Oral presentation

Bridging the gap between XRD, domain specific, and general, Machine Learning analysis

T. Degen[1], M. Sadki[1], M. Gateshki[1], E. Bron[1]

[1]Malvern Panalytical B.V. Lelyweg 1, 7603EA Almelo, The Netherlands
thomas.degen@malvernpanalytical.com

Nowadays the amount of XRD data to be analysed is drastically increasing by faster and more powerful Lab- as well as Synchrotron equipment. Additionally, the "depth" and complexity of structural analysis is increasing. In the past only well-ordered crystalline materials were deemed worth of analysis, for example by the Rietveld [1] method. Currently analysis shifts more and more to samples and methods for amorphous and semi-amorphous (nano) materials, in that case applying PDF analysis [2] and fitting. On top of this an increasing amount of Big Data methods are available, which can both simplify and speed up the analysis, like e.g. clustering, but may also be used to visualize hidden relations, using for example the t-SNE [3] method, Principal Component Analysis [4] or MMDS [5]. Combining all these highly domain specific and general methods can be a daunting task. In the analysis package HighScore(Plus) [6] all these methods can be automated by just dragging customizable analysis, treatment, report and scripting steps into a flowchart. Executing these complex flowcharts can be fully automated, and all time-consuming fitting tasks are highly parallelized. Automated workflows can include quick searches on free (COD database [7]) and commercial databases like ICDD PDF-5+ or the CSD [8] too, to find the closest matches and to extract the best physical models for subsequent fitting. In this presentation we give a short introduction how complete and complex workflows can be automated and easily transferred from one to another PC.

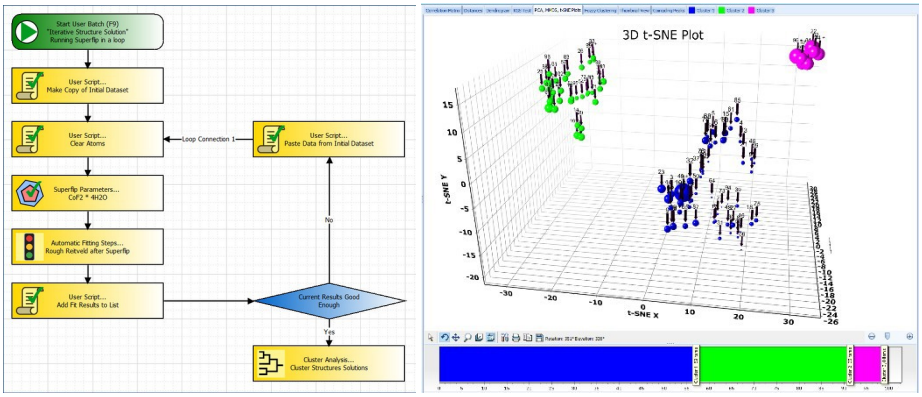

Figure 1. Example of an analysis flowchart (left) and a t-SNE plot (right), comparing fitting results of different structure solutions [9].

[1] H.M. Rietveld, A profile refinement method for nuclear and magnetic structures, *J. Appl. Cryst.* (1969), **2**, 65-71.

[2] A.K. Soper and E.R. Barney, Extracting the pair distribution function (PDF) from white beam x-ray total scattering data, *J. Appl. Cryst.* (2011), **44**, 714 -726

[3] L.J.P. van der Maaten and G.E. Hinton, Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* (2008)**, 9**, 2579-2605.

[4] H. Wold, (1966). Estimation of principal components and related models by iterative least squares, in P.R. Krishnaiaah, *Multivariate Analysis. New York: Academic Press.* (1966), 391-420.

[5] I. Borg, P. Groenen, Modern Multidimensional Scaling: theory and applications (2nd ed.). New York: Springer-Verlag, (2005), 207–212.

[6] T. Degen, M. Sadki, E. Bron, U. König & G. Nénert, The HighScore Suite, *Powder Diffr.* Vol. **29**, (2014), Supplement S2, 13-18.

[7] Grazulis, S., Chateigner, D., Downs, R. T., Yokochi, A. T., Quiros, M., Lutterotti, L., Manakova, E., Butkus, J., Moeck, P. & Le Bail, A. (2009). Crystallography Open Database – an open-access collection of crystal structures. *Journal of Applied Crystallography*, **42**, 726-729.

[8] C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Cryst.* (2016). **B72**, 171-179

[9] G. Oszlányi, A. Sütő, Ab initio structure solution by charge flipping, *Acta Cryst.* (2004), **A60**, 134-141.