MINTEK

A global leader in mineral and metallurgical innovation

Python-based machine learning tools for metallurgical data clustering

Quinn Reynolds, Markus Erwee and Pranusha Moodley

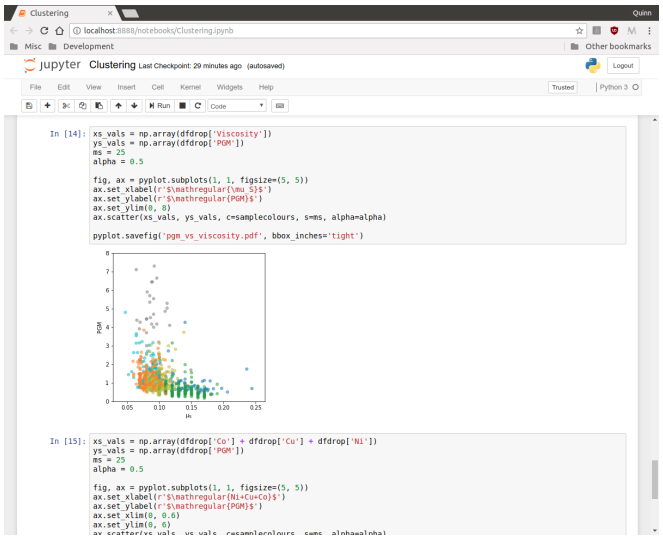
Pyrometallurgy Division, Mintek

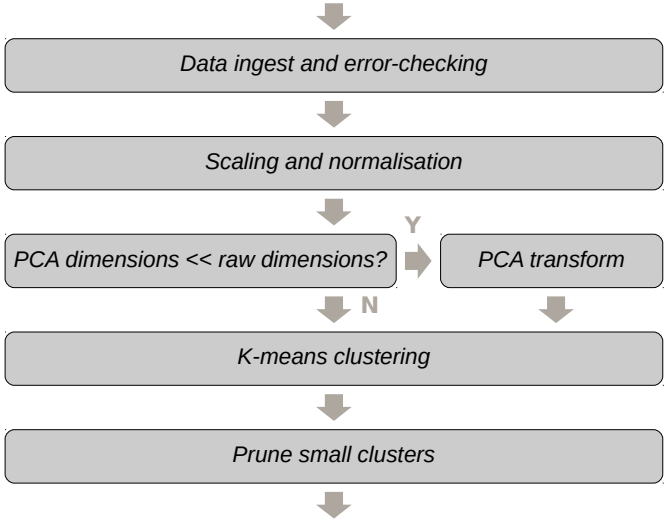
INTRODUCTION

- Pyrometallurgical smelters are engineering grand challenges…
  - Feed and product streams with **complex metallurgy and mineralogy**
  - **Multiple phases** (solid, liquid, gas) and phase changes between materials
  - Combination of **batch and continuous** processing for different streams
  - **Large range of temperatures**, ambient to > 1500°C
  - **Power input and cooling systems** balanced to maintain furnace integrity
- Plants generate large quantities of highly multidimensional data over time… but how best to use that data?

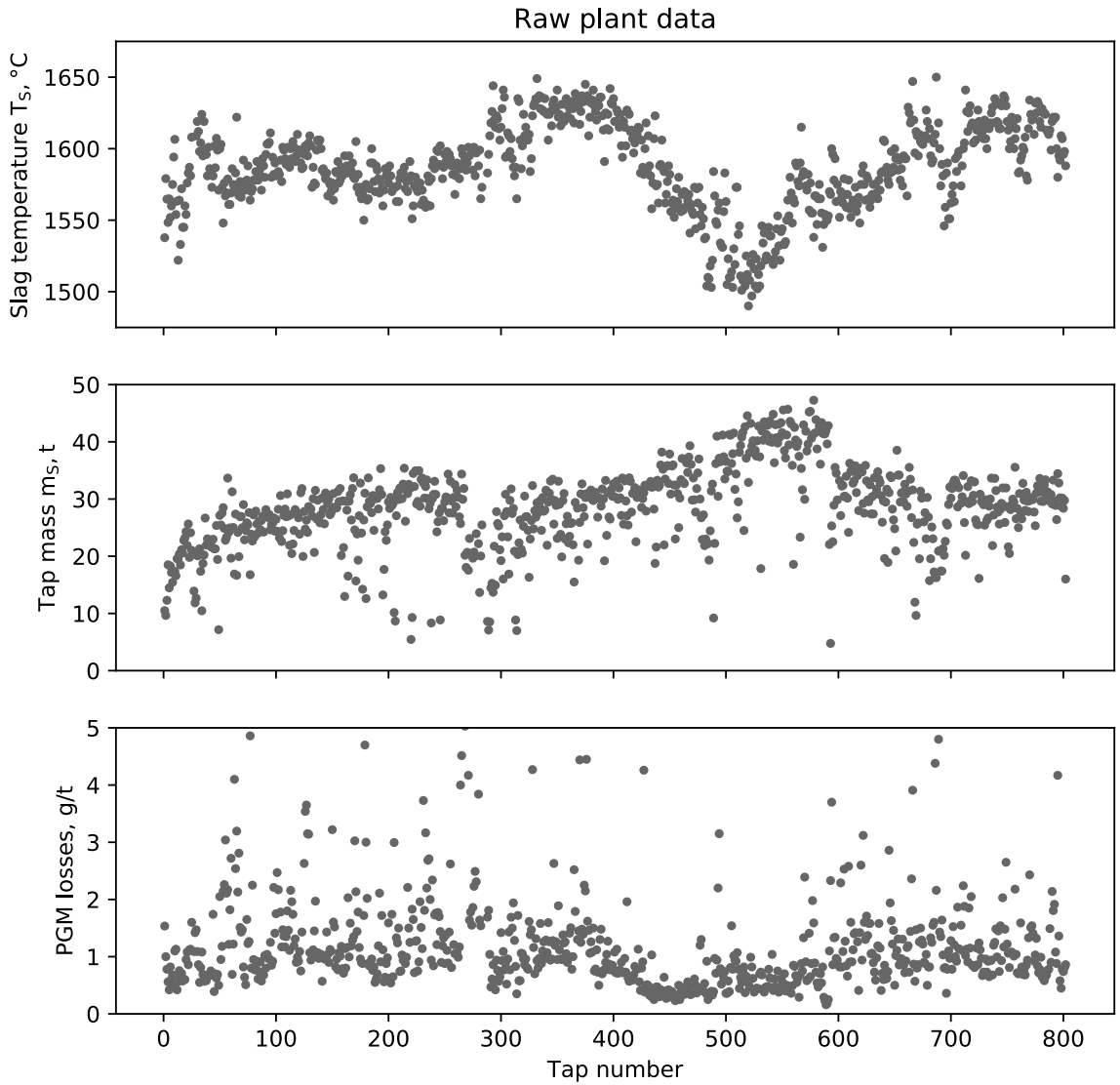
INTRODUCTION

- Modern data science tools can perform useful “**virtual assist**” functions by digesting large data sets into smaller summaries in rigorous and repeatable ways
- Plant operators and engineers are freed up to **develop intuition fast** and make good process decisions
- Python is a good language for this kind of analysis – heavy bias toward optimised scientific computing with powerful modules for **data science, data visualisation, and machine learning**

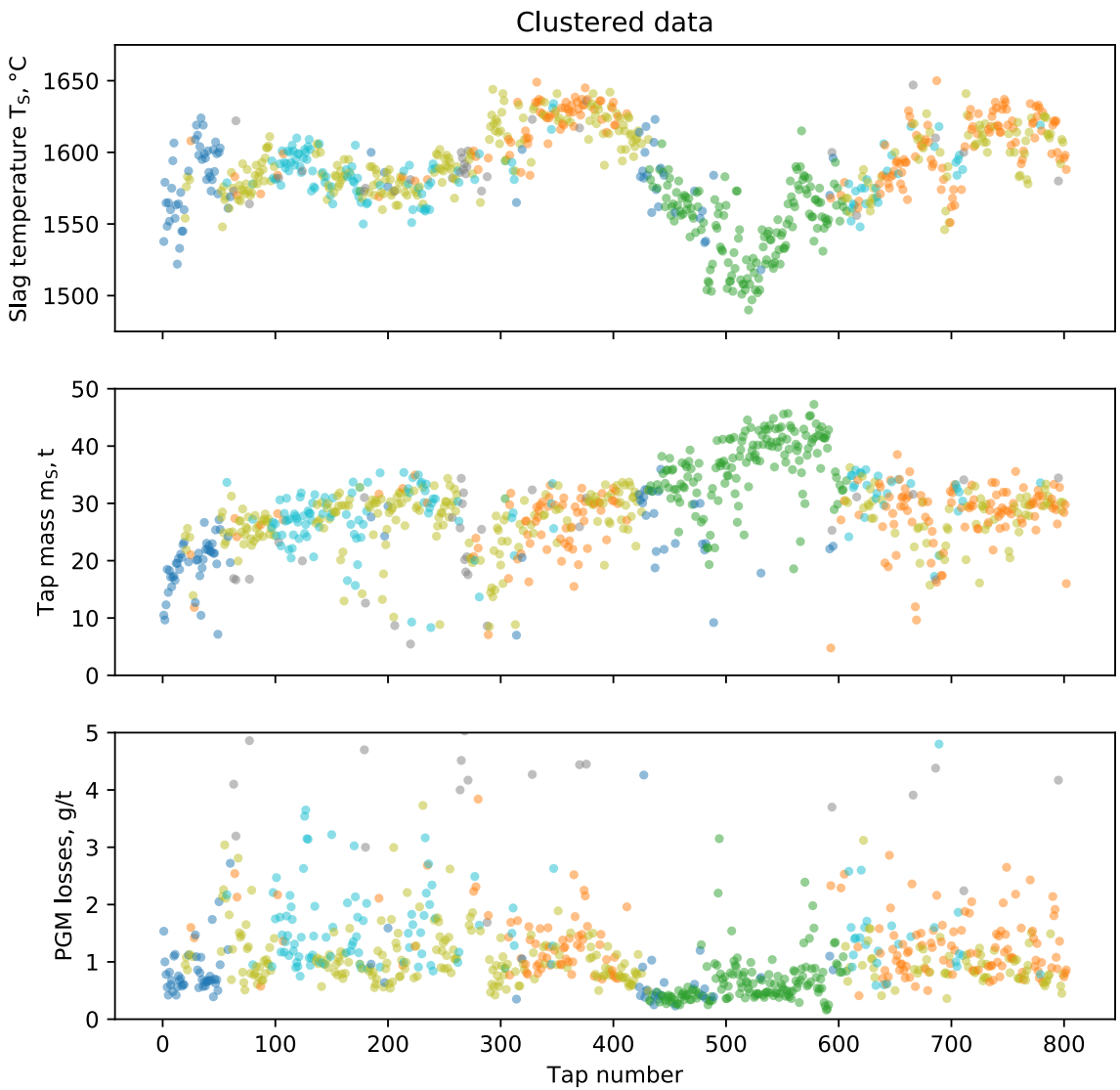




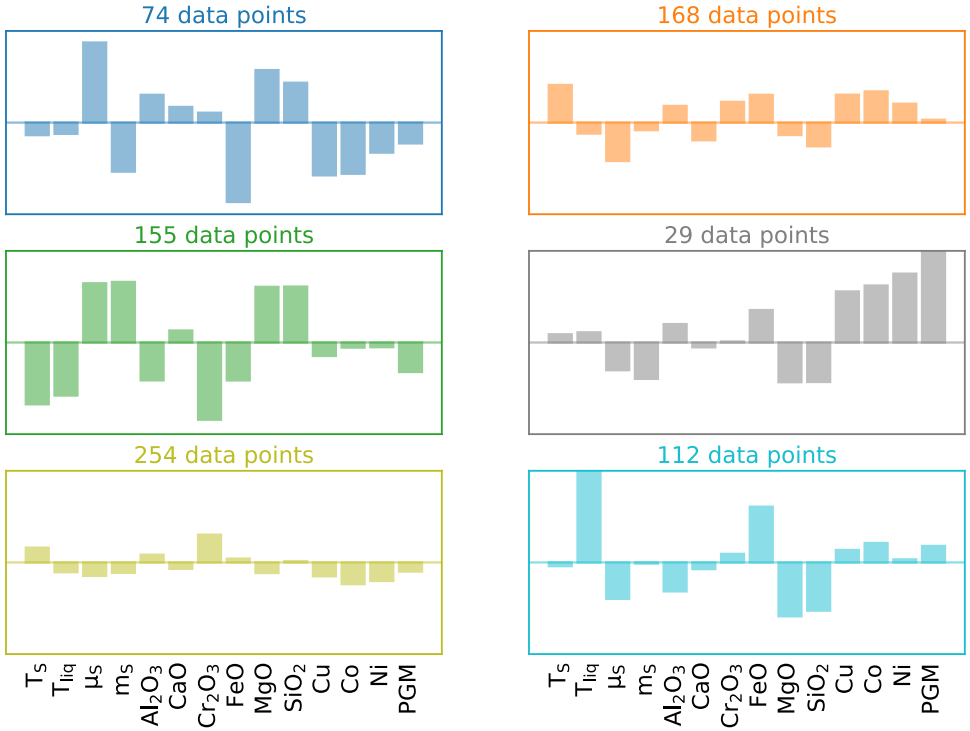
DATA CLUSTERING

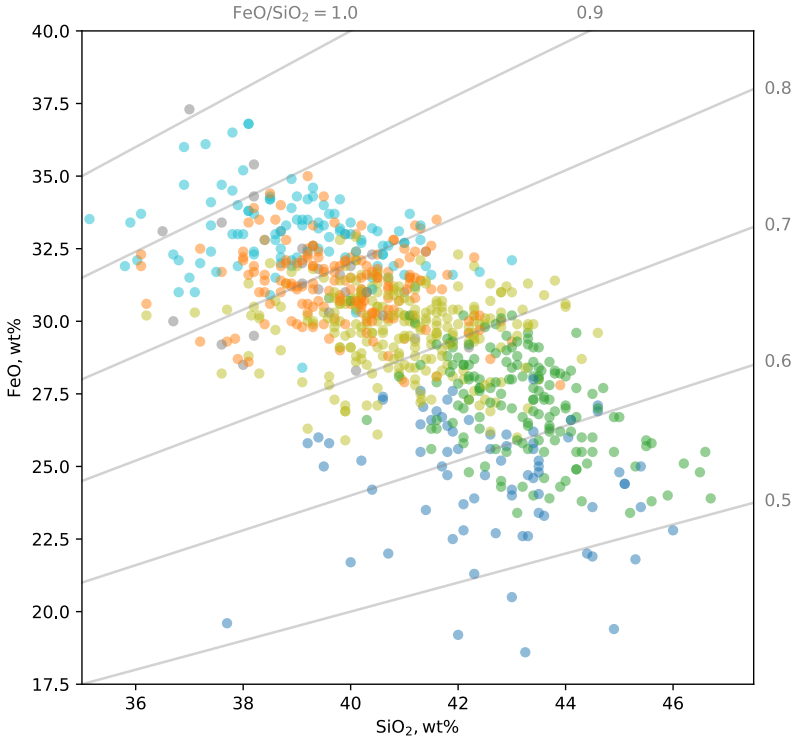


DATA CLUSTERING



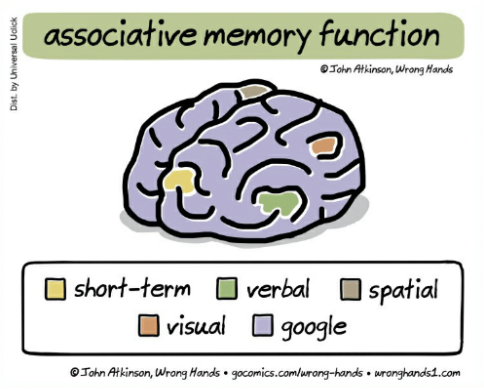
DATA CLUSTERING





CONCLUSIONS

- Nature and volume of data generated by modern pyrometallurgical plants are impractical for human beings to work with
- Data science and machine learning tools can provide rapid insight and condensed summaries of process behaviour for agile decision-making
- Highly controllable degree of data abstraction and filtering, tools can be adjusted to produce output appropriate for different user levels
- Need to be wary of machine tools being used to replace instead of complement user expertise – human brain is still the most flexible and intuitive decision-making instrument (for now…)



This is your brain on….

ACKNOWLEDGEMENTS

MINTEK

A global leader in mineral and metallurgical innovation

Python-based machine learning tools for metallurgical data clustering

QG Reynolds, MW Erwee, and P Moodley
Mintek, Pyrometallurgy Division, Private Bag X3015, Randburg 2125

Introduction

Pyrometallurgical process plants are notoriously complex in nature, and can occupy a large parameter space including the chemical compositions of multiple phases and materials, input and output stream conditions, state variables such as temperature and pressure, and many others[1]. These parameters are measured continuously on operating plants, generating massive, multidimensional data sets over time. Making sense of such data sets for rapid decision-making is a critical oversight and control function on pyrometallurgical plants, but is often challenging for less experienced engineers and operators in particular. This can lead to inefficiencies, off-spec product, and in extreme cases damage to plant equipment or danger to personnel. In order to assist with this process, a simple Python-based software tool for rapid interactive data ingest, clustering, and display was developed using the pandas data analysis module[2], scikit-learn machine learning module[3], and Jupyter Notebooks interface[4]. Some example screenshots are shown in Figure 1.

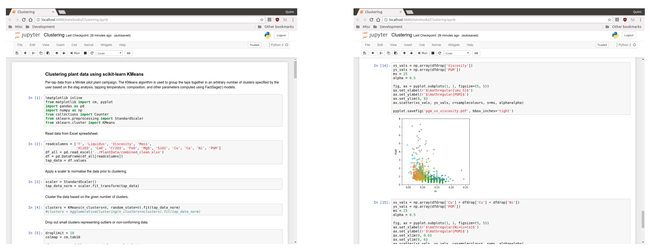


Figure 1: Jupyter Notebooks interface to clustering tool

Results and discussion

Data from an extended Mintek pilot plant campaign to smelt platinum-bearing raw material was mined to illustrate use of the software tool. To perform an analysis, measured data from slag taps (tapping temperature, tapped mass, slag composition by component) together with slag viscosities and liquidus temperatures calculated using the FactSage® thermochemical simulation package[5] are first read from a standard plant spreadsheet. The data is scaled to normalise the effects of variables with large differences in magnitude. Principal component analysis (PCA) is applied to the data to determine any internal dependencies, and dimensionality reduction by PCA transform is performed if required (in this case it was not). The resulting data is then grouped into a predetermined number of clusters using the K-means algorithm[6], with each element of the data set tagged with an identifier indicating which cluster it belongs to. During this process clusters with very small numbers of elements (< 10 in this case) representing outliers in the data set are optionally removed. The overall procedure is illustrated in Figure 2. It is clear that the data group into distinct periods when viewed as a time sequence with tap number. Comparing the groups with the coordinates of their cluster center vectors shown in Figure 3, it can be seen that the "dark blue", "green", and "yellow" conditions are linked with good results in terms of lower-than-average losses of platinum-group (PGM) and heavy metals (Co, Cu, Ni) to the slag, while the "light blue", "orange", and "grey" conditions perform more poorly.

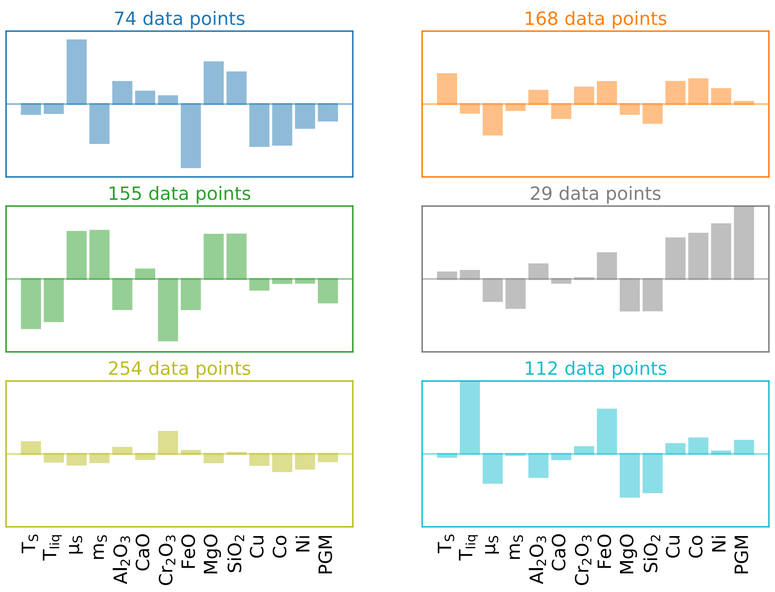


Figure 3: Cluster sizes and center coordinates (relative distances from overall mean values)

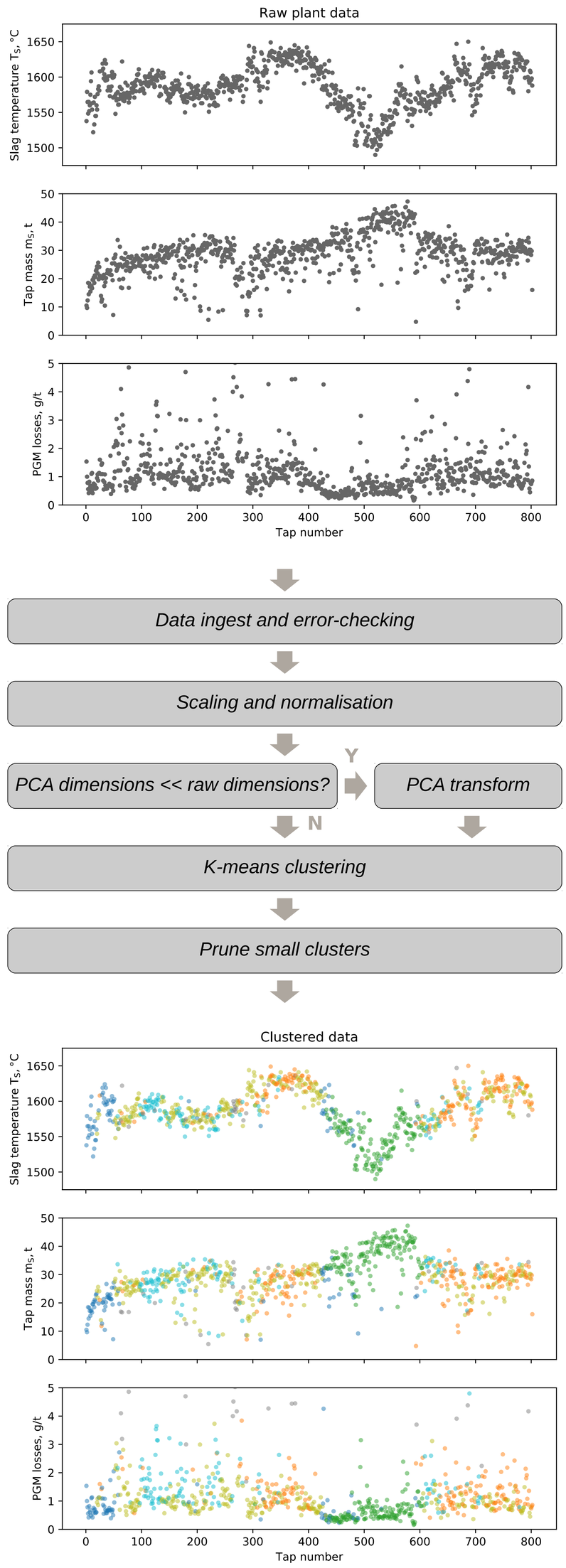


Figure 2: Procedure used to automatically group plant data from raw (top) to clustered (bottom)

Representative slag compositions are clearly identified by the cluster analysis, and suggest that relationships between certain slag components may be useful proxies for process performance. In this case human experts had already identified the FeO/SiO_2 ratio as a key indicator – this is supported by the results of the automated analysis, with low FeO and high SiO_2 correlating to clusters with better metal losses and vice versa. The data are presented graphically in Figure 4, and when combined with the cluster information show that lower FeO/SiO_2 ratios are indeed linked with better process performance. The dependence of some of the derived parameters on FeO/SiO_2 ratio is given in Figure 5, and also shows clear grouping by cluster. It is interesting to observe that better metal recoveries are associated with higher slag viscosity in this case, a somewhat counter-intuitive result.

Conclusions

Simple data reduction and visualization tools enable users to condense large sets of pyrometallurgical plant information into a small number of discrete conditions in a repeatable and consistent way, revealing important relationships in the data and providing rapid insight into the behaviour of the process. Such tools also offer a great degree of flexibility in the data sources they operate on – these can be sets of directly-measured plant parameters, numerical modelling results based on the measurements, or any combination of the two. It is expected that more refined "digital data assistant" systems using approaches similar to those discussed here will become increasingly commonplace in metallurgical plants in the future, as both the capability of machine learning tools as well as the complexity of ores and treatment processes increases.

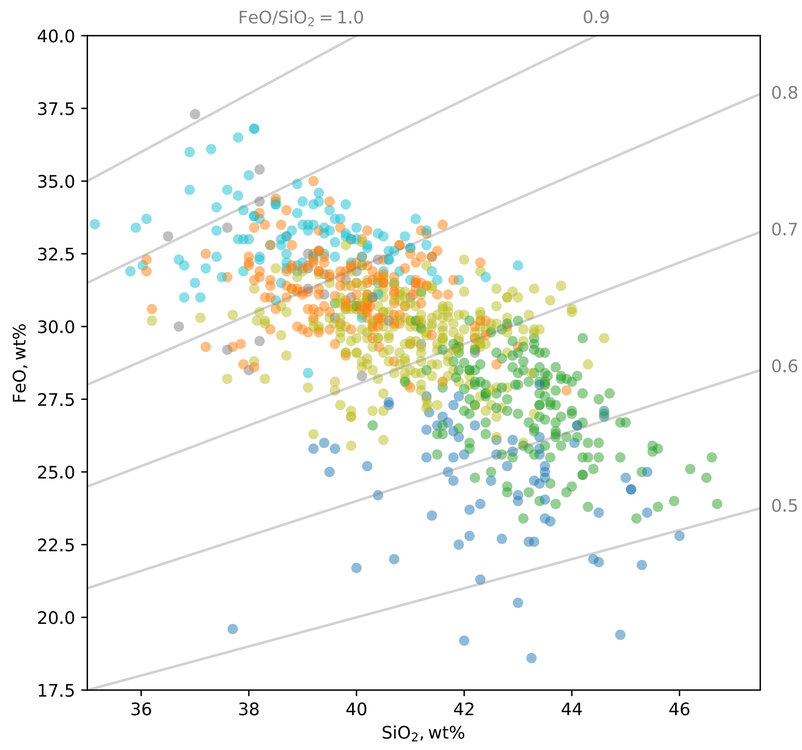


Figure 4: FeO/SiO_2 relationship showing cluster locations

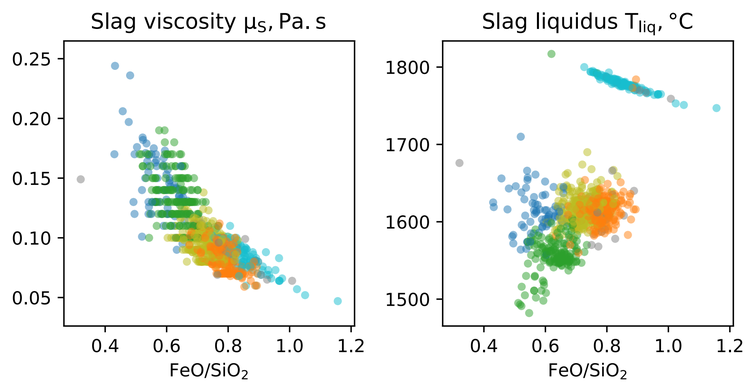


Figure 5: Grouping of slag properties with FeO/SiO_2 ratio

Literature cited

[1] S. Seetharaman, A. McLean, R. Guthrie, and S. Sridhar, "Treatise on Process Metallurgy vol. 2: Process Phenomena". Elsevier, Oxford (UK), 2014.
[2] "Python Data Analysis Library" [Online]. Available: https://pandas.pydata.org. [Accessed: 1-Jun-2018].
[3] "scikit-learn: Machine Learning in Python" [Online]. Available: http://scikit-learn.org. [Accessed: 1-Jun-2018].
[4]"Project Jupyter" [Online]. Available: http://jupyter.org. [Accessed: 1-Jun-2018].
[5] C. Bale, E. Bélisle, P. Chartrand, S. Decterov, G. Eriksson, A. Gheribi, K. Hack, I.-H. Jung, Y.-B. Kang, J. Melançon , A. Pelton, S. Petersen, C. Robelin, J. Sangster, P. Spencer, and M.-A. van Ende, "FactSage thermochemical software and databases 2010–2016". Calphad, 54, pp. 35-53
[6] S.P. Lloyd, "Least squares quantization in PCM". IEEE Trans. Inf. Theory, 28(2), pp. 129-137.

Acknowledgments

This poster is published by permission of Mintek.

For further information

Please contact markuse@mintek.co.za. More information on this and related research may be obtained at www.mintek.co.za/Pyromet/Index.htm#TechnicalPapers.

MINTEK
A global leader in mineral and metallurgical innovation

2030 NDP