Applying multivariate analysis to X-ray diffraction computed tomography: the study of medieval applied brocades

Pierre Bordet, Florian Kergourlay, Ariane Pinto, Nils Blanc, Pauline Martinetto

ARTICLE

Applying multivariate analysis to x-ray diffraction computed tomography: the study of medieval applied brocades.

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Pierre Bordet [*a], Florian Kergourlay [a], Ariane Pinto [ab], Nils Blanc [a], Pauline Martinetto [a]

X-ray diffraction tomography is a well-developed technique to study the structure of heterogeneous materials which makes it a tool of choice for the non-destructive investigation of cultural heritage microsamples. Characterizing such complex materials with a high enough spatial resolution requires acquiring large amounts of diffraction images, followed by a complex sequence of data management to localize the crystalline phases in the sample. Here we propose to use multivariate analysis in order to automatically decompose the data in a small set of components, each representing the diffraction pattern of one or a small number of phases. This makes phase identification and quantification of each component much more efficient and leads to a quantitative knowledge of the phase content in each voxel of the tomographic reconstruction. We show that Non-negative Matrix Factorization is very efficient for this purpose, with a computing time well compatible with in-line data analysis in order to assess the quality of measurements during experiments at synchrotron beamlines. Here, we apply the method to the investigation of microsamples from medieval applied brocade decors selected from a wooden statue, and the results are validated *a posteriori* by comparison with *ex situ* destructive techniques.

1. Introduction

For almost a century, X-ray powder diffraction has been the technique of choice for the structural investigation of solid matter, starting from simple phase identification to the ab initio structure solution of complex organic molecules, quantification of multiphase mixtures, or investigation of microstructures. These developments have been closely linked to the improved performances of x-ray sources, instruments and analytical methods, being in the laboratory or at dedicated synchrotron facilities.

More recently, with the progresses of powerful 3rd generation synchrotron sources offering micron size monochromatic beams of very high intensities, measurements allowing 3D computed tomographic reconstructions from x-ray diffraction data have become possible[1,2,3]. In such XRD-CT experiments, a small pencil x-ray beam is scanned through the sample and at each scan position, XRD patterns are recorded while the sample is rotated over an axis perpendicular to the beam. Reconstruction algorithms allow building 3D graphical representations of the crystalline and amorphous phase distribution in the sample, based on the selection of a characteristic diffraction peak or diffuse scattering signal. The quality of the reconstructions mainly depends on the size of the rotation angular step, and the spatial resolution is determined by the beam size and hence the translation scan step. Allowing to obtain a 3D map of the crystalline as well as amorphous phases in a sample, XRD-CT is complementary to other x-ray based tomographic techniques such as absorption-based x-ray tomography (X-CT) and fluorescence x-ray tomography (XRF-CT). The former informs on the variations of electron density in the sample, while the latter provides the distribution of chemical elements.

Since the appearance of the XRD-CT technique, a number of efficient software has been developed to perform the reconstruction and examination of the measurements[4,5]. The data are generally organized as a 3D matrix (or data cube), named as (T,R,D) hereafter in the text where 2 dimensions represent the coordinates of a scan points in translation (T) and rotation (R) and the 3rd dimension represents the XRD pattern (D) as function of the scattering parameter (e.g. the 2θ angle or $Q=4.\pi.\sin\theta/\lambda$, λ being the radiation wavelength) measured at each (T, R) point. The data cube can be represented graphically as 2D R-T sinograms for a particular slice in 2θ of the XRD pattern, showing the distribution of the crystalline phases having characteristic Bragg peaks in the selected slice. The same can be also visualized in real space after tomographic reconstruction using well-known algorithms. This procedure allows identifying and locating the different crystalline phases in the sample. Since nowadays 2D diffraction images are generally recorded, a preliminary step consists in azimuthally integrating the data to yield 1D diffraction patterns which can then be inserted in the data

[a] Univ. Grenoble Alpes, CNRS, Institut Néel, 38000 Grenoble, France.
[b] ARC-Nucléart, CEA, 38000 Grenoble, France.
Electronic Supplementary Information (ESI) available. See DOI: 10.1039/x0xx00000x

cube. In addition, the XRD patterns corresponding to an arbitrary volume of the sample can be obtained by summing the patterns from the voxels chosen in the real space reconstruction. By separating the contributions from the different phases, this so-called reverse analysis provides a way to identify the minor ones with better accuracy than would be feasible on the pattern from the whole sample, and to dramatically improve the detection limit of the technique.

Cultural Heritage Sciences are developing a growing interest for the technical skills shown by ancient artists and craftsmen in the fabrication of artworks and artefacts. In the need to identify recipes and material sources and to understand their degradation over time for restoration purpose, the contribution of x-ray based techniques has become more and more important. With the recent advent of mobile XRF and XRD instruments, non-invasive measurements can now be performed in situ, avoiding the need to move the artwork from its location. 2D mapping of artworks with flat surfaces (paintings, frescoes, manuscripts...) now allows observing the distribution of pigments, binders, etc. at micron size resolutions[6,7]. However, in many cases the detailed investigations of the sample design, stratigraphy or degradation damages require obtaining a three-dimensional picture from microsamples carefully selected from the investigated artwork. X-ray based tomographic methods are then very well suited to thoroughly determine the phase composition and stratigraphy of such microsamples in a non-destructive way, leaving them untouched for complementary analyses. They allow, for example, to visualize the full 3D arrangement of pictorial layers, without having to destructively prepare the microsample[8,9].

With the development of new generation synchrotron machines, the availability to submicron size beams allowing to rapidly perform such kind of experiments at higher resolutions is quite an enthusiastic perspective for the investigation of complex materials such as those from the Cultural Heritage field. However, with the foreseen new beamlines' performances, scanning a 1 mm size sample with 1 µm translation and 360 rotation steps (a quite typical case) yields in a few minutes 360 000 XRD patterns. This overwhelming amount of data has then to be processed in an efficient way in order to extract in a realistic timespan a usable, if possible quantitative information. Furthermore, despite the obvious interest of XRD-CT for Cultural Heritage Science, several technical difficulties still exist which can limit its use. Due to the relatively large gauge size, the voxels may contain more than one phase. It must be noted here that decreasing the gauge size for XRD-CT has a limit, since a sufficient number of grains must be contained in the gauge volume for powder averaging. Micron-size layers are not rare, for example in gilding layers. Hence, some quantitative (e.g. Rietveld) refinement of the XRD patterns calculated for each voxel is necessary to reach a precise, at least semi-quantitative description of the variation of sample composition and structure. Even when a quantitative knowledge of phase proportion is not required, such fitting may be necessary to identify minor phases. If sequential Rietveld refinement would be possible in theory, in practice it seems quite unrealistic given the very large amount of data, the random variation of phase contents (phases can appear and disappear abruptly from one voxel to the next) and the possibly large number of phases. Furthermore, such heavy data processing with no guarantee of rapid success requires a long time and cannot be considered for quality check of the data during experiments.

Bearing these problems in mind, we have investigated the use of multivariate analysis techniques for the processing of XRD-CT data collected on microsamples taken from ancient artworks. Here we compare the use of various algorithms, like Principal Component Analysis (PCA), supervised and non-supervised Multivariate Curve Resolution - Alternating Least-Squares (MCR-ALS) and Non-Negative Matrix Factorization (NMF). Using the latter, we show that a typical data cube can be automatically decomposed in a small number of crystallographically significant components within few minutes. These components which may contain more than one phase, can be submitted to quantitative analysis using Rietveld refinement, leading to a complete description of the 3D phase distribution in the sample.

Multivariate analysis techniques have been used recently in Materials Science in a wealth of problems describing the evolution of a chemical system as a function of an external parameter, such as time, temperature, pressure, etc., which involves processing large experimental data sets coming from various types of measurements. For these techniques to be relevant, it must be possible to find a small number of invariable components so that each individual measurement representing a state of the system can be described as a linear combination of the invariable components. Only the concentrations (or weights) of the components vary with the external parameter. The problem is thus to determine from the measurements the nature of the components and their weights for each value of the external parameter. This can be expressed mathematically by the relation:

$$D = CS+E$$

where D is a n.m matrix representing the measurements, n being the number of data sets (spectra, diffractograms, etc.) and m the number of points in each of them (as a function of 2θ, Q, etc.), C is a n.d concentration matrix containing the concentrations of each or the d components for each of the n measurements, S is a d.m matrix containing the d components of dimension m. E is the error matrix which represents the difference between the experimental and computed data and must be minimized by the algorithm.

Among many other applications, one of the most widespread technique in Chemometrics, PCA, has been used for more than 20 years for the processing of X-ray absorption data (XAS)[10]. For PCA, the components are considered as orthogonal and do not necessarily have a direct physical meaning. The optimal number of components to use for the decomposition (d) can be estimated from the norm of E(d) and by visual inspection of the components. For higher rank components, the signal becomes essentially noise and does not carry relevant information. This procedure is indeed a way to reduce the statistical noise from a dataset by reconstructing it using a

limited number of components. The selected number of components can be then used to launch a MCR-ALS minimization[11-14]. Like PCA, the MCR-ALS algorithm maximizes the explained variance in the data, but physical or chemical constraints can be applied to C and/or S matrices, rather than the orthogonality constraint used in PCA. In the ALS process, C is least-squares refined with fixed D and S, and then S is refined with fixed D and the previously refined C. The refinements are alternately carried out until convergence is reached. The main interest of the process lies in the possibility to input constraints such as e.g. positivity to S and/or C, which at the end will lead to physically/chemically sensible results[15]. A set of components defining a starting S matrix must be provided to launch the process, for which purpose, several kinds of algorithms have been devised[12]. One interesting feature of MCR-ALS is the possibility to introduce predefined components in the starting S matrix, for example diffraction patterns from isolated phases measured in the same condition. In this so-called "supervised" case, the components are not refined and only their weights are computed.

However, although they have been largely used for various types of spectral analyses, these techniques have been only scarcely applied to diffraction data. Burley et al. applied PCA to the analysis of time-resolved XRPD data[16]. Chapman et al. also applied PCA to the analysis of pair distribution function data[17]. Rodriguez et al. used MCR-ALS for an operando XRD study of battery discharge[18], and Taris et al. for time-resolved XRD data of cementing reactions[19]. Long et al. used a similar technique: Non-Negative Matrix factorization (NMF) for the identification of crystalline phase mixtures from thin film by XRD[20]. More recently, the same was successfully applied to combinatorial phase diagram investigations[21]. Opposite to PCA or MCR, the number of components must be supplied a priori to NMF but no starting values for the components are required. Therefore, no supervised process is possible. Due to the non-negativity constraints on the S and C matrices inherent to the algorithm, the components contained in S are expected to bear a physical meaning, i.e. they will consist in the scattering signal from one or a few phases from the mixture.

We report herein the application of multivariate analysis to XRD-CT data obtained from microsamples taken from Cultural Heritage artefacts. Such samples can be composed of many different phases of different natures, from mineral to organic, and with various crystallinity states, from fully amorphous to large crystal grains. They can be intimately mixed at the micron scale and the knowledge of their respective location, layer's thicknesses, etc., is important for the identification of the recipe used by the ancient craftsmen. Therefore, Cultural Heritage samples constitute an excellent playground for developing and testing these techniques.

Here, this experimental strategy has been applied to investigate the local specificities of manufacturing so-called applied brocade decors in the Savoy duchy. Applied brocade is a tin relief medieval polychromy technique developed during the first half of the 15th century in the Low Countries which then spread into all western Europe, which allowed to mimic the rendering of the sumptuous medieval clothes embroidered with gold or/and silver yarns. Its basic component is generally a sheet of tin foil pressed into a mold previously incised with the decorative pattern. The tin sheet is then filled with some pasty material, unmolded and applied on the surface to be decorated. It is then gilded with a gold/silver foil and sometimes highlighted with a colored glaze. These complex artefacts have been submitted to several kinds of degradations over time, such as extensive over-paintings, oxidation of metallic layers, etc. As part of an ongoing interdisciplinary study, our work aims at a detailed description of applied brocade in 15th and 16th century Savoy Duchy[22-25].

2. Experimental

2.1. Sample

For this study, we have focused our attention on a wooden Pièta statue coming from the Montrottier castle located in Lovagny, France (Figure S1), onto which applied brocade decorations had been identified by visual inspection. In the course of its restoration at the ARC-Nucleart laboratory (CEA, Grenoble), the statue was submitted to detailed *in situ* inspection using a portable XRD/XRF instrument, confirming at specific places the presence of tin-containing phases and gold, characteristic of the presence of applied brocades[25]. In order to gain insight into the composition, stratigraphy, manufacturing and degradation of these artefacts, several microsamples were collected for further investigations using synchrotron XRD/XRF-CT, at various locations were the presence of brocades had been confirmed using the portable x-ray instrument. These microsamples are chips of millimetric size in their larger dimension and a few hundred microns thick. They were sampled so as to preserve the entire stratigraphy from the preparation layers to the brocade system itself, including possible overpaintings. The selected microsample locations on the statue are shown in Figure S1, and a typical one is shown in Figure 1.

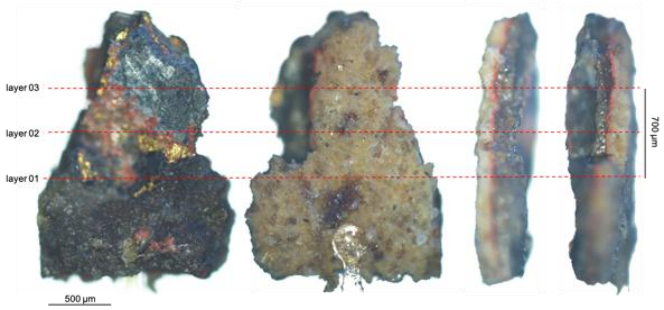

Figure 1: Reference microsample, Montrottier Pièta, Saint Mary Magdalene's dress, "continuous brocade", S2018-157. Left: front view showing the presence of gold traces. Center: Back view showing the preparation layer. The nylon Mitegen micromount is visible at the bottom. Right: side views revealing the sample stratigraphy

In addition, SEM-EDX measurements were also performed on some microsamples. The sample was cut at the level of a chosen layer, imbedded in resin and polished for Scanning electron microscopy (SEM) examination. SEM was carried out

using a Zeiss Ultra+ microscope. Microanalysis (EDX) measurements were performed using a Silicon Drift (SDD) detector. In order to identify the organic materials, Fourier Transform Infrared (FTIR) measurements were also performed.

2.2. XRD/XRF-CT measurements

XRD/XRF-CT measurements were performed at the BM02-D2AM French CRG beamline[26] of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The microsamples were mounted on a Mitegen micromount covered with a thin layer of wax avoiding any invasive intervention and allowing the possibility to perform further complementary analyses, fixed on a goniometer head and aligned at the center of the D2AM 6-circle diffractometer. For XRD measurements, a 2D ImXPAD D5 pixel detector was attached to the diffractometer 2θ arm at a sample-detector distance of 215 mm and a 2θ angle of 3.5°. The geometry and beam wavelength were calibrated from the XRD measurement of a NIST-660c LaB_6 capillary sample using the PyFAI library[27]. The XRF detector (KETEK SDD AXAS-M1) was located at 100 mm from the sample in reflection geometry at an angle of 40° from the direct beam. The energy scale was calibrated using the PyMCA software[28] from the positions of a few well identified fluorescence peaks. The energy of the monochromatic beam was tuned at 20 keV (or 0.61992 Å) with a flux of about 10^{10} photons/s. This energy has been chosen in order to allow good detection efficacy for a large number of elements using XRF, and to access a wide range of reciprocal space for XRD. The beam was focused by a set of Kirkpatrick–Baez mirrors down to a probe size of 25 x 40 μm^2 offering a gauge volume containing a sufficiently large number of crystallites for powder diffraction averaging and allowing to perform in a reasonable time frame a thorough study of the microsamples. With these conditions, each measurement point required a 1s exposure time. In order to improve the resolution of the image, and since we used the smallest available beam size on the instrument, we applied the oversampling method[3] and decreased the translation step size down from 40 µm to 25 µm, leading to a final voxel size of 25 x 25 x 25 μm^3 in the tomographic reconstructions. In order to reduce the time spent for sample rotation, a continuous rotation scan was used with a 2° rotation step. Thus, the tomographic data collection of a microsample slice of 1 mm maximum length, using e.g. 45 translation points each 25 µm (a few points are added on each side out of the sample), required approximately 2.5 hours. Five microsamples from the Montrottier Pièta were investigated, in which several slices were measured to obtain a more precise 3D description, their locations being chosen by visual inspection. For example, the positions of the slices for the microsample S2018-157 are shown in Figure 1. In the following, we will concentrate on the slice #2 of sample 2018-157.

2.3. Pre-treatment of the XRD/XRF-CT data

After collection, the data were treated to account for various effects and organized as data cubes (i.e. 3D matrices) for their subsequent investigation using multivariate analysis. All treatments were made using scripts written in Python 3, and the initial and treated data were stored in HDF5 format. For the XRF data, the recorded spectra (4096 points each) were stored as data cubes without modification. The XRD data required first the transformation from 2D to 1D diffraction patterns which was carried out by radial integration using PyFAI. The 1D patterns contain 3000 points extending from 2θ =0° to 33 ° with a step size of ≈0.011°. During this process, it was realized that many diffraction images showed the presence of strong Bragg spots coming from relatively large isolated grains. Although these grains (identified as mostly gypsum grains) represent a minor volume of the sample, the strong Bragg peaks they produce can greatly reduce the quality of the powder patterns and tomographic reconstructions. Therefore, their contribution was filtered out using the function 'separate' from the PyFAI library. The 1D XRD patterns were then stored as (T, R, D) data cubes. For our microsample S2018-157 which required a 2.05 mm translation scan, a data cube contained 82 x 180 = 14760 pattern of 3000 points each, i.e. 44 280 000 points. The XRD data cubes were submitted to further pre-treatments including:

-limiting the 2θ range;

-automatic background subtraction. In such tomographic experiment, the background level and shape can fluctuate quite strongly during the rotation of complex shaped samples, due to variations in x-ray absorption, shadowing of diffuse x-ray scattering, etc. This can greatly complicate the task of identifying a limited number of components for the multivariate analysis algorithms. Therefore, it is mandatory to remove background beforehand. After several attempts, we adapted a 1D rolling ball algorithm applied sequentially to all patterns, after Savitzky-Golay smoothing. This proved to be efficient and robust, especially to take account of abrupt background changes between adjacent points in the sinogram;

-zeroing the data outside the sample area, which improves the quality of the tomographic reconstruction;

-applying a center-of-mass correction to account for off centering of the sample rotation axis with respect to the x-ray beam, leading to a dissymmetry of the sinogram and artefacts in the reconstructed image.

The sinograms were then transformed to reconstructed images and displayed using the inverse radon transform *iradon* implemented in the scikit-image python library[29] (Figure 2). At this stage, it is possible to examine the sinograms or corresponding reconstructions to look for the location of specific features as phases (for XRD) or chemical elements (XRF).

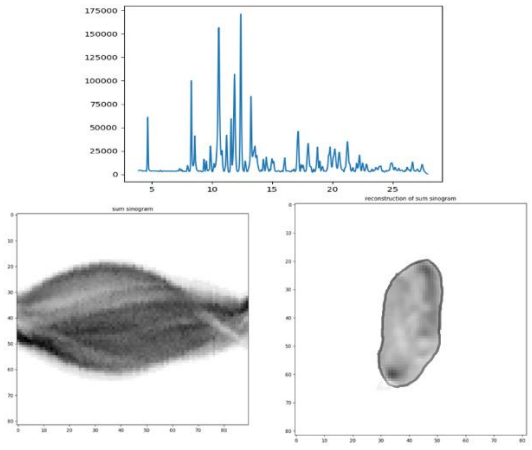

Figure 2: XRD pattern (top), sinogram (bottom left) and tomographic reconstruction (bottom right) for the sum of all XRD-CT data after data reduction. The approximate shape of the sample is shown as a grey line.

2.4. Element / phase identification and Rietveld refinements of summed data

The XRF spectrum summed over all voxels is shown in Figure S2 and was used to identify the chemical elements present in the sample. The spectrum is strongly dominated by heavy elements such as lead and mercury. Gold seems to be present as a shoulder on the left side of the mercury peaks. Iron and calcium are the most prominent lighter elements, with minor ones such as copper, nickel and chromium probably coming from contamination by the experimental setup rather than from the sample. Finally, sulfur, tin and traces of chlorine are detectable by their Lα lines at 3.44 keV (Sn) and Kα lines at 2.31 keV (S) and 2.62 keV (Cl), respectively. The peak at 2.98 keV may be attributed to ArKα from air scattering, though the Lα lines of silver may contribute at the same position. These data being the sum of a layered sample rotated in the beam during experiment, a tentative XRF quantification of the elements would be rather inaccurate. Moreover, the elements detected can participate to several phases located at different places and with different purposes. In order to get further insight into the sample organization, it is necessary to make use of the diffraction data.

The XRD data summed over all voxels of the sinogram were used for a first identification of the crystalline phases with the EVA software[30] and the PDF4+ database[31] (Figure S3). For complex samples such as those investigated here, which are made of several metallic and painted layers deposited on a substrate, the number of phases can be quite large. It may be difficult to identify all the phases, since Bragg peaks from minor constituents can be hidden among the strong peaks from major ones (see Figure S3). Here, seven phases could be identified: cinnabar (HgS), gypsum ($CaSO_4$,$2H_2O$), the two tin oxides cassiterite (SnO_2) and romarchite (SnO), two lead carbonates: cerussite ($PbCO_3$) and hydrocerussite ($Pb_3(CO_2)_2$,$(OH)_2$) and beeswax identified as n-paraffin. The presence of gold, which is expected from visual inspection of the sample and from the XRF data, can be only assessed by a small shoulder on the right side of a peak from cassiterite at 2θ = 15.1°. To ascertain the phase identification, a Rietveld refinement was carried out using the FullProf software[32] allowing to better identify the minor phases (Figure 3). Thanks to quantitative analysis, it also allows obtaining an estimate of the mass fraction of the crystalline phases in the sample. Finally, diffractograms from the individual phases can be extracted for subsequent use in the supervised MCR-ALS procedure. For these refinements, a Pseudo-Voigt profile function was used and the background was determined by linear interpolation between selected points. Only the scale factor, an overall atomic displacement parameter, the cell and profile parameters were refined for each phase, the atomic structure parameters being kept fixed. To compare with the results from multivariate analysis (see below) three additional phases were also included in the refinement, although they were not clearly identified in the search-match analysis: minium (Pb_3O_4), goethite (FeOOH) and chlorargyrite (AgCl) resulting in a total of 11 phases. The latter two phases are also compatible with the observation of iron, silver and chlorine in the XRF data. In decreasing order of weight fraction: beeswax (41%), gypsum

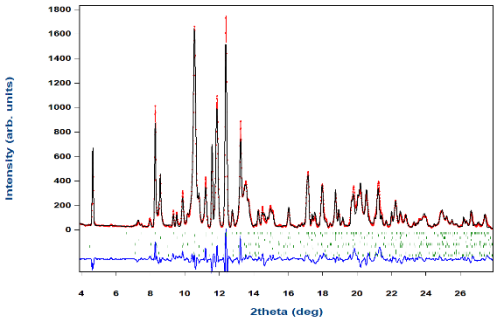

Figure 3: Rietveld refinement of the summed XRD-CT data including 11 phases. Rp=12.4, Rwp=16.0, Rexp=9.78, Chi2=2.69.

(23%), cassiterite (15%), romarchite (8%), cinnabar (8%), goethite (2%), hydrocerussite, cerussite and chlorargyrite (1%), minium and gold (<1%).

In the case of beeswax, for which no crystal structure has been published, we recorded a XRPD pattern on a Bruker D8 Endeavor diffractometer, with CuKα radiation in Bragg-Brentano geometry, for a natural beeswax sample (purchased from C.T.S., Italy) mixed with corundum powder. This allowed us to check that the XRPD pattern of natural beeswax is very close to those observed in our samples as well as to that of n-paraffin in the PDF4+ database[33]. The corundum powder was used to calibrate the beeswax pattern Bragg peak intensities for the determination of its weight fraction by Rietveld analysis of the tomographic data.

2.5. 3D analysis of the XRD-CT data using PCA/MCR-ALS and NMF

The data cube was used for PCA decomposition using the scikit-learn python library[34]. Figure 4a shows that most of the data variance is explained by a small number of components. The plot of the first two components (figure 4b&c) also demonstrates that this type of decomposition is not suitable for phase identification, with strong negative signal in component 2, since the orthogonality constraint used in PCA is not adapted to this kind of physical data. Nevertheless, this can be used to estimate the number *d* of components to be used in a further analysis using e.g. MCR-ALS.

MCR-ALS was applied using the pyMCR python package[35] and non-negative least-squares regressor. Non-negativity constraints were applied to both the ***C*** and ***S*** matrices (i.e. concentrations and components).

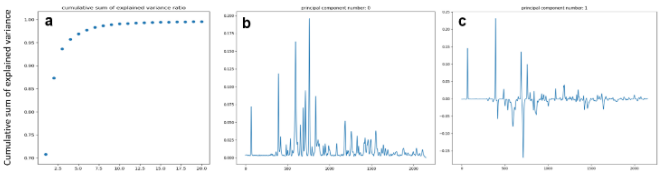

Figure 4: Results of the PCA decomposition: a) cumulative sum of explained variance vs component number; b & c) first and second components.

To provide a starting estimate of the ***S*** matrix, we used the first *d* components found by PCA and change all negative values to a small positive one to respect the non-negativity constraint. Different numbers of components were tested. In principle, *d* cannot exceed the number of phases in the sample, i.e. close to 9 for the present sample if we consider the 11 phases used in the Rietveld refinement of the sum data. However, given the large size of the voxels, the presence of components consisting in a mixture of phases expected to be adjacent in the sample (e.g. the two tin oxides) is highly probable, which will reduce the number of components to be found. Figure S4 shows that the data can be successfully described using either 6 or 9 components.

The individual patterns for the components can be seen in Figures S5 and S6, as well as their concentration maps for the sinogram and real space reconstructions. For the 9 components case, components 1 and 4 consist almost exclusively in gypsum, while components 0 and 8 contain mainly cinnabar, components 6 and 7 are mixtures of cassiterite and romarchite, etc. As suggested above, the process is not able to decompose the data into individual phases. Figure S7 shows a part of the XRD patterns for components 1 and 4. The patterns are very similar and can be indexed as gypsum, but the peaks are obviously shifted between the two patterns. This might be attributed to an apparent change of position of the diffraction center occurring during sample rotation, due to the very anisotropic sample shape, which modifies the distance to the detector. Since the data can be equally well described with only 6 components a choice of a larger number is not justified. In any case, the location of individual phases in the sample requires an additional step consisting in the identification and quantification (e.g. by Rietveld refinement) of the different phases in each component pattern. However, this task is now much easier than the previous one carried on the sum data, since the number of phases in each single pattern is much smaller and less peak overlapping occurs. This also allows identifying additional phases which could have been overlooked previously.

In principle, the MCR-ALS method with these positivity constraints used here should give equivalent results as NMF. In order to check this, we have used the NMF python implementation in the scikit-learn library[29]. Here only the number of components has to be provided. As shown in figure S8 for the case of 6 components, the set of patterns obtained after NMF analysis and their location in the reconstructed slice are similar to those obtained by MCR-ALS, although the order and scaling are different.

This provides an excellent test of the procedure reproducibility. It is also quite interesting to compare the execution times for both methods: using a standard laptop computer under Windows10® (i7 at 2.6 GHz, 16 Go of RAM) the MCR-ALS process took 863 s and the NMF 9.3 s. Weather such a large difference is due to the algorithm or to the programming or implementation efficiency is beyond the scope of this paper, but obviously, the NMF algorithm used here is able to decompose a large data cube in a quite short time, well compatible with the requirements for assessing data quality during a synchrotron experiment.

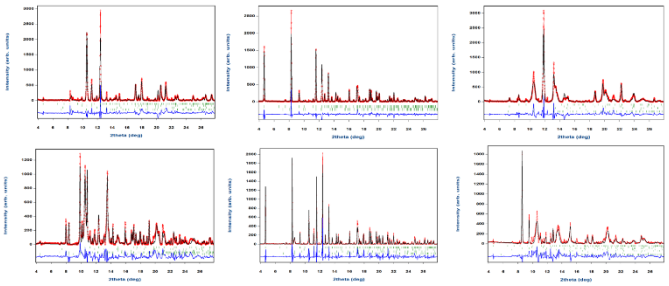

Figure 5: Rietveld refinements of the 6 component patterns extracted by NMF. Component # 0 to 5 from left to right and top to bottom. Chi2=19.5, 6.0, 12.4, 11.3, 11.4 and 14.6, respectively.

The physico-chemical significance of the obtained component patterns found by NMF with 6 components has now to be checked by using them for phase identification and quantification. The phase matching search results can be seen in figure S9, while the Rietveld plot are shown in Figure 5. In figure S9, the main phases found in the sum pattern analysis can be clearly identified in the component patterns. Moreover, minor phases which cannot be directly observed on the sum pattern can also be detected here. For example, on the pattern of component 5, gold can be clearly seen with 3 peaks at 2θ=15.11°, 17.5° and 24.8° and chlorargyrite with 3 peaks at 2θ=11.1°, 12.8° and 18.2°. A small peak at 2θ=5.7° can be tentatively attributed to minium. In components 0, 2 and 4, two peaks at 2θ=8.5° and 14.5° suggest the presence of goethite. Rietveld refinements of the 6 component patterns were carried out to validate their presence and determine the mass fractions of the phases in each component, which are reported in Table 1. From this table, the phase concentration matrix $\boldsymbol{C_p}$ representing the concentration of each phase in each voxel of the sample tomographic reconstruction can be obtained by the $\boldsymbol{C_p}=\boldsymbol{C}.\ {}^{t}\boldsymbol{P}$ matrix multiplication, where ***C*** is the concentration matrix of the components found by NMF and ${}^{t}\boldsymbol{P}$ is the transpose of the ***P*** matrix containing the phase concentrations of each components. The ***P*** matrix, corresponding to Table 1, is of dimensions *p.d*, *p* being the number of phases and *d* the number of components. However, before proceeding to this calculation, the components have to be properly scaled with respect to each other, since no physical scaling is provided by the algorithm.

Table 1: Mass fractions in % of the different phases for each of the 6 components found by NMF after Rietveld refinement.

Component	0	1	2	3	4	5
Gypsum	7(1)	99(2)	-	-	72(10)	0.22(7)
Cinnabar	64 (2)	-	-	8.2(8)	7(1)	0.25(3)
Beeswax	-	-	51(4)	-	7(14)	94(12)
Goethite	26(2)	-	3.0(3)	-	13(3)	-
Cerussite	-	0.48(3)	-	7.3(7)	-	-
Minium	-	-	-	-	-	0.12(2)
Hydrocerussite	-	-	-	23(2)	-	-
Chlorargyrite	-	-	-	-	-	1.0(1)
Gold	-	-	0.11(3)	-	-	0.39(4)
Romarchite	2.6(5)	-	22(1)	23(7)	-	0.30(3)
Cassiterite	-	-	24(1)	39(3)	-	3.8(3)

In order to scale the components, we derived a method similar to the Reference Intensity Ratio (RIR) method which uses the ratio of the strongest peak of a given phase *a* to the strongest peak of a standard ***c*** (usually corundum)[36]. However, using the strongest peak height here proved to be quite inaccurate because of large variations between the peak widths of different phases, and we rather used the scale factors obtained from the Rietveld refinement instead. For each component, the scaling is calculated using the phase ***a*** with the largest weight fraction. The ratio between the scale factors of phase ***a*** and corundum ***c*** for an equimass mixture can be easily computed as:

$$Ra = \frac{Sa}{Sc} = \frac{ATZc \times Vc}{ATZa \times Va}$$

where *ATZ* and *V* are the molecular mass and the volume of the unit cell as computed by the Fullprof program, respectively. The scale factor to be applied to each component ***i*** can then be obtained as:

$$Si = \frac{Sa.0.5}{Wa.Ra}$$

where *Wa* is the mass fraction of phase ***a*** in the component ***i***. The intensity values from the pattern of each component ***i*** is divided by the corresponding *Si* value, while the component concentrations from matrix ***C*** are multiplied by the same values, so that the calculated ***C.S*** matrix containing the reconstructed patterns for all pixels remains unchanged. The correctly scaled ***C*** matrix can then be used to compute the ***Cp*** matrix containing the phase concentrations in all voxels as shown above. The distribution of the 11 phases in the layer are shown in Figure 6. The composition of any voxel or its evolution along line cuts can be obtained to facilitate the interpretation of the results.

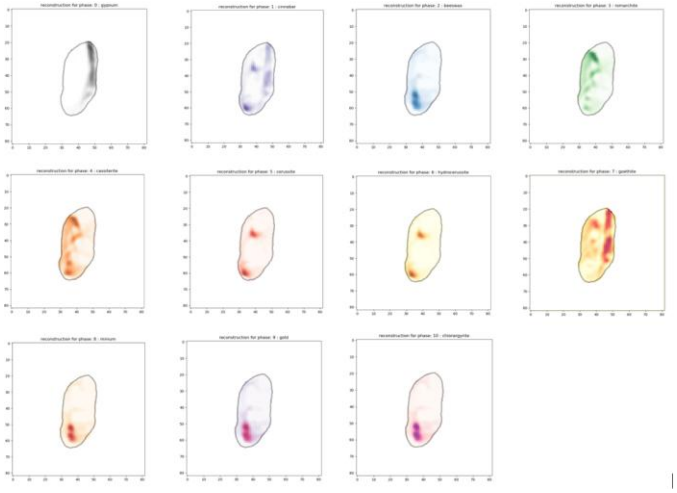

Figure 6: Tomographic reconstruction of the concentrations of the 11 phases quantified by Rietveld analysis of the 6 component patterns obtained by NMF. From left to right and top to bottom: gypsum, cinnabar, beeswax, romarchite, cassiterite, cerussite, hydrocerussite, goethite, minium, gold, chlorargyrite. The approximate sample shape is shown as a grey line. The color hue is scaled to maximum concentration for each plot.

3. Discussion

Supervised MCR-ALS analysis can provide a good check for the validity of the phase concentrations extracted using (unsupervised) NMF. For this analysis, the patterns of the 11 individual phases introduced in the *S* matrix were simulated using the results of the previous Rietveld refinement of the sum pattern, with a 0 intensity background. MCR-ALS was applied using the pyMCR python package[35] and non-negative least-squares regressor. Non-negativity constraints were applied to the ***C*** matrix, while the ***S*** matrix was kept fixed.

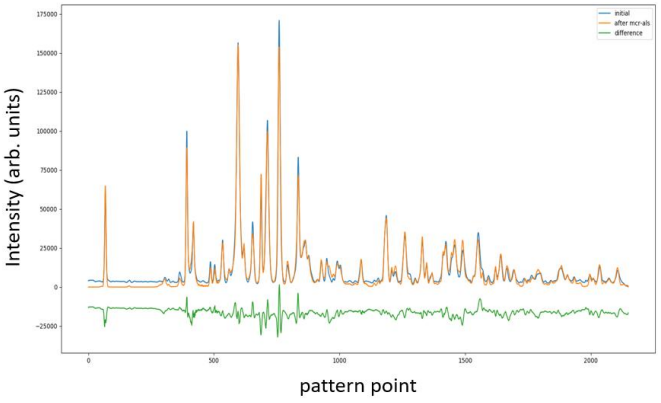

Figure 7: Comparison of the experimental sum data with the reconstructed sum pattern after supervised MCR-ALS decomposition using 11 components from Rietveld refinement. Blue: experimental pattern, orange: calculated, green: difference (shifted by -10% of the maximum).

One can see in Figure 7 that the global agreement on the sum pattern is similar to the one obtained by Rietveld refinement. Moreover, this process directly yields for each voxel of the sample the concentration of the phases which have been introduced in the calculation. A small remaining background is

still present at low angles which cannot be accounted for by the component set for which the background is fixed at 0.

The observed background is the sum of the remaining signal after background subtraction of the 7380 individual patterns, and will be hard to completely eliminate. Since it varies in an erratic way from pattern to pattern, it cannot be accounted for by introducing an additional background component in the calculation. Also, the peak profile from the Rietveld refinement of the sum data may have difficulty to represent small variations of profile or scattering center during the tomographic data collection. Figure S10 shows the distribution of phase concentrations to be compared with those of Figure 6 found by NMF. The distributions of the phases with the strongest concentrations (gypsum, cinnabar, beeswax, romarchite, cassiterite, cerussite and hydrocerussite) are quasi identical for the two methods. Gold, goethite and chlorargyrite are only partly similar, and the distribution of minium for the supervised analysis is somewhat erratic. Given the small concentrations of these phases and the only approximate agreement of the decomposition using the supervised method as seen from Figure 7, it is no surprise that these phases could not be modelized accurately. On the other hand, the very good agreement for the major phases determined independently by the two methods constitute a satisfactory validation test. Table 2 shows the mass fractions of the phases as determined by Rietveld refinement of the sum data and by multivariate analysis using NMF decomposition (6 components) and supervised MCR-ALS. The overall agreement is quite satisfactory given the large differences between the methods. The NMF values show a somewhat poorer agreement with the Rietveld ones, especially due to an overestimation of beeswax, a phase that mainly appears as 2 Bragg peaks in the diffraction patterns and may be difficult to estimate accurately. As phase quantification through NMF requires an additional step (Rietveld fitting of the components) as compared to supervised MCR-ALS, the obtained precision maybe decreased.

Table 2: Mass fractions in % of the different phases determined by Rietveld refinement of the sum data and by multivariate analysis using NMF decomposition (6 components) and supervised MCR-ALS.

Phases	Rietveld (sum data)	NMF (6 components)	MCR-ALS supervised
Gypsum	23.1(1)	18.0	25.6
Cinnabar	7.8(3)	1.3	8.3
Beeswax	40.9(4)	57.5	40.0
Goethite	1.9(1)	1.8	3.4
Cerussite	0.9(7)	0.5	1.1
Minium	<0.01	0.1	0.3
Hydrocerussite	1.3(1)	1.6	1.5
Chlorargyrite	1.0(1)	0.5	0.8
Gold	0.2(2)	0.2	0.3
Romarchite	8.3(4)	7.6	8.8
Cassiterite	14.6(7)	10.9	9.9

Although it can look appealing, using the supervised method seems to be prone to several drawbacks. It first requires to correctly identify all the phases present in the sample, through search-match analysis of the sum pattern and to prepare the individual patterns for all these phases. This necessitates to perform the Rietveld refinement of the sum pattern, since the line profiles depend on the experimental resolution function and on the crystalline quality of each phase. This work can be rather long and the description of minor phase profiles cannot be very accurate. Furthermore, as we have seen above, the diffraction center of a given phase can vary during the tomographic experiment, depending on the phase distribution in the sample. The peak shifts due to these variations will yield a broadening and distortion of the peak profiles of the sum pattern. Finally, the presence of a remaining background in the sum pattern may lead to inaccuracies in the localization of minor phases. This method should thus be restricted to relatively simple and small samples where such drawbacks would be limited.

The phase distribution in the sample chip can now be examined on the basis of the NMF multivariate analysis results. From bottom to top of the chip (the bottom is on the right side in Figure 8a) the first feature consists in a preparation layer made of gypsum, followed by a layer of cinnabar mixed with some gypsum and goethite. One does not observe a strong signal in the center of the sample, which may be due to the presence of a non-diffracting/amorphous phase, mixed locally with cinnabar, lead white and goethite. More to the top, one observes a strong layer of beeswax, mixed with a somewhat disturbed distribution of tin oxides. Gold mixed with silver chloride and maybe minium are located as two strong spots on top of the tin oxide layer. Finally, on top of the gold spot, a mixture of cinnabar and lead white is observed. This stratigraphy can be interpreted in the following way: a preparation layer made of gypsum was covered by a vermillion paint, on which the brocade system was applied. It consisted in two filling layers, one made of an amorphous phase mixed with various pigments, and a second one made of beeswax on which the tin foil and gilding were attached. The tin layer is now fully degraded into a mixture of the two tin oxides. The gilding was made of gold and silver, possibly zwischgold, consisting in two thin sheets of gold and silver hammered together[37]. Finally, the brocade is covered by an overpainting layer containing lead white and cinnabar. Despite their limited spatial resolution, these observations (such as the nature of the preparation layer, the presence and composition of two brocade filling layers and the use of zwischgold) already constitute important markers which can be used to identify fabrication techniques specific to local craftsmen or schools.

In order to compare these results with those which can be obtained by more classic investigation techniques, we sacrificed the sample 2018-157 which was cut at the level of layer 2, for SEM-EDX and FTIR measurements. The results of optical microscopy and SEM are shown in Figure 8, and the results of the analyses are summarized in Table 3. The

observations made with these techniques are in very good agreement with the results of the XRD-CT investigation. The organic phases present in the two brocade filling layers could be identified by FTIR as beeswax (as found by XRD-CT) and an organic layer loaded with lead soaps. For x-ray diffraction, lead soap compounds are known to yield strong Bragg peaks only at d spacing larger than 40 Å[38], leading to too small angles to be detected in our geometrical configuration. This is also seen as a dip in phase concentrations observed on the profile of Figure 8f showing the major phase concentrations obtained by NMF. On this profile taken from a virtual cut at the middle of the sample, the succession of phases and the layers thicknesses correspond well to what is observed on the optical microscopy image. This demonstrates that the NMF method for analysis of XRD-CT data is able to perform quantitative phases reconstruction in very complex systems without the need to destroy the sample.

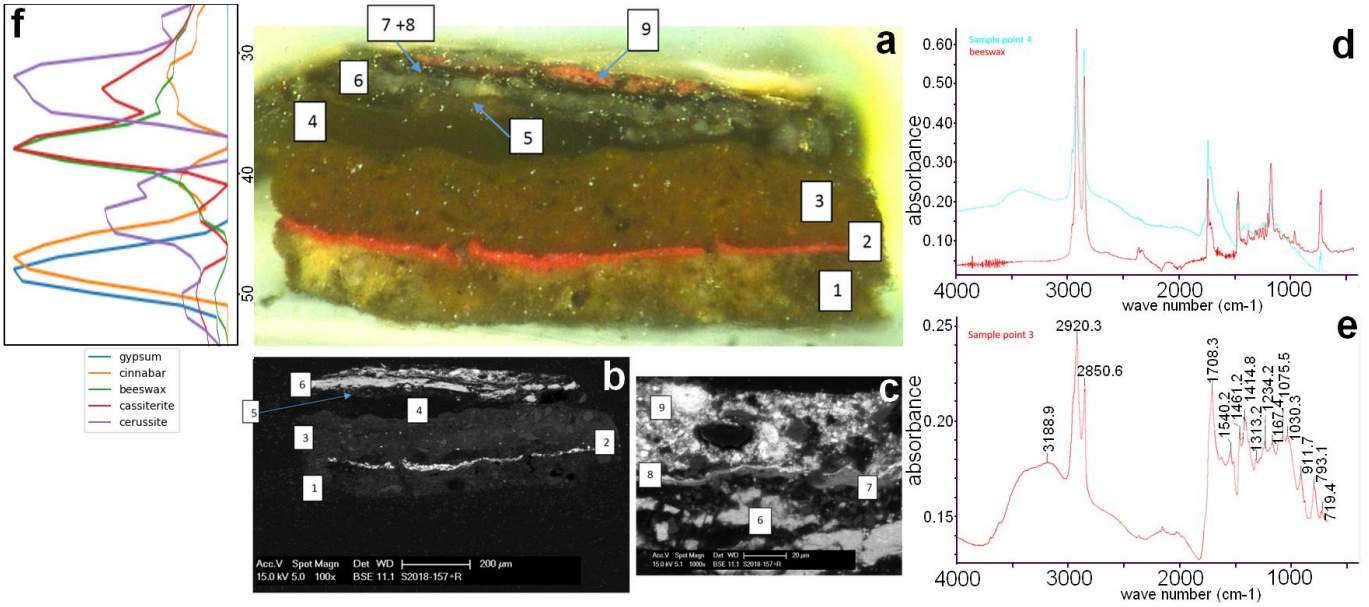

Figure 8: Observation with optical microscopy (a) and SEM (b & c) of the stratigraphy of sample 2018-157 at the level of layer 2. The number refers to analysis points detailed in Table 3. Right FTIR spectra measured at points 3 (e) and 4 (d), compared with a measurement of pure beeswax (in red). In f, a profile taken at the middle of the sample showing the major phase concentrations obtained by NMF (the scale is in pixel, 1 pixel=24 µm). The concentrations are scaled to 1.

Table 3: results from the optical microscopy, SEM and FTIR investigations of sample S2018-157 layer 2.

layer	color	nature	Main chemical elements (from EDX)
9	red	Red overpainting (20 µm) : with vermillion (HgS) and lead white	Pb, Si, Hg, S, Ca
8	golden	Gilding (1-2 µm) made of zwischgold or a gold-silver alloy	Au, Ag, Cl
7	grey	Organic gluing layer between tin and gilding	-
6	grey	Tin layer (≈50 µm)	Sn
5	grey/ brown	Organic layer containing lead. Gluing paste for the tin foil?	Pb, Cl, Si
4	black	First brocade filling layer, pure beeswax as from FTIR	Cl, Ca
3	dark brown	Second brocade filling layer, mixture of lead soap in an inorganic binding (most likely saponified oil, as from FTIR) and various pigments.	Si, Fe, Ca, Al, Pb
2	red	Colored layer made of vermillion mixed with gypsum.	Hg, S
1	white	Preparation layer (75-110 µm), gypsum	Ca, S

4. Conclusion

The results shown above demonstrate the utility of using multivariate analysis methods, and particularly NMF, for the treatment of large data sets from XRF/XRD-CT experiments. Especially concerning XRD, the method allows to obtain in very reasonable computing time and in an automatic process the decomposition of the data cube into a small set of physically meaningful components representing diffraction patterns which can be used for search-match identification of phases and even Rietveld refinements and quantitative analysis. Depending on the sample complexity and experimental resolution (i.e. beam size vs characteristic size of sample inhomogeneity) the extracted components will consist of single or mixed phases. Using the quantitative analysis results for the components, the complete set of phase concentrations can be retrieved. Applied here to XRD-CT, the same method can be used to any kind of XRD data collection leading to large data sets, such as 2D mapping of large areas which is now often performed on artworks such as paintings or frescoes, etc. and can also be applied to XRF data.

Author Contributions

The manuscript was written through contributions of all authors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors wish to thank the team at ARC-Nucléart (Florence Lelong and Fédéric Fabre) for the choice and preparation of the micro samples. They are grateful to Sophie Omère, Heritage Curator at the Direction Régionale des Affaires Culturelles Auvergne-Rhône Alpes and Julien Coppier, member of the Académie Florimontane (Annecy) and owner of the castle of Montrottier (Lovagny, Haute-Savoie, France), for authorizing access to the sculpture. We thank Pierre-Olivier Autran for his help in the center of mass corrections.

This work was financially supported by the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02, Cross Disciplinary Program Patrimalp).

The microfocusing set-up was funded by the French National Research Agency (ANR) under the "Investissements d'avenir" program with the grant number: ANR-11-EQPX-0010.

References

1 P. Bleuet, E. Welcomme, E. Dooryhée, J. Susini, J-L. Hodeau, P. Walter, *Nature Mater,* 2008, 7**,** 468.
2 G. Artioli, T. Cerulli, G. Cruciani, M. C. Dalconi, G. Ferrari, M. Parisatto, A. Rack, R. Tucoulou, *Anal Bioanal Chem*, 2010, **397**, 2131.
3 M. Álvarez-Murga, P. Bleuet, J.-L. Hodeau, *J. Appl. Cryst*., 2012, **45**, 1109.
4 W. De Nolf, F. Vanmeert K. Janssens, *J. Appl. Cryst.*, 2014, **47**, 1107.
5 J. Kieffer, S. Petitdemange, T. Vincent, *J. Synchrotron Rad.,* 2018, 25, 612.
6 E. Dooryhee, M. Anne, M. I. Bardies, J.L. Hodeau, P. Martinetto, S. Rondot, J. Salomon, G. Vaughan, P. Walter, *Appl. Phys. A, Mater. Sci. Process.,* 2005, **81**, 663.
7 F. K Mürer, S. Sanchez, M. Alvarez-Murga, M. Di Michiel, F. Pfeiffer, M. Bech, D. W. Breiby, *Sci. Rep.* 2018, **8**, 10052
8 F. Vanmeert, G. Van der Snickt, K. Janssens, *Angew. Chem.,* 2015, **127**, 3678.
9 S. W. Price, A. Van Loon, K. Keune, A. D. Parsons, C. Murray, A. M. Beale, J. F. W. Mosselmans, *Chem. Commun.,* 2019, **55**, 1931.
10 S.R. Wasserman, *Journal De Physique IV*, 1997, **7**, C2-203.
11 R. Tauler, *Chemom. Intell. Lab. Syst.,* 1995, **30**, 133.
12 R. Tauler A. de Juan, *Data Handling in Science and Technology,* 2015, **29**, 247.
13 P. Conti, S. Zamponi, M. Giorgetti, M. Berrettoni, W.H. Smyrl, *Anal. Chem.,* 2010, **82**, 3629.
14 C. Ruckebusch, L. Blanchet, *Analytica Chimica Acta,* 2013, **765**, 28.
15 W.H. Cassinelli, L. Martins, A.R. Passos, S.H. Pulcinelli, C.V. Santilli, A. Rochet, V. Briois, *Catalysis Today*, 2014**,** **229**, 114.
16 J.C. Burley, D. O'Hare, G.R. Williams, *Anal. Methods*, 2011, **3**, 814.
17 K.W. Chapman, S.H. Lapidus, P.J. Chupas, *J. Appl. Cryst.,* 2015**,** **48**, 1619.
18 M.A. Rodriguez, M. R. Keenan, G. Nagasubramanian, *J. Appl. Cryst.,* 2007, **40**, 1097.
19 A. Taris, M. Grosso, A. Viani, M. Brundu, V. Guida, *Chem. Eng. Trans.*, 2015, **43**, 895.
20 C.J. Long, D. Bunker, X. Li, V.L. Karen, I. Takeuchi, *Rev. Sci. Instrum.,* 2009, **80**, 103902.
21 V. Stanev, V.V. Vesselinov, A. Gilad Kusne, G. Antoszewski, I. Takeuchi, B.S. Alexandrov, *npj Computational Materials* 2018, **4**, 43.
22 F. Lelong, E. Pouyet, S. Champdavoine, T. Guiblain, P. Martinetto, P. Walter, H. Rousselière, M. Cotte, *CerOArt,* 2021. https://doi.org/10.4000/ceroart.7802
23 F. Lelong, S. Champdavoine, E. Pouyet, T. Guiblain, *La rubrique des patrimoines de Savoie,* 2015, **35**, 24.
24 F. Lelong, S. Champdavoine, M. Lefèvre, S. Peurichard, E. Pouyet, C. Terpent, *La rubrique des patrimoines de Savoie,* 2015, **36**, 8.
25 P. Martinetto, N. Blanc, P. Bordet, S. Champdavoine, F. Fabre, T. Guiblain, J.-L. Hodeau, F. Lelong, O. Leynaud, A. Prat, E. Pouyet, E. Uher, P. Walter, *J. Cult. Herit.* 2021, **47**, 89.
26 G.A. Chahine, N. Blanc, S. Arnaud, F. De Geuser, R. Guinebretière, N. Boudet, *Metals*, 2019, ***9***, 352.
27 G. Ashiotis, A. Deschildre, Z. Nawaz, J. P. Wright, D. Karkoulis, F. E. Picca, J. Kieffer, *J. Appl. Cryst*., 2015, **48**, 510.
28 V. A. Solé, E. Papillon, M. Cotte, P. Walter, J. Susini, Spectrochimica *Acta B,* 2007, **62**, 63.
29 S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, *PeerJ* 2014**,** **2**:e453.
30 BRUKER-AXS, software DIFFRAC.EVA V5.0, 2019
31 S. Gates-Rector, T. Blanton, *Powder Diffr.*, 2019, **34**, 352.
32 J. Rodriguez-Carvajal, *Physica B*., 1993, **192**, 55.
33 R.D. Heyding, K.E. Russell, T.L. Varty, D. St-Cyr, *Powder Diffraction*, 1990, **5**, 93.
34 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *JMLR* 2011, **12**, 2825.
35 C. H. Camp Jr, *J. Res. Nat. Inst. Stan.* 2019, **124**, 124.018
36 I. C. Madsen, N. V. Y. Scarlett, in *Powder Diffraction: Theory and Practice*, Editors: R. E. Dinnebier, S. J. L. Billinge, The Royal Society of Chemistry, Chapter 11, 2008, 298.
37 M. Eveno, E. Ravaud, T. Calligaro, L. Pichon, E. Laval, Herit. Sci., 2014, **2**, 17.
38 M.-C. Corbeil, L. Robinet, *Powder Diffr.*, 2002, **17**, 52.

