

Artificial intelligence as a new tool for qualitative and quantitative phase analysis from X-ray powder diffraction data

Mattia Lopresti¹, Fabio Mora², Luca Palin^{1,3}, Marco Milanese¹

¹Università del Piemonte Orientale, Dipartimento di Scienze e Innovazione Tecnologica, Viale T. Michel 11, 15121 Alessandria, Italy; ² Italian Agile Movement, Via Guglielmo Oberdan 24, 25128 Brescia, Italy;

^{1,3}Nova Res s.r.l., Via D. Bello 3, 28100 Novara, Italy.

mattia.lopresti@uniupo.it

Abstract

X-ray powder diffraction (XRPD) is an important technique for qualitative, quantitative and textural analyses, which exploits the scattering of X-rays by matter. The output of the method consists in a pattern, unique for each material, that can be analyzed to extract precious information on the investigated sample. When polycrystalline materials are analyzed, the phase recognition can be carried out comparing the resulting pattern, collected on a mixture, to those associated to pure components. Even if this procedure, known as qualitative analysis, may appear as a relatively simple procedure, many experimental factors can heavily affect the performances of phase recognition, such as the peak superposition and the presence of preferred orientations. Quantitative phase analysis is an even more difficult task to be carried out, since the relations between signal and quantity in polycrystalline mixtures can be heavily affected by many parameters, being caused by sample features (preferred orientations, microabsorption, presence of low ordered and/or amorphous phases) or the use of different instruments or measurement setups (Debye-Scherrer geometry, Bragg-Brentano geometry). In the past century, many methods had been developed to try to overcome these difficulties, such as single-peak and whole profile fitting methods. Despite the efforts, quantitative analysis is still almost impossible to be carried out in many common situations in the real world industrial environment. The purpose of this project is to build an AI-based approach for both qualitative and quantitative phase analysis from XRPD data and to make it available for both industrial and academic communities as a new tool to be integrated among the already existing methods.

1 Introduction

X-ray powder diffraction (XRPD) is an important technique spread in both the industrial and academic worlds. The main advantage consists in the capability of analyzing materials using a sample-conservative approach giving qualitative, quan-

titative and textural information [Dinnebier e Billinge, 2008], exploiting the scattering of X-rays caused by crystalline materials. The result of the analysis is described by a pattern in which intensities are function of scattering angles, expressed as 2θ [Giacovazzo, 2011]. Patterns are unique for each phase and which can be used as a fingerprint [Winchell, 1927; Hanawalt e Rinn, 1936] for materials, as reported in figure 1. In this context, different phases can be recognized in a relatively easy approach by comparing the result of the XRPD analysis to patterns contained in databases of known substances. This approach is often used for distinguishing polymorphs of the same chemical compound (e.g. anatase, rutile and brookite for titanium oxide), but also components in polycrystalline mixtures. An important contribution to the qualitative analysis by XRPD was given by Hanawalt and colleagues [Hanawalt *et al.*, 1938]. The Hanawalt method consisted in calculating a set of parameters (e.g the d-spacings calculated from Bragg's law [Giacovazzo, 2011]) using the positions of the most intense peaks. These parameters, collected in dedicated books, are unique for each phase and can be used for phase recognition. The method was used for long time, but had some evident limitations, such as the impossibility of reporting millions of materials and their parameters, but also a problematic in obtaining correct d-spacings. In fact, if the sample is not accurately positioned in the sample holder and its height is not at the perfect height, the pattern present a shift in the peaks at higher or lower angles that affect the calculation of the d-spacings. Quantitative phase analysis is a more difficult task to carry out, since the intensity of the signal of each substance is bound to its quantity by very complex equations that vary with the sample composition, the instrumental setup and other kind of effects. For this purpose, different methods had been developed, some based on the measurement of single peaks and some based on the whole profile fitting approach [Dinnebier e Billinge, 2008]. Advances have been made in the past years, to which some of us have contributed with various published works. In particular, thanks to the use of multivariate methods it has been shown how it is possible to monitor changes within the crystalline structures [Palin *et al.*, 2015; Conterosito *et al.*, 2020] and how these methods are suitable for the study of mixtures, even complex ones, both in a qualitative and quantitative way [Guccione *et al.*, 2020; Mangolini *et al.*, 2021; Lopresti *et al.*, 2022]. Even using

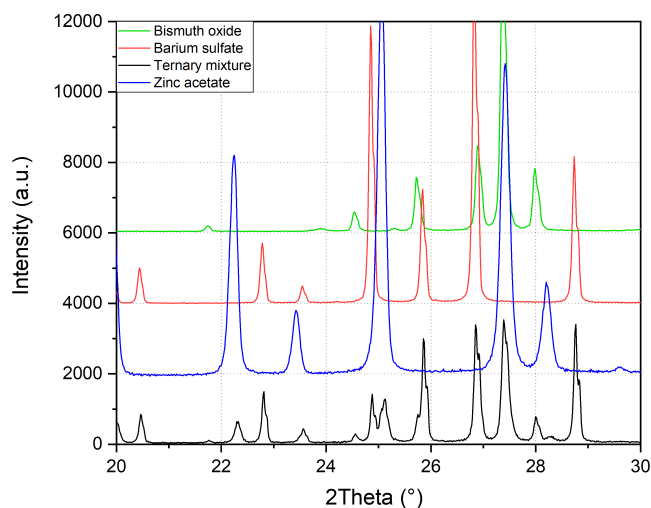


Figure 1: Example of XRPD patterns of three pure substances (barium sulfate, bismuth oxide and zinc acetate) compared to a ternary mixture of equal weighted components.

these methods, the task can be particularly difficult if in the presence of mixtures with very different densities and atomic numbers of the constituting elements. In this context, the use of artificial intelligence can play a crucial role in overcoming traditional methods as, if properly trained, it is able to overcome problems related to non-linear equations. Furthermore, the fact that the analyzes, both qualitative and quantitative, depend on many parameters, instrumental, structural and environmental, means that a system capable of evaluating enormous quantities of variables at the same time is the most suitable for this type of analysis.

1.1 Methods and common issues for qualitative phase analysis

At the present day, the qualitative analysis is carried out using dedicated softwares that processes the diffraction pattern (i.e. QualX [Altomare *et al.*, 2015]), identifying the peaks and their positions in 2θ . These values, uniquely linked to the structure of the phase, are compared with those contained in a database in a procedure that also provides for small deviations at greater or lesser angles for a greater probability of identification of the corresponding substance. As aforementioned, for small deviations (zero errors), softwares are generally able to compensate, but when these are added to different substances, distinguishing different substances becomes particularly difficult. Moreover, if the samples are prone to show preferred orientations (PO), the relative intensities of the peaks may be offset. In the worst case a subset of peaks does not appear on the diffraction pattern, while relative intensities of other ones can be heavily influenced [Giacovazzo, 2011]. For the said reasons, even if the approach is conceptually simple, the task of phase recognition can be difficult even for common industrial situations.

1.2 Methods and common issues for quantitative phase analysis

Methods for quantitative phase analysis from XRPD data can be divided in three groups: single peak methods, whole profile fitting methods and multivariate methods. The first group, as suggests the name, involves the use of single peaks belonging to different substances in a polycrystalline mixture pattern. The quantification is then made by comparing the intensities of such peaks to the ones belonging to pure phases to extract the relative ratio of each substance contained in the polycrystalline mixture. This method can be performed also by adding a known quantitative of a known phase to the mixture, in order to quantifying the others by comparing the intensities to the pure ones. This method is efficient in mixtures of polymorphs of the same chemical substance, which have very similar mass absorption coefficients (μ_m^*) and consequently scatter X-rays affecting the outgoing intensities in the same way. The second method involves building complex nonlinear regression model having *a priori* information about the crystal structure of the phases contained in the analyzed sample. The most famous and exploited one is known as Rietveld refinement [Dinnebier *et al.*, 2018]. This method is one among of the most used, but can be heavily affected by PO and MA, as reported in a recent article of some of us [Mangolini *et al.*, 2021]. Moreover, Rietveld refinement has its greatest issue in requiring a known crystal structure for simulating the powder diffraction pattern and the knowledge on the instrument on which the sample has been measured, to estimate the effects of the instrumental parameters on the quality of the measurement. Multivariate analysis (MA) is a relatively new approach to the study of XRPD data for quantitative purposes and its result are very promising. The greatest advantage in using multivariate analysis consists in not requiring known crystal knowledge for the analysis, therefore avoiding biases due to the forecast peak intensities for each specie that can be obtained by calculating a XRPD pattern starting from the crystallographic information. In a our recent article, submitted to the Journal of Applied Crystallography, we exploit the potentialities of the multivariate methods and give different approaches that can be used to analyze difficult system and comparing them to the traditional Rietveld refinement method and to PONKCS (Partial or no-known crystal structure) method [Scarlett e Madsen, 2006; Scarlett e Madsen, 2018], which is an hybrid Rietveld-MA approach to the quantitative phase analysis.

2 Purposes and challenges

The purpose of the project is to create a brand-new AI-based methodology for qualitative and quantitative analysis from XRPD data. Such an approach is completely new in the crystallographic field and has huge growth perspectives for the next decades. Both industrial and research communities of many different areas can benefit from the development of such a new tool. Challenges of this project can be summarized in the following list of partial objectives:

- Evaluation of the *a priori* information required by an AI system to perform qualitative analysis (i.e. availability of structure knowledge or usage of patterns only).

Data about phases can be collected from open repositories such as the Crystallography Open Database.

- Training of an AI for the qualitative phase analysis on XRPD data starting from methods reported in scientific literature on smaller datasets [Lee *et al.*, 2020].
- Evaluation of the performances of the qualitative analysis and, if the expectations are met, analysis of the upgrade for performing quantitative phase analysis.
- Training of an AI for the quantitative phase analysis on XRPD data.

References

- [Altomare *et al.*, 2015] Angela Altomare, Nicola Corriero, Corrado Cuocci, Aurelia Falcicchio, Anna Moliterni, e Rosanna Rizzi. QUALX2.0: a qualitative phase analysis software using the freely available database POW_COD. *Journal of Applied Crystallography*, 48(2):598–603, February 2015.
- [Conterposito *et al.*, 2020] Eleonora Conterposito, Mattia Lopresti, e Luca Palin. In situ x-ray diffraction study of xe and CO₂ adsorption in y zeolite: Comparison between rietveld and PCA-based analysis. *Crystals*, 10(6):483, June 2020.
- [Dinnebier *et al.*, 2018] Robert E Dinnebier, Andreas Leineweber, e John SO Evans. *Rietveld refinement: practical powder diffraction pattern analysis using TOPAS*. Walter de Gruyter GmbH & Co KG, 2018.
- [Dinnebier e Billinge, 2008] R E Dinnebier e S J L Billinge, editors. *Powder Diffraction*. Royal Society of Chemistry, 2008.
- [Giacovazzo, 2011] Carmelo Giacovazzo. *Fundamentals of crystallography*. Oxford University Press, Oxford New York, 2011.
- [Guccione *et al.*, 2020] Pietro Guccione, Mattia Lopresti, Marco Milanese, e Rocco Caliendo. Multivariate analysis applications in x-ray diffraction. *Crystals*, 11(1):12, December 2020.
- [Hanawalt *et al.*, 1938] J. D. Hanawalt, H. W. Rinn, e L. K. Frevel. Chemical analysis by x-ray diffraction. *Industrial & Engineering Chemistry Analytical Edition*, 10(9):457–512, September 1938.
- [Hanawalt e Rinn, 1936] J. D. Hanawalt e H. W. Rinn. Identification of crystalline materials. *Industrial & Engineering Chemistry Analytical Edition*, 8(4):244–247, July 1936.
- [Lee *et al.*, 2020] Jin-Woong Lee, Woon Bae Park, Jin Hee Lee, Satendra Pal Singh, e Kee-Sun Sohn. A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic XRD powder patterns. *Nature Communications*, 11(1), January 2020.
- [Lopresti *et al.*, 2022] Mattia Lopresti, Beatrice Mangolini, Marco Milanese, Rocco Caliendo, e Luca Palin. Multivariate vs. traditional quantitative phase analysis of x-ray powder diffraction and fluorescence data by mixtures showing preferred orientation and microabsorption. *Submitted to J.Appl.Cryst*, 2022.
- [Mangolini *et al.*, 2021] Beatrice Mangolini, Luca Palin, Marco Milanese, e Mattia Lopresti. XRF and XRPD data sets in ternary mixtures with high level micro-absorption and/or preferred orientations problems for phase quantification analysis. *Data in Brief*, 36:107043, June 2021.
- [Palin *et al.*, 2015] Luca Palin, Rocco Caliendo, Davide Viterbo, e Marco Milanese. Chemical selectivity in structure determination by the time dependent analysis of in situ XRPD data: a clear view of xe thermal behavior inside a MFI zeolite. *Physical Chemistry Chemical Physics*, 17(26):17480–17493, 2015.
- [Scarlett e Madsen, 2006] Nicola VY Scarlett e Ian Madsen. Quantification of phases with partial or no known crystal structures. *Powder Diffraction - POWDER DIFFR*, 21, 12 2006.
- [Scarlett e Madsen, 2018] Nicola V. Y. Scarlett e Ian C. Madsen. Effect of microabsorption on the determination of amorphous content via powder x-ray diffraction. 33(1):26–37, March 2018.
- [Winchell, 1927] A. N. Winchell. “Finger prints” of minerals. *American Mineralogist*, 12(6):261–262, 06 1927.