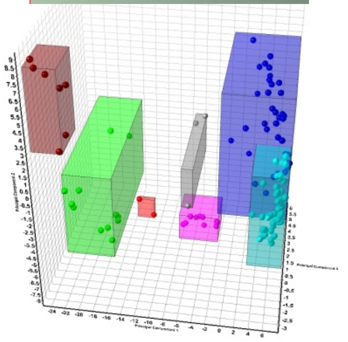
Cluster Analysis Basics

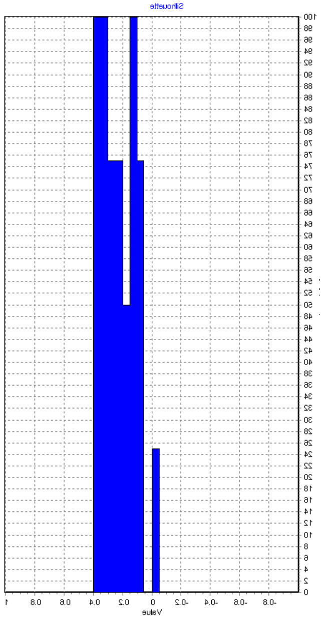
Next to these basic methods lots of useful additional features and graphical views are available:

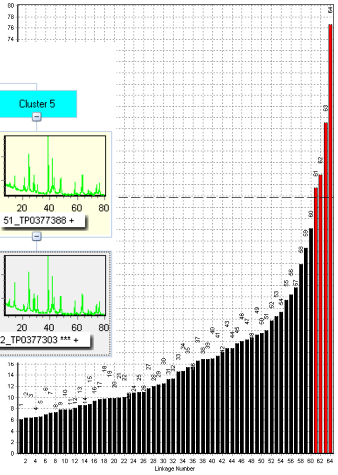
- Determination of the most representative dataset
- Determination of the two most different datasets
- Calculation of Silhouettes to analyze the quality of clusters
- Fuzzy clustering to handle mixtures
- Lots of different graphical views

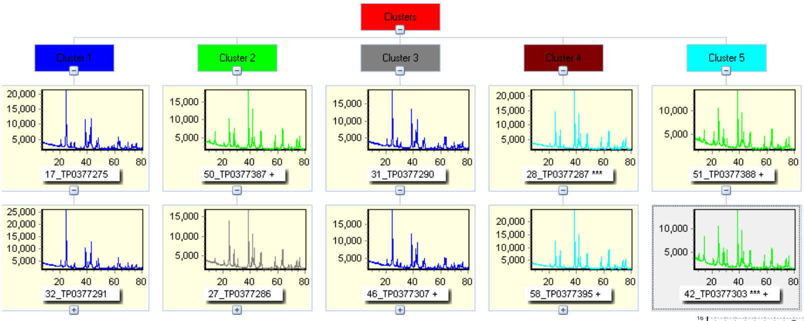






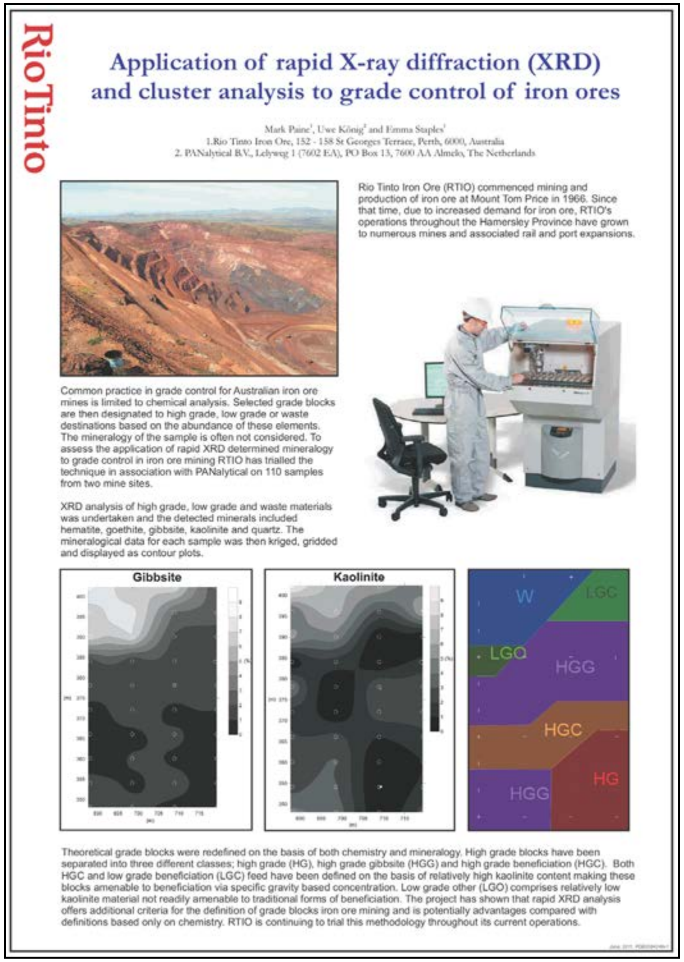






Case study - Rio Tinto Iron Ore

- Analysis of approximately 110 iron ore samples from two mine sites for grade control using XRPD



RioTinto

Application of rapid X-ray diffraction (XRD) and cluster analysis to grade control of iron ores

Mark Paine[1], Uwe König[2] and Emma Staples[1]
1.Rio Tinto Iron Ore, 152 - 158 St Georges Terrace, Perth, 6000, Australia
2. PANalytical B.V., Lelyweg 1 (7602 EA), PO Box 13, 7600 AA Almelo, The Netherlands

Rio Tinto Iron Ore (RTIO) commenced mining and production of iron ore at Mount Tom Price in 1966. Since that time, due to increased demand for iron ore, RTIO's operations throughout the Hamersley Province have grown to numerous mines and associated rail and port expansions.

Common practice in grade control for Australian iron ore mines is limited to chemical analysis. Selected grade blocks are then designated to high grade, low grade or waste destinations based on the abundance of these elements. The mineralogy of the sample is often not considered. To assess the application of rapid XRD determined mineralogy to grade control in iron ore mining RTIO has trialled the technique in association with PANalytical on 110 samples from two mine sites.

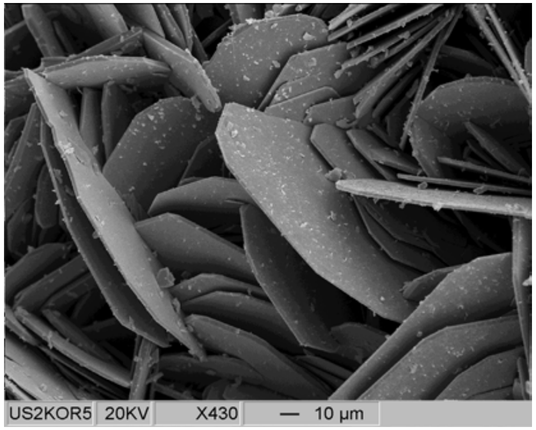
XRD analysis of high grade, low grade and waste materials was undertaken and the detected minerals included hematite, goethite, gibbsite, kaolinite and quartz. The mineralogical data for each sample was then kriged, gridded and displayed as contour plots.

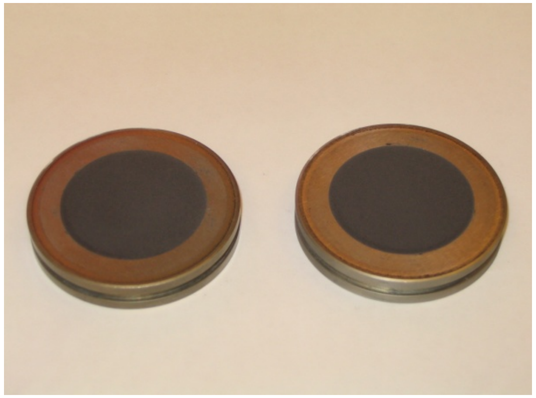
Theoretical grade blocks were redefined on the basis of both chemistry and mineralogy. High grade blocks have been separated into three different classes; high grade (HG), high grade gibbsite (HGG) and high grade beneficiation (HGC). Both HGC and low grade beneficiation (LGC) feed have been defined on the basis of relatively high kaolinite content making these blocks amenable to beneficiation via specific gravity based concentration. Low grade other (LGO) comprises relatively low kaolinite material not readily amenable to traditional forms of beneficiation. The project has shown that rapid XRD analysis offers additional criteria for the definition of grade blocks iron ore mining and is potentially advantages compared with definitions based only on chemistry. RTIO is continuing to trial this methodology throughout its current operations.

Sampling and measurement conditions

- Milling time 30 seconds
- Pressed for 30 seconds with 10 tons into steel ring sample holders

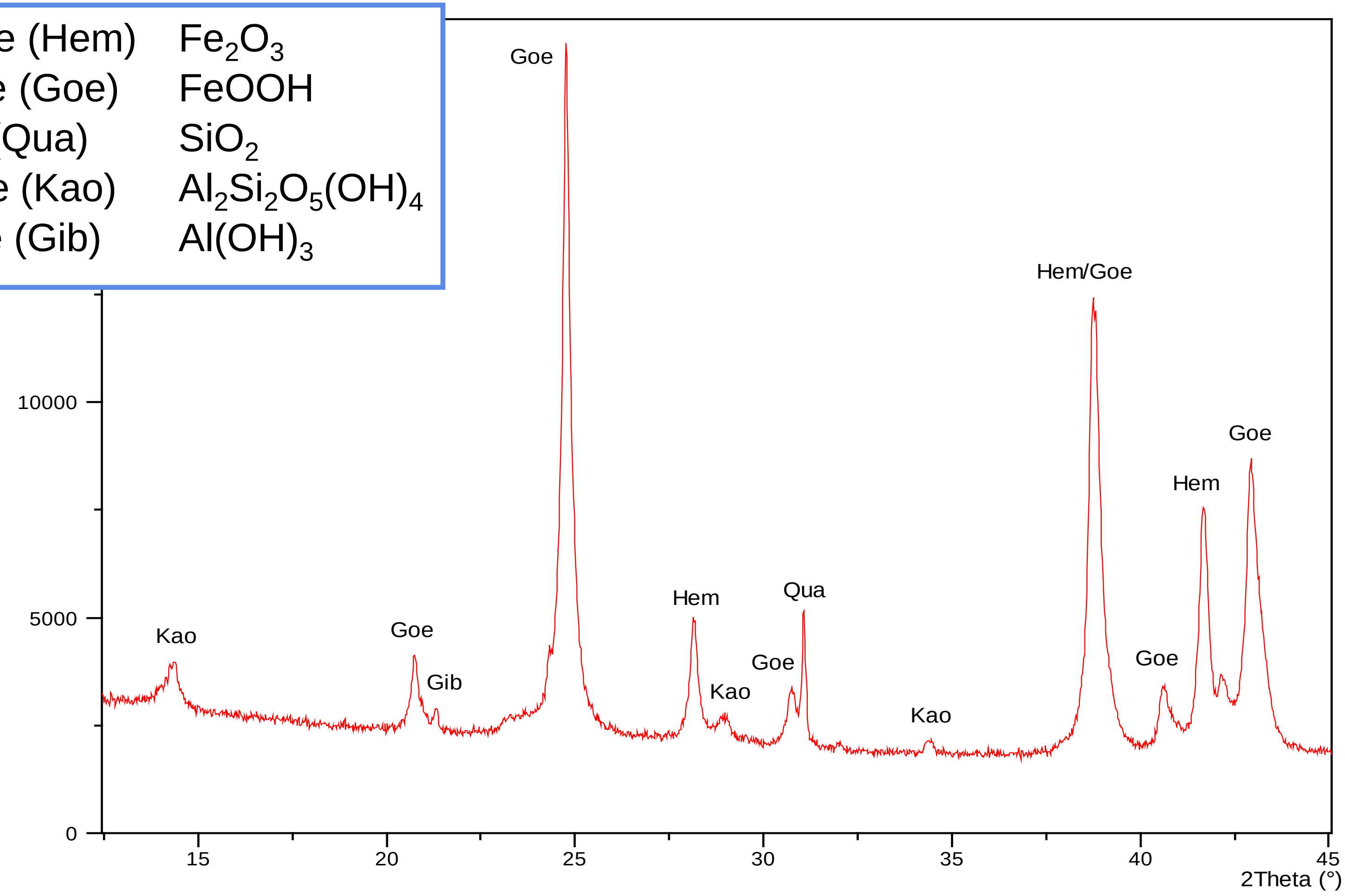
- Measurement conditions:
  - Co tube
  - Spinning sample stage
  - Range 10 – 70 deg 2Theta
  - X'Celerator detector
  - Measurement time 5 min



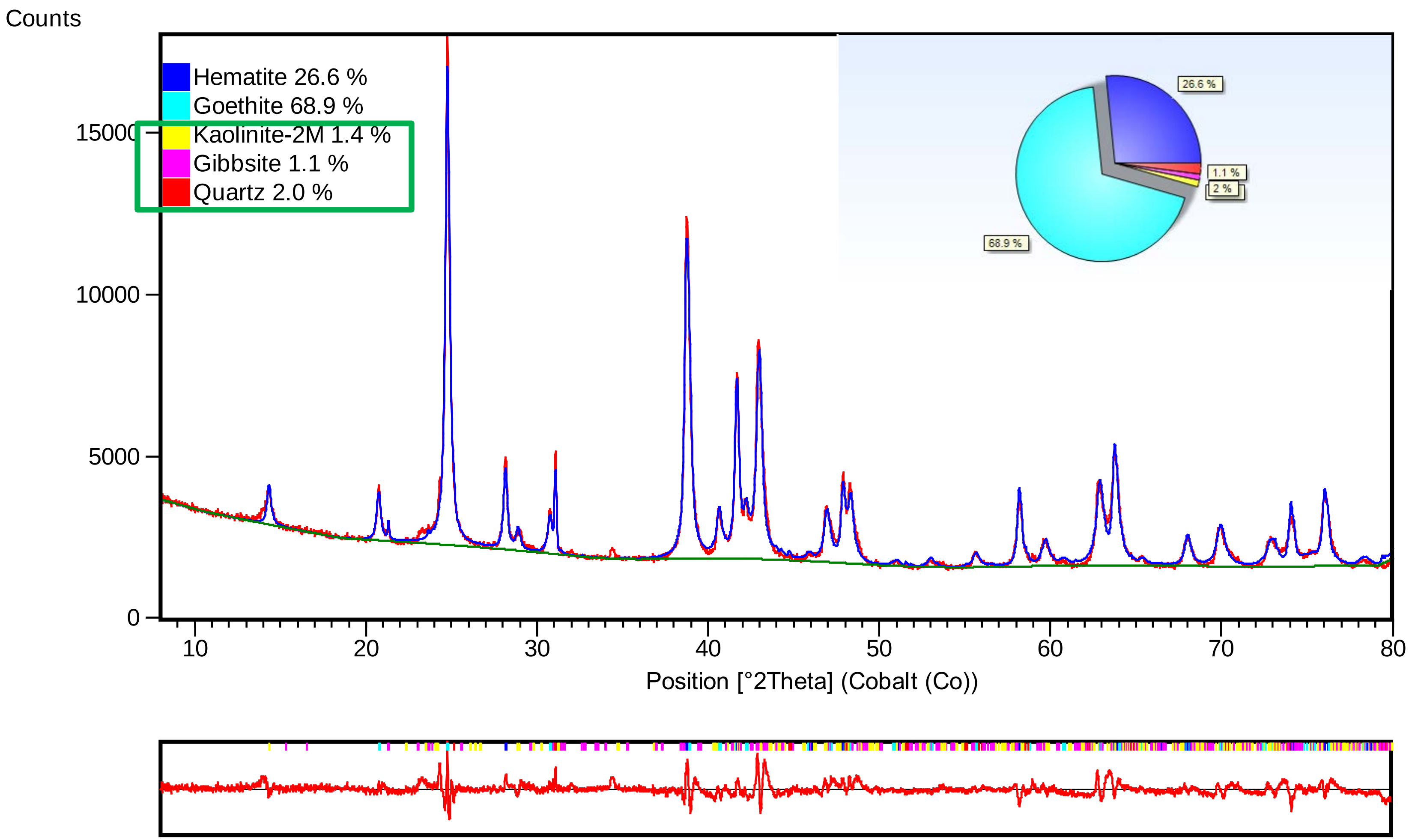


Case study - Rio Tinto Iron Ore – Phase ID (typical oxidic ore)

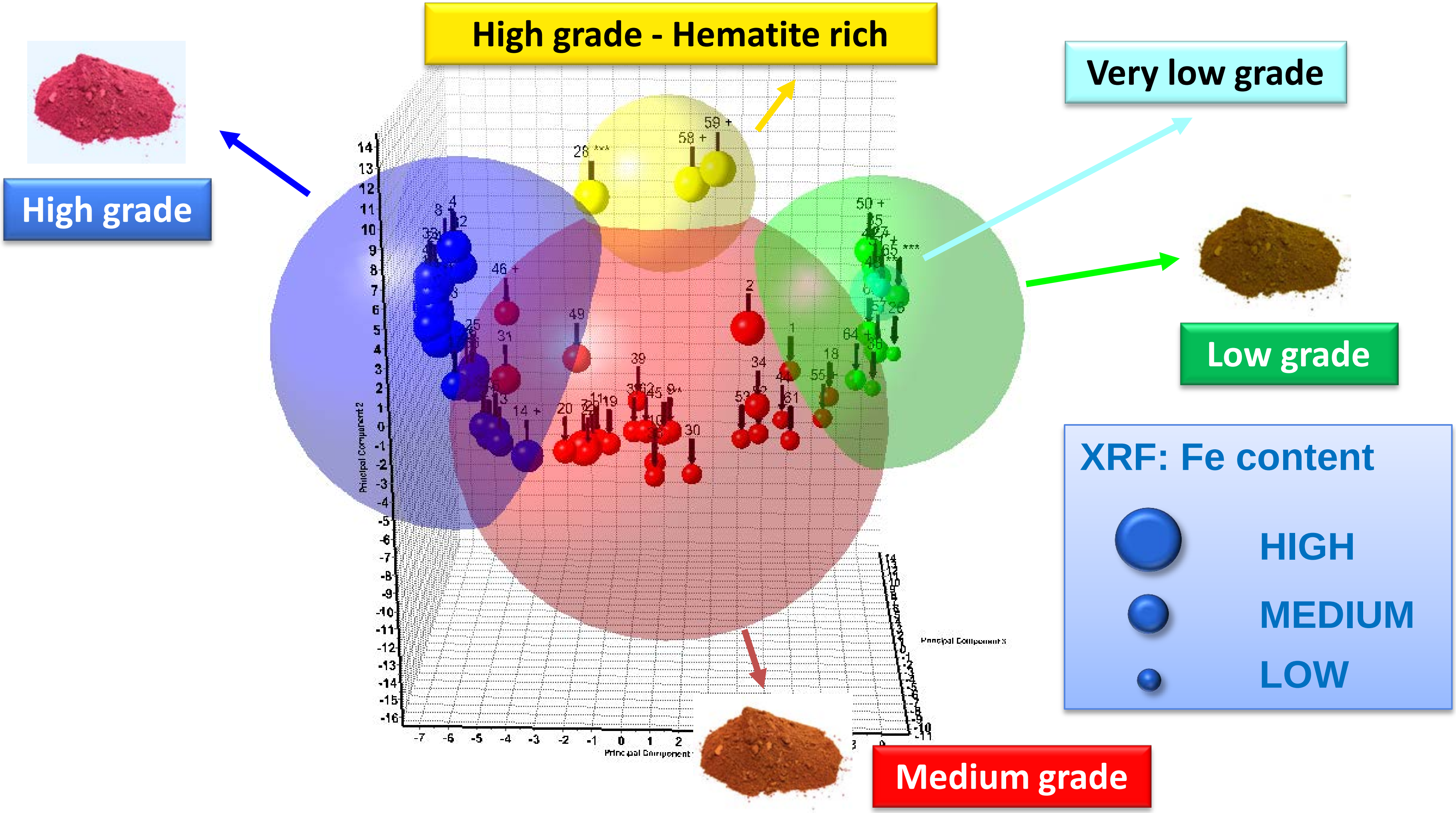
- Hematite (Hem) Fe_2O_3
- Goethite (Goe) FeOOH
- Quartz (Qua) SiO_2
- Kaolinite (Kao) $Al_2Si_2O_5(OH)_4$
- Gibbsite (Gib) $Al(OH)_3$



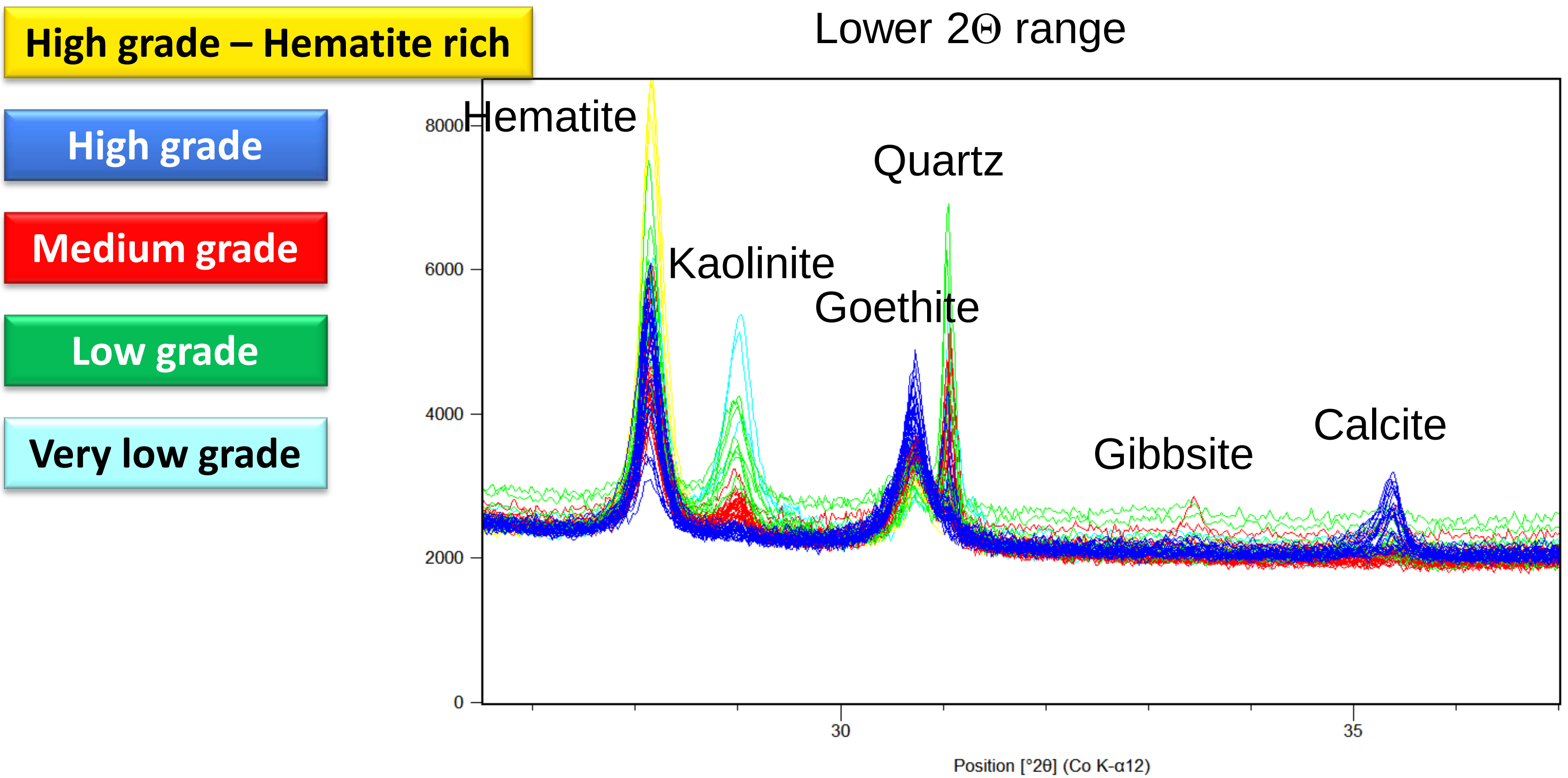
Case study - Rio Tinto Iron Ore – Rietveld quantification



Data evaluation of Rio Tinto samples – 5 Clusters



Data evaluation of Rio Tinto samples



- Scans are color coded according to their cluster color, text labels are indicating peaks of different mineral phases

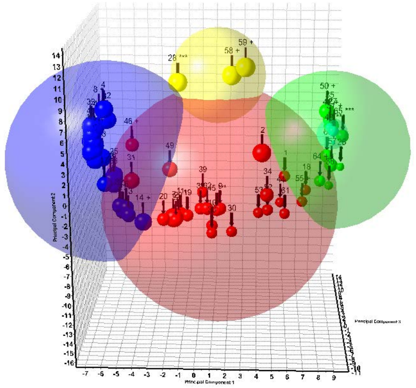
Data evaluation of Rio Tinto samples - Rietveld

- Rietveld quantification results for all representative and most outlying scans of all clusters:

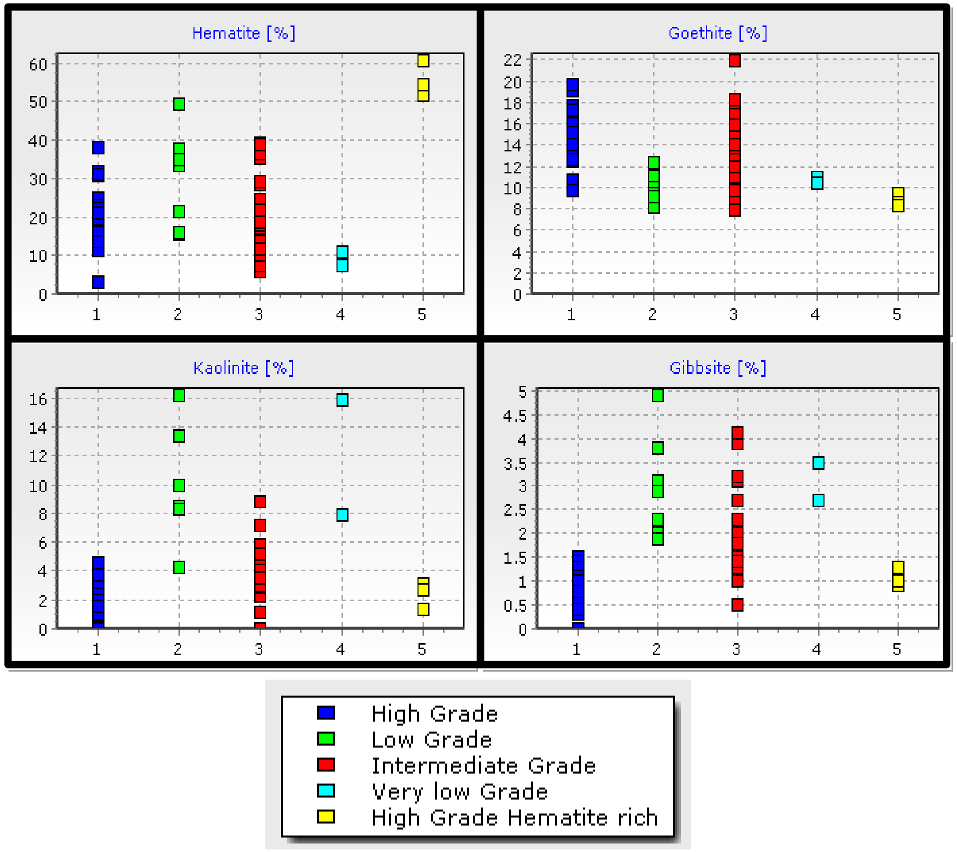
	Hematite [%]	Goethite [%]	Goethite, Al [%]	Gibbsite [%]	Kaolinite [%]	Quartz [%]	Magnetite [%]
TP0377287 ***	51.6	9.4	35.1	0.9	3.1	0	0
TP0377395 +	60.6	8.6	28.2	1.3	1.3	0	0
TP0377396 +	54.2	8.3	33.3	1	2.7	0.4	0
TP0377301 ***	23.1	16.4	57	0	2.9	0	0.6
TP0377291 +	3.1	14.8	79.5	0.5	2	0	0
TP0377262 +	38	19.4	36.7	1.5	4	0	0.4
TP0377399 ***	17.8	12.6	62.7	1.8	5.2	0	0
TP0377286 +	39.1	11.5	39.2	2.3	5.7	2.2	0
TP0377271 +	6.4	16.8	73	3.9	0	0	0
TP0377404 ***	35.1	11.1	40.6	4.9	8.3	0	0
TP0377402 +	16	10.5	66.3	2.9	4.3	0	0
TP0377303 +	34.3	8.1	37.7	3.1	13.4	3.5	0
TP0377258 +***	11	11	67.4	2.7	7.9	0	0
TP0377276 +	7.3	10.4	62.9	3.5	15.9	0	0

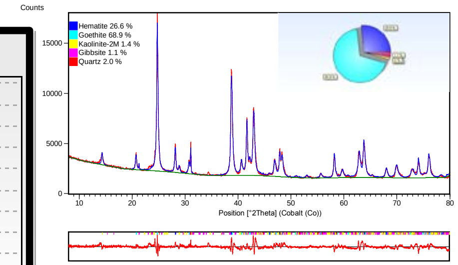
PANalytical

Case study - Rio Tinto Iron Ore – Cluster vs. Rietveld



Cluster analysis



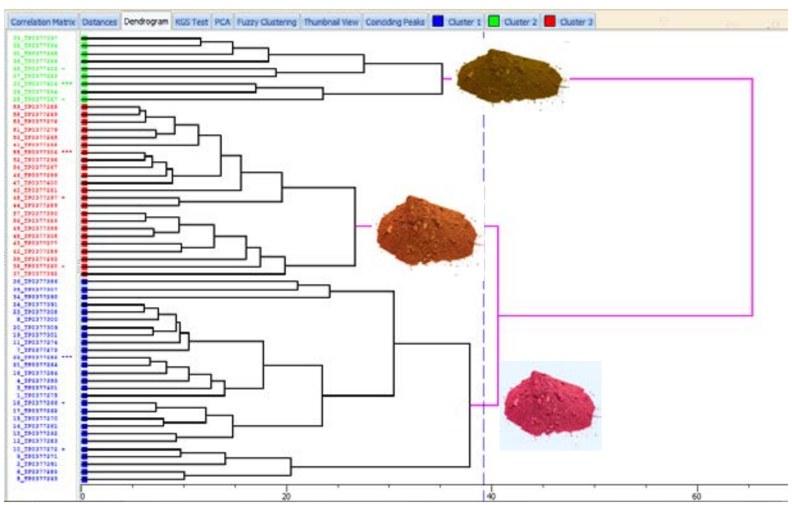


Rietveld analysis

Possible automation of Cluster Analysis

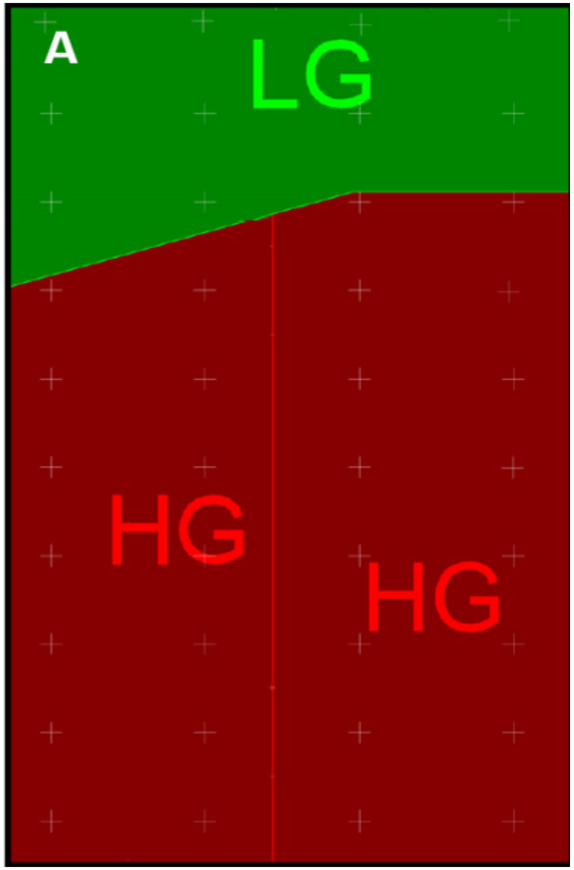
- Once a broad set of typical samples has been measured and analyzed, this set can be used for an automatic classification via cluster analysis:

Open master document containing training set → Add scan from the unknown sample → Execute cluster analysis using predefined parameters → Output to which cluster the new scan belongs



Case study - Rio Tinto Iron Ore – Grade blocks

Without XRPD



With XRPD

XRPD: Automatic Clustering & Rietveld

Detailed insight into deposit mineralogy

Cost-saving by enhanced beneficiation

Conclusion / Benefits

- Cluster analysis offers fast and reliable grade control
- No “human factor” for grade estimation
- Cost-effective
  - Wrong classification of 1 grade block represents a loss of 100 000 Euro
- Short measurement times (6min) for quick quality check of concentrator feed, concentrate and tailings
- Fine grained grade control enables enhanced (cost saving) beneficiation

- Fully automatable analysis, also in combination with XRF
- No chemicals involved
- Simple sample handling

PANalytical

X-RAYS HELP TO SEE !

Thank you for your attention!

Literature

Polysnap:
C. Gilmore, G. Barr and J. Paisley, High-throughput powder diffraction. I. A new approach to qualitative and quantitative powder diffraction pattern analysis using full pattern profiles, J. Appl. Cryst., 2004, 37, 231-242.
G. Barr, W. Dong and C.J. Gilmore, High-throughput powder diffraction. II. Applications of clustering methods and multivariate data analysis, J. Appl. Cryst., 2004, 37, 243-252.

Agglomerative clustering:
H. LOHNINGER, Data Analysis, Springer-Verlag, Berlin-New York-Tokyo, 1999, ISBN 3-540-14743-8.
G.N. LANCE, W.T. WILLIAMS,A general theory of classification sorting strategies 1., Hierarchical systems, Comp. J. (1966), ***9, 373 - 380.***

Number of clusters:
L.A. Kelley, S.P. Gardner and M.J. Sutcliffe, An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies, Protein Engineering (1996), ***9 1063 - 1065.***

PCA:
K.V. Mardia, J.T. Kent, J.M. Bibby Multivariate analysis. Academic Press, London, 1979.

Silhouettes:
J.P. ROUSSEEUW, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J. Comp. Appl. Math (1987), ***20, 53 - 65.***

Fuzzy Clustering:
6*M. SATO, Y. SATO and L.C. JAIN, Fuzzy Clustering Models and Applications, Studies in Fuzziness and Soft Computing vol.* ***9, Springer group (1997), New York, 122 pages.***

MMS:
K.V. Mardia, J.T. Kent, J.M. Bibby Multivariate analysis. Academic Press, London, 1979.