

CrossMark  
click for updatesCite this: *J. Anal. At. Spectrom.*, 2015, **30**, 777

# Strategies for processing mega-pixel X-ray fluorescence hyperspectral data: a case study on a version of Caravaggio's painting *Supper at Emmaus*

Matthias Alfeld<sup>\*ab</sup> and Koen Janssens<sup>a</sup>

Technical progress in the fields of X-ray sources, optics and detectors is constantly enhancing the pace of data acquisition in XRF imaging. This enlarges the size of the hyperspectral datasets and the number of their sub-parts. This paper describes the challenges in processing large XRF datasets featuring several million pixels/spectra and the strategies developed to overcome them. During the investigation of historical paintings by scanning macro-XRF the main challenges are the correct identification of all spectral features in a dataset and its timely processing. For the identification of spectral features different approaches are discussed, *i.e.* the use of sum spectra, maximum pixel spectra and of  $\chi_r^2$  maps. For the time-efficient, artefact-free evaluation of XRF imaging data, different software packages are evaluated and intercompared (AXIL, PyMCA, GeoPIXE and the in-house written datamuncher). The process of data evaluation is illustrated on a large dataset (3.4 MPixels) acquired during the investigation of a version of Caravaggio's *Supper at Emmaus* (143 × 199.5 cm<sup>2</sup>). This 17th century painting is currently the largest object entirely scanned with macroscopic XRF.

Received 31st October 2014  
Accepted 19th January 2015

DOI: 10.1039/c4ja00387j

www.rsc.org/jaas

## 1 Introduction

Technical progress in the fields of X-ray sources, optics and detectors has made the acquisition of elemental distribution images with dwell times of a few milliseconds per pixel commonplace.<sup>1,2</sup> Recent development in X-ray fluorescence scanning<sup>3–5</sup> and full-field<sup>6</sup> imaging instrumentation is accelerating this pace further, reporting XRF datasets with several tens of million pixels.<sup>3</sup>

The hyperspectral XRF datasets acquired can encompass several millions of pixels, each featuring a complete spectrum of several thousand channels. The evaluation of such datasets faces two main challenges. First a timely evaluation of the data is desirable, as it allows considering the results of previous experiments when planning the following and second a model needs to be defined that accounts for all spectral features present.

The authors faced these challenges first hand during their work on the investigation of historical paintings by scanning macro-XRF (MA-XRF) from 2008 to 2013. In these experiments large areas (up to several square meters) of historical paintings are scanned with a focused or collimated X-ray beam, achieving a lateral resolution of typically a few hundred micrometers. Even a small area of 10 × 10 cm<sup>2</sup> scanned with 100 μm step size

will produce a dataset featuring 1 million energy dispersive XRF spectra with typically 2048 or 4096 channels. After processing, each of the resulting 10–15 elemental distribution images comprises 1 million pixels.

On the one hand, historical paintings are favourable objects for investigation with XRF, as the main elements of the pigments used are present at a concentration level of several mass percent. This allows us to acquire low-noise elemental distribution images with dwell times of a fraction of a second, even when employing low power X-ray tube sources of only a few tens of Watts power.

On the other hand, historical paintings are challenging objects to be investigated with XRF, as the distribution and elemental composition of pigments is often highly heterogeneous. Also the scattering characteristics of the material, which contribute strongly to the shape of the spectral background, particularly in the case of polychromatic excitation, may vary strongly between areas of the painting. Furthermore, many paintings are too large to be investigated in a single scan, in which case several sub-areas need to be scanned separately, constituting sub-parts of the dataset. The elemental distribution images obtained from these subsets need to be recombined in order to obtain images of the entire scanned surface. The presence of several sub-parts further complicates the interpretation of the XRF data.

The intensity of the recorded X-ray fluorescence radiation and the absolute abundance of elements in a sample are correlated; thus XRF imaging allows in principle for the acquisition of quantitative elemental distribution images. However,

<sup>a</sup>Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium. E-mail: koen.janssens@uantwerpen.be

<sup>b</sup>Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany. E-mail: matthias.alfeld@desy.de; Fax: +49 40 8998 4475; Tel: +49 40 8998 4177

this requires a good characterization of the instrument as well as knowledge on the sample's composition and stratigraphy. Historical paintings are complex samples, often with an unknown stratigraphy, so that their investigation by XRF imaging is in general focused on the detection of relative, but not absolute differences in concentration. For this reason the following discussion will be limited to qualitative imaging.

The first experiments on historical paintings attributed to Vincent van Gogh, Rembrandt van Rijn and other artists were done at synchrotron radiation sources.<sup>7–10</sup> Synchrotron experiments are in general done in short periods of time (a few days) with often months in between. So, while sufficient time for the (final) data processing and optimization of the fitting model is available, near real-time data analysis, possibly preliminary, is needed to make efficient use of the time available for experiments. The latter is supported by the availability of high performance computing resources.

Mobile instruments, which were developed in the last years, allow investigation of historical paintings *in situ*, *i.e.* in a museum or gallery.<sup>9,11,12</sup> As these scanners are nearly constantly available, a large stream of raw spectral data is produced. This stream of data needs to be processed in a time efficient manner, *i.e.* fast, compared to the measurement time, and requiring only a minimum amount of the evaluating person's attention. The computing resources available during an *in situ* experiment for data processing are typically limited.

It is obvious that a software package for data evaluation is in this case expected to provide (a) (largely) artefact-free elemental distribution images, while it should also (b) allow for fast processing of huge datasets, preferably with the computing hardware present at the site of the experiment.

The iterative improvement of the spectral fit model requires the repeated processing of several or all sub-parts of the dataset. A suitable software package should (c) allow for the scheduling of data processing batch jobs of sub-parts to avoid time loss between the end of one fitting process and the (otherwise manual) start of the next.

Several software packages were employed for the evaluation of the data acquired. Initially AXIL<sup>13–15</sup> was used, followed by PyMCA<sup>16</sup> and finally GeoPIXE.<sup>17–23</sup> While all software packages allowed for the reliable processing of the data, none was fully satisfying the criteria above, which motivated the in-house development: datamuncher. A fifth software package, not discussed here as the authors have no experience with it, is MAPS, developed at Argonne National Laboratory by Stefan Vogt.<sup>24</sup> It is freely available as compiled IDL code, featuring a GUI.<sup>†</sup> Further, a number of routines based on MAPS are available as open source command line tools.<sup>‡</sup>

In the following, after a brief recapitulation of the mathematical background, the identification of relevant spectral features in a large MA-XRF dataset with several sub-parts will be discussed. The different software packages available will be presented and the motivation for the development of

datamuncher highlighted. The features of interest of the software packages will be compared in the last section. The dataset used for this paper is derived from a painting called *Supper at Emmaus*, a 17th C. version of the painting by Caravaggio, part of a private collection.

## 2 Experimental

### 2.1 The painting

A 17th century version of Caravaggio's *Supper at Emmaus* from a private American collection is pictured in Fig. 1. An autograph version of this composition documented as being painted by Caravaggio is in the National Gallery, London. The painting depicts a scene described in the Gospel of Luke (24 : 13–35). After his resurrection Jesus meets two of his disciples on the way to the town of Emmaus. The disciples, who are at first oblivious of his identity, recognize him at the moment he breaks and blesses the bread. The disciples sit on the left and right of Jesus, who is in the act of blessing. The standing man in the background is referred to as 'the Innkeeper'.

The painting was investigated to gain insight into its creation process, which is expected to provide new arguments for (or against) Caravaggio's authorship. As this paper is focussed on the technical aspects of data processing, the art-historical interpretation of the results will be subject of future publications.

### 2.2 Scanning macro-XRF investigation

*Supper at Emmaus* was investigated with an in-house built XRF scanner of the University of Antwerp, labelled Instrument D, described in detail elsewhere.<sup>9</sup> The scanner consists of a measurement head, which is mounted on a 60 × 60 cm<sup>2</sup> (hor × ver) sample stage (Newport Corporation, Irvine, CA, USA). The measurement head consists of four Vortex EX SD-detectors (formerly: SII, Northridge, CA, USA, now: Hitachi High-Technologies Science America, Inc., Northridge, CA, USA), positioned closely around the X-ray tube to record fluorescence radiation from a large solid angle, thus enhancing the sensitivity of the instrument. The analogue signals of the SD-detectors are processed with a DXP-XMAP (XIA LLC, Hayward CA, USA). The primary radiation emitted by a 10 W Rh-anode "Magnum" X-ray tube (Moxtek, UT, USA) is collimated with a lead pinhole.

For the measurements, the painting was fixed on a height adjustable easel and repositioned for each scan with respect to the scanner, which was installed on a table. The entire surface of the painting was scanned in 12 subsections of approx. 55 × 55 cm<sup>2</sup> with a step size of 1 mm and an average dwell time of 0.42 s per pixel, resulting in a total scanning time of 17 days for the entire painting. The 12 subsections featured together 3.4 million pixels with sufficient overlap between scans to allow for good quality stitching of the elemental distribution images into the final 2.9 MPixel images. For data processing the spectra acquired by the four individual detectors were scaled to the same energy calibration and consequently summed. The spectral data were saved for evaluation in the ESRF Data Format

<sup>†</sup> [http://www.aps.anl.gov/Xray\\_Science\\_Division/Xray\\_Microscopy\\_and\\_Imaging/Software\\_and\\_Tools/maps.html](http://www.aps.anl.gov/Xray_Science_Division/Xray_Microscopy_and_Imaging/Software_and_Tools/maps.html)

<sup>‡</sup> <http://code.google.com/p/mapsy/>

(EDF) for all software packages discussed with the exception of AXIL, for which its native SPE format (ASCII text files) was used. 0.7  $\mu$ s were employed as the peak shaping time, so that the spectra were acquired with an energy resolution of approx. 200 eV. For all images shown, the software package used to generate them is indicated in the caption.

While for the overview scans a collimator of 0.8 mm diameter was used, yielding a beam size of approx. 1.2 mm, additional details were scanned with a 0.5 mm aperture, which provided a beam of approx. 0.7 mm diameter. The detail shown in Fig. 3 was acquired with a dwell time of 0.45 s and 0.5 mm step size in 40 hours.

The average output count rate per detector was typically 30 000 counts per second. Thus a summed spectrum of all four detectors per pixel featured on average 50 000 counts. A summed spectrum of all pixels of a sub area is shown in Fig. 7. The rather long dwell time (0.42 s) was chosen in order to clearly visualize also less abundant/less well detected elements such as K and Mn. To visualize only the distribution of elements of high abundance for which the scanner features a high sensitivity, such as Fe, Cu, Pb and Hg, dwell times below 10 ms would have been sufficient.<sup>12</sup>

In the colour scale of all images shown a higher brightness indicates a stronger signal. For all images the minimum and maximum intensity values shown were selected to enhance the readability and highlight the features that are discussed. The final contrast was adjusted by means of gamma-correction of the grey value scale.

## 3 Results

### 3.1 Elemental distribution images of the entire image

*Supper at Emmaus* features a classical 16–17th century palette, as shown in the elemental distribution images of Fig. 1 and 2. Fe and Mn are present in the reddish and brownish earth pigments, used throughout the painting. In both elemental distribution images damaged areas of the canvas are visible, which were filled during restoration works. These are mainly gaps in the background of the painting, but also a cut through the head of the left disciple is discernible, visible in the Mn distribution image. In the bottom part of the painting, slightly above the edge of the table cloth more restored areas are discernible, which result from the presence of a joint in the canvas.



Fig. 1 *Supper at Emmaus* (143 × 199.5 cm<sup>2</sup>, oil and tempera on canvas), private collection. Photograph and elemental distribution images of Mn, Fe and Hg. The red and green shapes indicate the areas discussed in detail in Fig. 3 and 8, respectively. Data processed with the in-house written software (datamuncher).



Fig. 2 *Supper at Emmaus* (143 × 199.5 cm<sup>2</sup>, oil and tempera on canvas), private collection. Elemental distribution images of Pb and Cu. Data processed with the in-house written software (datamuncher).

Hg is present in the red pigment vermilion (HgS), which was used either in the pure form to model the red cloth, or in mixtures with other pigments to model skin tones and to accentuate brown tones, such as on the chair of the left disciple, the bread and the fowl on the table. It is present throughout the painting, probably in the ground layer to give it a warmer touch. Pb is present in lead white (basic PbCO<sub>3</sub>) and to a minor amount in lead tin yellow. It is present in the ground layers and used throughout the painting mixed with other pigments, but also in the pure form, *e.g.* in the table cloth. Also in this distribution image the damaged areas mentioned above are well discernible. Cu is present in green and blue pigments, which can be either natural (malachite (Cu<sub>2</sub>CO<sub>3</sub>(OH)<sub>2</sub>) and azurite (Cu<sub>3</sub>(CO<sub>3</sub>)<sub>2</sub>(OH)<sub>2</sub>)) or their synthetic equivalents (green and blue verditer). Beyond that, Cu is present in the green pigment Verdigris, which consists of chemically ill-defined corrosion products of Cu.<sup>25</sup> A distinction between these pigments by XRF alone is not possible. Cu pigments are used to model the blue clothing of the right disciple as well as the bottles and dishes on the table and the fruit. Further, it is used throughout the painting in mixture with other pigments to model the shadows, especially visible in the face of the innkeeper.

### 3.2 Elemental distribution images in detail

Fig. 3 shows elemental distribution images for the sub-area indicated by the red shape in Fig. 1. This sub-area was chosen as it contains a wide range of elements detectable by XRF, as well as it allows discussion of the spectral features without getting lost in the details of the painting. The area depicts the sleeve of the left disciple and an elaborately decorated ceramic blue jug.

K is assumed to be present in green earth, used to model the cloth of the disciple, and the blue pigment smalt (see below). Ca is present in bone black and in chalk and/or gypsum, which can be found in the ground layers and was used during conservation treatments to fill holes and gaps in the paint. Cr is not a main constituent of pigments made before the 19th century and is present in repaired parts. Earth pigments (containing Fe and Mn) were used throughout the area, with the darker variant used to model shadows. Cu pigments were not used to model shadows in this area. Instead they were used in the bluish jug and bottle. Vermilion (HgS) was used in combination with earth pigments to model the loaf of bread and the glass with amber coloured content. Lead white is used in combination with other pigments throughout this area. It was used undiluted in the table cloth and to set highlights in reflections on the glass vessels. Both Hg and Pb are present throughout the painting, but to enhance the readability of the images the lowest recorded signal is displayed as black. As Hg is present in the ground layer of the painting, the intensity recorded of this element reflects the absorption power of covering paint layers. Thus the elemental distribution images of Hg and the strongly absorbing Pb are notably anti-correlated. K, Co, Ni, As and Bi are all co-localized in the thick paint layers used to model the jug and the bottle. This is due to the use of the pigment smalt, a ground K-rich glass coloured blue by Co. Smalt contains, depending on the raw material used and the manufacturing process employed, varying amounts of minor components, such as Fe, As, Ni and Bi.<sup>26</sup>

Bi was found present only in the smalt containing areas of the painting at a rather low concentration level. The intensity of the recorded Bi-L fluorescence radiation in Bi hotspots is 1/30th of that of the Pb-L fluorescence. Pb and Bi are neighbours in the periodic table, and so the intrinsic sensitivity of the scanner for them is essentially equal. Due to the modest energy resolution (200 eV) and shifts in peak ratios due to absorption effects (see Section 4.3.5), Pb-L and Bi-L lines cannot be completely resolved here. Hence, the Bi-L distribution image contains visible contributions of the Pb-L radiation. The contribution of the lesser intense Bi-L radiation to the Pb-L distribution image is negligible.

The authors experienced that As elemental distribution images of many paintings contained artefacts, due to significant overlaps of the As-K<sub>α</sub> line (10.53 keV) with the Pb-L<sub>α</sub> line (10.55 keV) and the As-K<sub>β</sub> line (11.72 keV) with the Hg-L<sub>β</sub> line (11.82 keV). If only one of Pb or Hg co-locates with As, its intensity can be correctly determined by the interference-free As line. However, when both Pb and Hg contribute to the same spectrum, the As signal is extremely difficult to quantify and often overestimated. In this painting As is overestimated in the bright red clothes of Jesus and the innkeeper (not displayed).

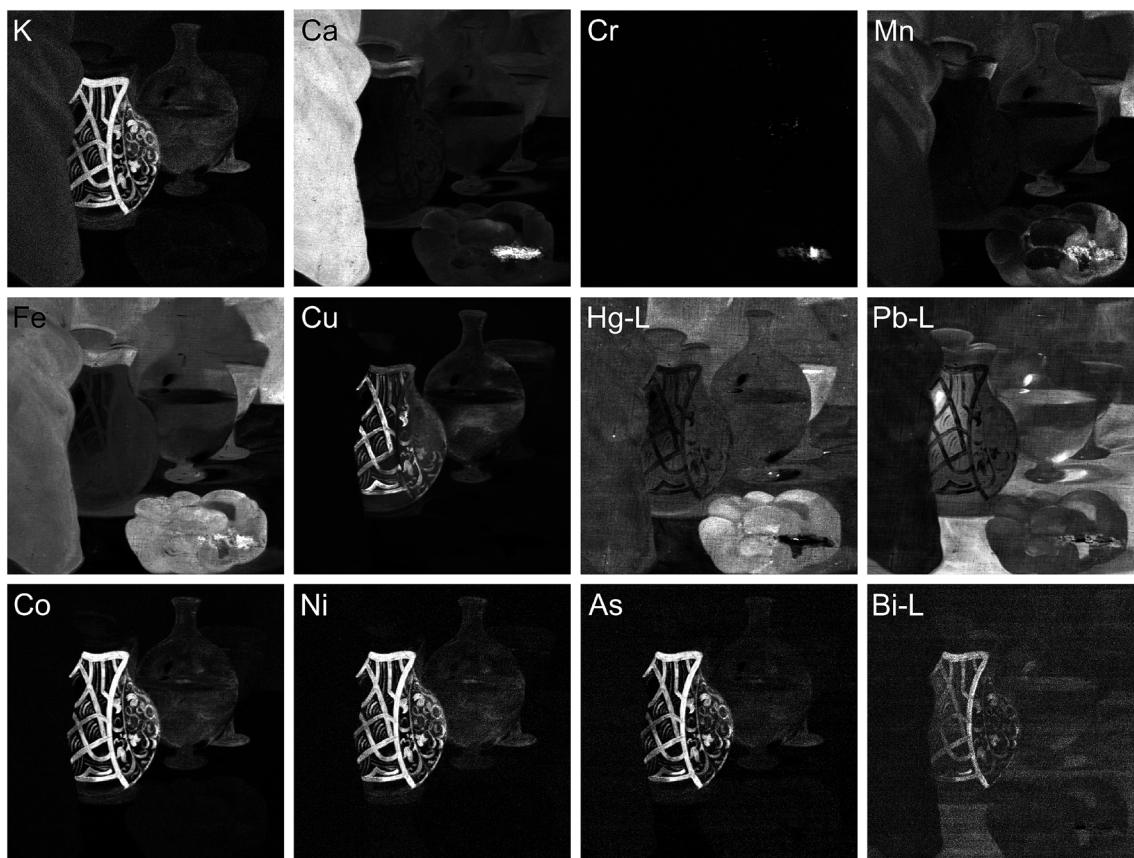


Fig. 3 Elemental distribution images of the area indicated in Fig. 1 with the red square. Results were acquired with a step size of 0.5 mm and a dwell time of 0.45 s. Results obtained with linear least squares fitting (L-LS) with PyMCA.

## 4 Discussion

### 4.1 Mathematical background

For spectral deconvolution the XRF spectrum  $S$  is described by a function  $f$  (see eqn (1)).  $f$  is a linear combination of the spectral background  $B$  and peak profiles  $y_e$  of the elements  $e$  included in the fit. While  $y_e$  depend on the non-linear parameters  $p$ , e.g. energy calibration, energy resolution of the detector and peak shape parameters, the linear factors  $a$  represent the intensity of the recorded fluorescence lines and are the main result of fitting an XRF spectrum for qualitative imaging.

$$f = B + \sum_{e=0}^e a_e y_e(p_0, p_1, \dots) \quad (1)$$

Region Of Interest (ROI) integration allows quick estimation of  $a$  and leads to fast visualization of elemental distributions. ROI imaging can correctly estimate the intensity of a peak  $e$  by integrating the spectrum  $S$  over the selected channels from ROI

(0) to ROI (1)  $\left( a_e \approx \sum_{i=ROI(0)}^{i=ROI(1)} S_i \right)$ , if eqn (2) holds true. This is the

case, if the intensity of peak  $e$  is considerably larger than the sum of all interfering lines  $k$  and the spectral background  $B$ . However, if the fluctuation of the spectral background  $B$  throughout the scanned area is less than that of the element  $e$ , it

has no significant influence on the qualitative ROI image of  $e$ . In this case, a constant offset is added to all pixels of the ROI image, which can be removed by adjusting its grey scale.

$$\sum_{i=ROI(0)}^{i=ROI(1)} a_e y_{i,e} \gg \sum_{i=ROI(0)}^{i=ROI(1)} B_i + \sum_{i=ROI(0)}^{i=ROI(1)} \sum_{k=0, k \neq e}^{k=ROI(1)} a_k y_{i,k} \quad (2)$$

If eqn (2) does not hold true the elemental distribution images contain significant contributions from interfering lines. This is shown in Fig. 4. In the leftmost image the influence of the Fe- $K_\beta$  line on the Co- $K_\alpha$  elemental distribution image is shown. Likewise, the As distribution images are either a combination of the signals of As- $K_\alpha$  and Pb- $L_\alpha$ , or of As- $K_\beta$  and Hg- $L_\beta$  due to the interference discussed in the previous section.

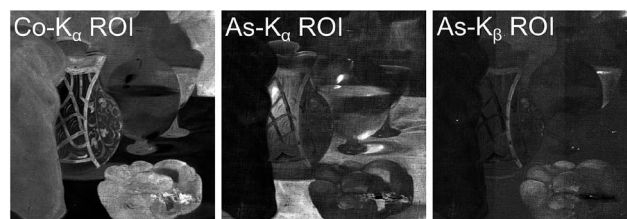


Fig. 4 Region Of Interest (ROI) images of Co- $K_\alpha$  (6.92 keV), As- $K_\alpha$  (10.53 keV) and As- $K_\beta$  (11.72 keV) with a ROI width of 100 eV. Images obtained by direct integration of the raw spectral data.

The As-K $\beta$  ROI image contains also a broad vertical band, which results from contributions of the primary radiation scattered by the wooden stretcher present behind the canvas.

However, eqn (2) holds true for many elements, e.g. Fe-K $\alpha$  and Pb-L $\beta$  in the exemplary dataset. If the composition of the investigated object is well known and when it is thus unlikely to miss an element or mis-interpret a spectral interference, also ROI imaging can provide imaging data of acceptable quality.<sup>27</sup> It is sometimes possible to largely correct for inter-element interference by subtracting ROI images, e.g. correct the As-K $\alpha$  ROI image for the interference from Pb-L $\alpha$  by subtracting the Pb-L $\beta$  ROI image multiplied with a scaling factor. However, these operations require, in general, a high degree of user intervention.

For the investigation of historical paintings ROI imaging was found not suitable to deduce the final elemental distribution images. However, it proved to be a valuable tool for the interactive inspection of datasets to determine the fit parameters.

In order to determine  $a$  and  $p$  the value of the reduced weighted squared difference  $\chi_r^2$ , summed over a user-defined range of channels  $i$ , is minimized (see eqn (3)).  $n - m$  are the degrees of freedom, with  $n$  the number of channels and  $m$  the number of parameters. This normalisation is not necessary for determining  $a$ , but to yield an objective criterion for the quality of the fit. A detailed discussion of the physical phenomena and mathematical operations involved can be found elsewhere.<sup>28</sup>

$$\chi_r^2 = \frac{1}{n - m} \sum_{i=0}^n w_i (S_i - f_i(a_0, a_1, \dots, p_0, p_1, \dots))^2 \quad (3)$$

Commonly, as weights  $w_i$  the inverse variance is used, which is in general approximated as  $w_i = 1/S_i$ . Yet, it has been argued that this approach induces errors in the case of low counting statistics and that  $w_i = 1/f_i$  is a more suitable approach.<sup>29</sup> Weighting the data enhances the quality of the fit for peaks of lower intensity. When not using weights, i.e.  $w_i = 1$ , the quality of the fit is mainly determined by the most intense peaks. In this case, a small relative, but large absolute deviation of the experimental data and the fit function in the area of an intense peak has more weight than a larger relative, but smaller absolute deviation in the area of a lesser intense peak or the background.

Non-Linear Least Squares fitting (NL-LS) allows minimization of  $\chi_r^2$  and determination of  $a$  and  $p$ . However, it presents two limitations during the processing of low statistics datasets with a large number of spectra. First, the individual spectrum typically features insufficient statistics to improve the non-linear parameters  $p$ . Second, NL-LS is an iterative process that requires considerable computing resources and thus results in a relatively slow data processing pace. The first limitation can be dealt with by starting the fit of each spectrum with the same set of non-linear parameters and thus preventing a drift of parameters during the processing of a dataset. Given that  $p$  will be close to the optimal values the main improvement will be in the linear factors  $a$ , which is desired.

To facilitate a faster processing of the dataset, the non-linear parameters ( $p$ ) can be determined by fitting of a spectrum that is representative for the entire dataset and kept fixed during processing of the dataset by Linear Least Squares fitting (L-LS).<sup>16</sup>

During NL-LS the parameters  $p$  are in general constraints to physically meaningful values. It is also possible to constrain  $a$  to positive values in NL-LS and L-LS. This has the advantage that unphysical results are avoided. However, if an appropriate model is defined, the fit of an XRF spectrum will not produce significant negative values for  $a$ . Thus, areas in elemental distribution images featuring significant negative values are indicators of spectral features not included in the fit model. In the case of elements contributing weakly to a small number of pixels in the dataset, such a constraint results in a positive bias on the average calculated intensity. And, finally, L-LS with constraints is more demanding in terms of programming effort during its implementation and computing resources during the data processing.

Unrestricted L-LS can be realised by means of a matrix multiplication, which is a fast operation in computing systems. This is the concept of Dynamic Analysis (DA), developed by C. Ryan in the 1990s.<sup>17</sup> Here a vector  $A$ , comprised of all linear factors  $a$ , is obtained by matrix multiplication of the DA matrix  $T$  with the full spectrum  $S$ . As an example, selected elemental profiles (Fe, Co, As and Pb-L) of the DA matrix  $T$  used for fitting the data of *Supper at Emmaus* are shown in Fig. 5.

$$A = TS \quad (4)$$

While mathematically sound, the working of dynamic analysis is counter-intuitive and physically unsound, as the DA matrix contains significant negative parts. During the operation in eqn (4) the analysed spectrum, which has no negative

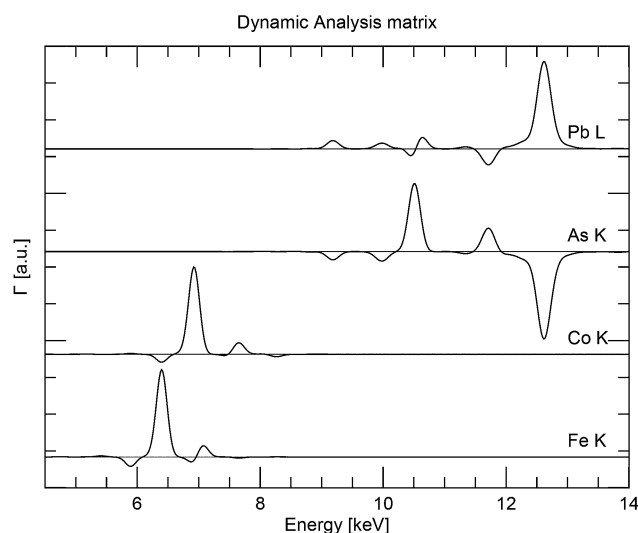


Fig. 5 Excerpt of the DA Matrix used to process the dataset discussed below, featuring selected elemental profiles. The background shape was not considered in the calculation of these profiles for enhanced readability.

channel values, is channel-wise multiplied with each profile in the DA matrix. The product of the spectrum and profile is summed to obtain the net intensity of the group of fluorescence lines in question. Ranges where the profile is zero do not contribute to the calculated intensity, while ranges with a positive value add to the net intensity calculated and those with a negative value reduce it. The Fe and Co profiles in Fig. 5 feature negative ranges to take the interference with neighbouring elements into account (Mn and Co in the case of Fe and Fe and Ni in the case of Co). The As DA profile is strongly influenced by the interference between the As-K $\alpha$  (10.53 keV) and Pb-L $\alpha$  (10.55 keV) lines. While the As-K $\alpha$  and As-K $\beta$  (11.77 keV) lines give rise to positive ranges in the As profile, all Pb L-lines (with the exception of Pb-L $\alpha$ ) are negative. Thus, when a spectrum containing only Pb-L signals is multiplied with the As-K profile, the positive and negative energy ranges cancel out, so that the contribution to the elemental distribution image of As is zero. Likewise the Pb-L profile takes the interference with As into account by weighting the interference free Pb-L $\beta$  line stronger than the Pb-L $\alpha$  line and having a negative energy range at the location of As-K $\beta$ . It is obvious, that DA, as presented here, assumes fixed line ratios in the elemental profiles.

One particularly attractive feature of the DA analysis is that it is equally applicable to the analysis of per-pixel spectra or individual photons. As such, DA is a very suitable tool for real-time data analysis. The formation of the DA matrix is described in detail by Ryan *et al.*<sup>18</sup>

In eqn (1) the background  $B$  is not defined. It can be approximated in three ways during the processing of imaging data. (a) The background shape of the sum spectrum can be used throughout the scan, (b)  $B$  can be estimated by digital filtering of the spectrum and subtracted before deconvolution and (c) it can be modelled during the fit to the individual pixel spectra.

Method (a) is suitable for homogeneous and/or weakly scattering samples, and so is mainly suitable for paintings with thin paint layers on a homogeneous support, *e.g.* a panel painting. A filter often used for the background estimation (b) is Statistics-sensitive Non-linear Iterative Peak-clipping (SNIP).<sup>30</sup> The filter starts with the measured spectrum and erodes the peaks on the background by setting each channel to the average of two neighbouring channels, if the latter is less than the channel's value:  $B_i = \min[B_i, B_{(i+n)}/2 + B_{(i-n)}/2]$  with the initial value of  $B = S$ . Typically  $n$  is set initially to a value of twice the Full Width at Half Maximum of the peaks eroded and reduced during later iterations. With a suitable value for  $n$  the peaks can be eroded and the background shape estimated in less than 25 iterations.<sup>28</sup>

Method (c) uses mathematical functions to describe the background, whose parameters are determined during the fit of the individual pixel spectra. Often polynomials of degree  $d$  are used, with either linear  $\left(B_i = \sum_{d=0}^d a_d(i - i_0)^d\right)$  or exponential variables  $\left(B_i = \exp\left(\sum_{d=0}^d a_d(i - i_0)^d\right)\right)$ .<sup>16</sup>

Generally speaking, background modelling is faster and less influenced by low statistics in the spectrum. However, it also enhances the number of parameters in the fit, making it less stable. The authors refrain, in general, from using polynomials with exponential variables in the processing of individual pixel spectra, but for background correction in sum spectra featuring good statistics these polynomials are well suited.

In the processing of individual pixel spectra, polynomials with linear variables of 5th to 8th degree were found to be satisfying in order to describe noisy backgrounds with a low level of discontinuities. Depending on the statistics of the individual pixel spectra, higher degrees of polynomials were observed to destabilize the model and result in noisy images, at times containing visible artefacts. Polynomials of lesser degree were found to not correctly represent the shape of the background.

Background estimation (and subsequent subtraction) is better suited for complex background shapes with many discontinuities,<sup>15,16,31</sup> but it is also slower than background modelling, due to its iterative nature. The main reason for background discontinuities in XRF spectra acquired from historical paintings is the scattered primary radiation.

## 4.2 Identification of minor elements

In order to correctly fit a spectrum, a model is needed that includes all spectral features present in the dataset. It is, however, not time efficient to inspect a few hundreds of thousands spectra manually. Different approaches help to identify the elements whose signals contribute to the dataset. These approaches notwithstanding, the design of the fit model is strongly supported by knowledge of the sample. A scientist experienced in the investigation of 17th century paintings knows that the blue Co containing pigment is small, provided it is original paint and not a later addition (*e.g.* during a restoration treatment), and will include its characteristic minor components (K, Ni, As and Bi) in the fit model.

In the case of homogeneous samples, the elements present can be easily identified in a sum spectrum of all individual pixel spectra. If the lateral distribution of elements is very heterogeneous and some elements are present only in a small area, they might be easily missed. This can be due to interfering lines, which contribute stronger to the sum spectrum, or natural

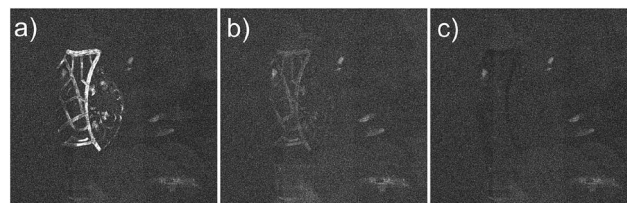


Fig. 6 Three  $\chi_r^2$  maps of the area shown in Fig. 3. In (a) As and Bi were omitted in the model, in (b) only Bi was omitted, while in (c) both elements were included. Remaining hotspots of  $\chi_r^2$  are due to absorption effects, as discussed in the text. Results obtained with Pymca.

statistical fluctuations of the background shape, which may mask lines of minor elements.

A simple approach to identify elements that are only present locally, but at a moderate to high concentration level is the calculation of the Maximum Pixel Spectrum (MPS). Here each channel is equal to the maximum value of that channel in the entire dataset. This approach was already described by Bright in 2004 for EDX,<sup>32</sup> but has found use only in part of the XRF imaging community, in spite of its ease of application. In Fig. 7 the sum spectrum and MPS of the area shown in Fig. 3 are shown. The elements Cr, Co, As and Bi were not included in the fit model used for the processing. Cr is not clearly discernible in the sum spectrum as the statistical fluctuations of the background mask its signals. As and Bi are not clearly discernible due to interferences with Fe and Pb, respectively. On the other hand, in the MPS Cr and Co are easily identified. As and Bi contribute in their most intense pixels much less than Pb and Hg in their hot spots and are thus hardly discernible.

Another approach to identify components not included in the fit is the inspection of  $\chi_r^2$  maps, as shown in Fig. 6. In (a) both As and Bi were left out of the fit model, in (b) As was included and in (c) both. The inspection of local sum spectra of areas with high local  $\chi_r^2$  thus allows for the identification of missing elements.

Some areas of higher  $\chi_r^2$  value remain in (c). These are due to absorption effects inside thick lead white layers. Here the Pb-L $_{\gamma}$  lines (around 15.0 keV) are overestimated, as their energy is above the Pb-L3 edge (13.0 keV) and radiation of this energy is consequently absorbed and gives rise to secondary fluorescence. This effect results in wrongly estimated peak profiles, which in turn results in enhanced  $\chi_r^2$  values. In a homogeneous sample this self-absorption could be easily modelled, but the heterogeneous nature of historical paintings hampers this, as a model that would better describe these hotspots would describe all other pixels less well.

The  $\chi_r^2$  maps presented here are rather favourable cases. The spectral range was limited from 2 keV to 16.5 keV, so that the

scattered characteristic radiation of the Rh tube was not included. The scattered radiation is more difficult to model and results in structured  $\chi_r^2$  maps in that a distinction between features resulting from missing elements, background shape, absorption effects and scattered radiation is challenging. Also, in data with low counting statistics the  $\chi_r^2$  map tends to be very noisy as well, complicating its interpretation further. The inspection of  $\chi_r^2$  maps after processing the data and iterative improvement of the fit model obviously benefits from fast fitting algorithms and automatic scheduling of fit processes for sub-parts of the dataset.

A third approach towards the identification of missing elements is to perform Principal Component Analysis (PCA) on the raw spectral data. A convenient routine is included in the PyMCA ROI imaging tool. A close inspection of Eigenimages and Eigenvectors provides clues of elements only weakly contributing to the acquired dataset. Local hotspots in the Eigenimages allow detection of their location and their identification from the corresponding Eigenspectrum and local sum spectra. However, due to the fact that the non-negative nature of the XRF spectral data is not preserved, the interpretation requires some experience. Also, while it is perfectly feasible to inspect, *e.g.*, 10 pairs of Eigenimages/spectra for a single dataset (or a representative sub-part), doing so for all 12 sub-parts of *Supper at Emmaus* is very time consuming. For this reason PCA was scarcely used by the authors when fitting datasets of historical paintings.

### 4.3 Performance of software

Unless otherwise noted, processing speeds achieved refer to data processing on a small server featuring 8 Intel® Core™ i7 CPUs 930 @ 2.8 GHz, 12 GB RAM and 15 WD20EARS hard disks with 2 TB space each mounted in three Chieftec SNT-3051SS SATA backplanes. The operating system was Fedora Linux 16.

**4.3.1 AXIL.** AXIL was developed at the University of Antwerp from the end of the 1970s until the mid of the 1990s. The limited computing resources available in the time of AXIL's development necessitated efficient programming to allow least squares fitting of non-linear functions with a broad range of background models. With current computing resources, AXIL allows for the processing of 30 spectra per second with NL-LS, which is considerably faster than the other XRF software packages discussed. AXIL is proprietary software that requires hardware keys (dongles) and its source code is not public. The software version used was LINUX SUSE-AXIL 1.01.

To date, the AXIL software package does not include features to create elemental distribution images from the fitted data. This has to be done by external routines, *e.g.* MICROXRF2.§

AXIL was originally intended for the deconvolution of single spectra. This heritage is visible in a limitation of the software: raw spectral data and fit results are read and written as human readable ASCII text files. The writing and subsequent reading of the saved result files to and from a local hard disk adds 20 ms

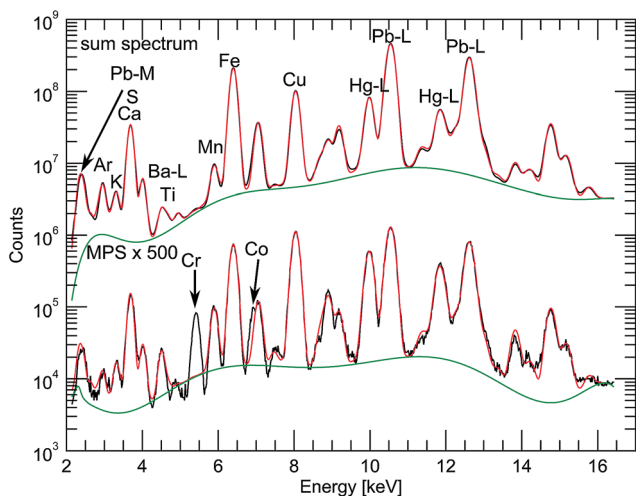


Fig. 7 Fit of sum spectrum and Maximum Pixel Spectrum. In the fit model Cr, Co, As and Bi are not included. Spectra fitted with PyMCA.

§ By Bart Vekemans (University of Gent, formerly: University of Antwerp), Version beta April 18, 2009.



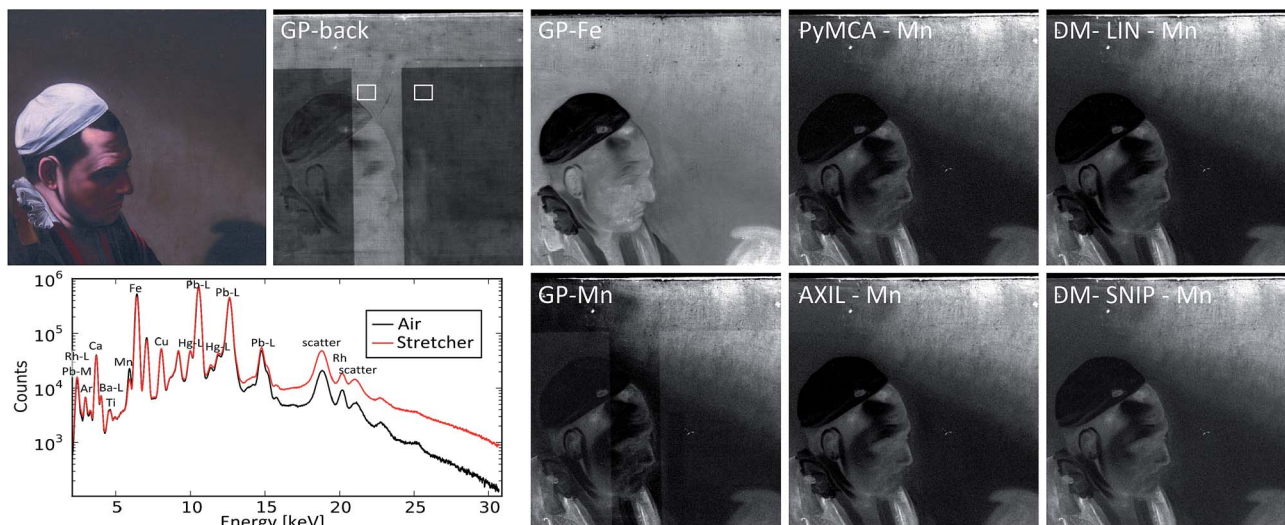


Fig. 8 Area around the innkeeper's head as photograph and distribution images of Fe and Mn as well as the intensity of the spectral background. The location of the local sum spectra is indicated in the background image. GP indicates results of GeoPIXE, DM-LIN results of datamuncher using an 8th degree polynomial with linear variables, DM-SNIP results of the same software package obtained with 15 iterations. PyMCA uses a SNIP filter, while AXIL uses its included "LIN" background model.

per spectrum to the processing time. This disadvantage of AXIL is even more severe on distributed systems, where the data are accessed *via* the network. This read/write access constitutes the major bottleneck when processing data with AXIL, which can be "widened" by the use of Solid State Disks or RAM disks.

MICROXRF2 allows for splitting the data processing into parallel processes. While this allows enhancement of the overall pace of data processing one has to be aware that the speed does not linearly scale with the number of processes. With 6 parallel processes on the above mentioned computing system the maximum gain in performance is 2.

AXIL was initially used by the authors for the processing of XRF imaging data acquired on historical paintings, especially those acquired at synchrotron radiation sources. It was not suitable for *in situ* data processing, but well suited for processing on the aforementioned server. The scans acquired in this time had typically just a few thousands to ten thousands of spectra.

Its use was discontinued when technical progress in the regime of scanners allowed for the routine acquisition of large scans ( $>10^5$  spectra). With so many spectra/files the access to the folders slowed down and data processing became impractical.

**4.3.2 PyMCA.** PyMCA is being developed and maintained by the Software Group of the European Synchrotron Radiation Facility (ESRF). Its source code is public and it is easy to install on modern Windows and Linux systems. The fact that it is, as opposed to the other software packages discussed, open source and freely available cannot be overestimated, as it allows a widespread use of the software.

PyMCA is based on the same mathematical routines as AXIL and features next to a modern GUI for XRF fitting and image manipulation many convenient tools. It provides a ROI imaging tool, which allows for the interactive inspection of data with a range of useful plug-ins.

Given its recent date of development, PyMCA was not under as much pressure to fully exploit the capabilities of computing resources as AXIL, so that in NL-LS it is considerably slower than AXIL. When version 4.4.0 of PyMCA was evaluated in 2010, its L-LS fitting routine performed at a comparable pace to AXIL and the quality of the results was satisfying. Given the fact that writing and reading fit results is not the bottleneck of data processing with PyMCA, it makes better use of parallel processing than AXIL and was found suitable for the final data processing on the aforementioned small server. However, version 4.4.0 was found too slow for *in situ* data processing as it did not drastically outperform AXIL in terms of processing pace.

The most recent version of PyMCA (version 4.7.3) achieves a multiple of the pace of AXIL in L-LS with a simple model, *e.g.* assuming fixed line ratios of all elements. If the ratios of the different fluorescence lines of the same element are not fixed (to take shifts in the line ratios due to absorption effects into account) the pace is roughly comparable to that of AXIL, when comparing single process operations.

Version 4.7.1 PyMCA contains a plugin to the ROI imaging tool that allows for the fast processing of XRF imaging data at a speed comparable to GeoPIXE (see below), which makes it suitable for *in situ* data processing. The fast fitting plug-in is not discussed here, as its inner workings are not described in the literature yet.

**4.3.3 GeoPIXE.** GeoPIXE has been developed at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) division of Earth Science and Resource Engineering since the late 1980s. It was the first software package for the evaluation of X-ray fluorescence spectra to include DA and provides a powerful GUI for data processing and evaluation. While, as the name suggests, it was written for the evaluation of PIXE data, it also found use in the evaluation of XRF data,

especially those acquired at synchrotron radiation sources.<sup>23</sup> It is, like AXIL, proprietary software and requires a licence.

Its DA approach to data evaluation allows for a data processing pace that is far above that of the conventional L-LS approach. The version 6.6 of GeoPIXE achieves a pace of 4800 spectra per s. However, one has to be aware of the fact that the EDF file format is not native to it, so that in this example GeoPIXE was not run under optimal conditions.

GeoPIXE's approach has one major limitation when analysing samples with heterogeneous scattering properties: it assumes the shape of the background to be constant, which can be a major limitation in the case of excitation with polychromatic X-rays.

In Fig. 8 elemental distribution images of the area around the head of the innkeeper are shown. This area is highlighted in Fig. 1 by the green shape. In the Mn distribution image obtained by GeoPIXE a dark, T-shaped shadow appears, which is not visible in those obtained by PyMCA and AXIL. The shape is also visible as a bright area in the image of the spectral background obtained with GeoPIXE. This T-shape is caused by the wooden stretcher present behind the canvas, resulting in a different shape of the spectral background due to stronger scattered primary radiation. The two local sum spectra shown illustrate this difference as one sums an area on top of the stretcher, while in the second area, just air is present behind the canvas.

The DA matrix assumes the background shape determined in the sum spectrum to be valid in an unmodified form in every individual spectrum. Thus, on top of the stretcher the intensity of the fluorescence lines is overcompensated for the contribution of the background. The strength of this effect is dependent on the signal-to-background ratio. Strong signals, such as Fe in the given case, are less affected, while weaker ones, such as Mn, are more strongly perturbed. In the image shown the fit parameters were chosen to highlight the effect by extending the energy range of the fit from 2.2 to 30.2 keV. When fitting the local sum spectrum above the stretcher with NL-LS and DA in GeoPIXE an attenuation of the Mn-K fluorescence intensity by 33% is observed. By limiting the energy range from 2.2 to 16.4 keV (slightly below the incoherent scatter peak of the Rh-K $\alpha$  line) the visibility of the effect is reduced, but the Mn signal is still attenuated by 14%.

The case discussed here is one of the most pronounced observed by the authors in the investigation of more than 100 historical paintings. In *Supper at Emmaus* the strongly absorbing paint layers are rather thin, compared to other paintings investigated. In the case of thicker paint layers or a homogeneous support (as with panel paintings) the shown effect is often not obvious. Also, if monochromatic excitation, whose scattering can be modelled in GeoPIXE, is used the effect is negligible. For this reason GeoPIXE was found satisfying for a limited group of datasets.

**4.3.4 Datamuncher.** For the reasons given above, none of the software packages discussed was found satisfactory for the processing of large datasets acquired in the analysis of historical paintings by scanning macro-XRF.

AXIL satisfies only the first (artefact-free elemental distribution images) of the three requirements, as it is too slow for the others. The evaluation of the entire dataset of *Supper at Emmaus* by AXIL would have taken 32 hours in one process or 16 hours with six parallel ones, requiring the better part of a week for the iterative improvement of the fit model. The individual scan of 550 × 550 pixels would have been evaluated in 2.8 hours in a single process, a time not acceptable to wait before making a decision on the next scan.

In the past years PyMCA only satisfied the first requirement; with the recent fast fitting plug-in it satisfies the second (fast *in situ* processing). However, the use of the fast fitting plug-in requires a high degree of user intervention. A uniform processing of several sub-parts of a dataset is time consuming and does not satisfy the third requirement (scheduling of data processing batches).

GeoPIXE fully satisfied the second requirement and partly the first, especially in the case of paintings with homogeneous X-ray scattering properties. It also features a batch mode for fitting different datasets with one model, so it partly satisfies the third requirement.

Overall none of the programs alone was found satisfying for the processing of data, so that an in-house software was developed at the University of Antwerp in IDL.<sup>¶</sup> It was labelled datamuncher, as it was meant to figuratively “eat” through a series of sub-sets of a scan. It started out as a simple routine calculating sum and maximum pixel spectra of datasets in a user defined folder and scheduling batch fits with AXIL or PyMCA.

Later DA routines were added. For this the sum spectrum of the dataset is first fitted with PyMCA. The resulting peak profiles are exported and treated as described in ref. 18. This routine allowed processing of more than 20 000 spectra per second, allowing for the processing of the entire dataset in less than 5 minutes. However, it features the same artefacts as mentioned above for GeoPIXE. Also, datamuncher does not include the pile-up rejection routines GeoPIXE employs,<sup>33</sup> as most elements of interest in the investigation of historical paintings are in the energy range below 20 keV and only in a few cases significant contributions from pile-up effects were observed.

To overcome the limits in background compensation two approaches were followed. In the first approach the fixed background  $B$  in eqn (3) was replaced with the members of a polynomial with linear variables and included in the DA matrix. The extension of the DA matrix resulted in a 20% loss of processing pace but allowed correction for fluctuations of the background shape. The results labelled “DM-LIN” in Fig. 8 were processed using a polynomial of 8th degree and show no artefacts resulting from the shape of the spectral background.

As a linear polynomial background is not suitable to model complex background shapes also the SNIP background correction was implemented, following the worked example from ref. 28. This approach requires *ca.* 10 times the processing time compared to the use of a fixed background shape but allows

¶ Interactive Data Language, Exelis Visual Information Solutions, Boulder, CO, USA.

taking more complex background shapes in data with sufficient statistics into account. The results labelled “DM-SNIP” in Fig. 8 were obtained with a SNIP filter with 15 iterations.

The software further includes routines for the (manual) stitching of elemental distribution images of sub-parts of the data and the correction of acquisition artefacts similar to those in ref. 34. Furthermore, a routine for the interpretation of two element scatter plots of elemental distribution images is implemented. This tool proved to be of value to distinguish between original smalt paint and 19th century Co salts added as a drier in Saul and David from the workshop of Rembrandt.<sup>35</sup> datamuncher is available from sourceforge.||

**4.3.5 Comparison of fit quality.** In the processing of the *Supper at Emmaus* dataset, all software packages yielded results of comparable quality in the energy range discussed. While the quality of the results is comparable, the results themselves are not. This is due to the fact that especially for L-lines the exact transition of which the intensity is returned differs between the software packages. Further, the fit results of PyMCA are constrained to positive values and those of AXIL to integers. Both features accurately represent the photon counting process of XRF detection, but one has to be aware of this difference. The non-negative restriction is biasing the average results towards a higher value and complicating the identification of spectral interferences in distribution images. However, it prevents the misinterpretation of incorrectly scaled elemental distribution images.

From our experience, PyMCA is superior when fitting L- and M-lines in the energy range below 3.5 keV. In the same software, the ratio of L-lines resulting from a vacancy in the same shell is calculated based on the fundamental parameters approach. On the one hand, this allows the user to model different detection geometries and stratigraphies, as it was used in the study of Leonardo da Vinci's painting technique.<sup>36</sup> On the other hand, it requires a certain degree of user intervention and knowledge of the sample. The absorption effects for the model used to fit the detail shown in Fig. 3 were estimated from the sum spectrum of the entire dataset, so that, in areas of high Pb abundance, the Pb-L<sub>γ</sub> lines were overestimated. By modelling self-absorption effects based on local sum-spectra of these areas a better separation of Pb-L and Bi-L lines can be achieved at the price of a slightly inferior description of Pb-L lines in pixels with moderate Pb abundance.

The authors found it easier to resolve the contributions of these two elements in AXIL. Here L-lines are not grouped by the vacancy they result from, but by their Siegbahn name. This allows us to treat absorption effects without the complicated modelling. While it allows an easier processing of imaging data of heterogeneous samples, it provides less insight into the sample during the processing of an individual spectrum.

Neither GeoPIXE nor datamuncher is capable of fitting fluorescence lines originating from the same inner shell vacancy independently in the processing of imaging data and thus are inferior to AXIL and PyMCA in the resolution of Pb-L

and Bi-L lines. Furthermore, the absence of this feature prohibits the study of absorption effects, as discussed elsewhere.<sup>9</sup>

## 5 Conclusions

The XRF dataset acquired by scanning *Supper at Emmaus* constitutes a considerable challenge for evaluation, as (a) it contains 12 sub-parts, (b) it features spectral interferences of weakly and strongly contributing elements (Co with Fe, As and Bi with Pb) and (c) it features, due to the heterogeneous scattering properties and the polychromatic excitation, a high level of variability in the shape of the spectral background. *Supper at Emmaus* is currently the largest object entirely mapped with XRF and the size of the dataset (3.4 MPixel) constitutes a further challenge, although it is by far not the largest reported in the literature.

Due to the complexity of the dataset, ROI imaging was not found to be a suitable tool for data processing, as only a few elemental distribution images can be obtained without significant artefacts, making full spectral deconvolution necessary.

In the case of such a complex XRF dataset the definition of an initial fit model that takes all minor elements and spectral features into account is unlikely. The initial model is in general based on the sum spectrum of all pixels and on *a priori* knowledge. This model can be easily improved by using the maximum pixel spectrum (MPS) to identify elements not obvious in the sum spectrum. After a first processing of the dataset, the results are inspected for anomalies, *e.g.* areas of strongly negative fit results, and areas of high  $\chi_r^2$  values. Local sum spectra of these areas are inspected and the fit model is improved until these anomalies are either removed from the elemental distribution images or sufficiently explained and found not to be misleading.

It is obvious that for an iterative improvement of a fit model for a dataset with many sub-parts a software package is needed that allows for fast processing data with a minimum of user intervention. When the authors started with the investigation of large areas on historical paintings by scanning macro-XRF in 2008, the only software package capable of processing the data at a sufficient pace for this was GeoPIXE. Today PyMCA and datamuncher also provide these capabilities.

This illustrates the constant progress that is made in terms of data evaluation routines, which is not the least driven by development in instrumentation. Given this constant development, this article can only present a snapshot in time, as new features are added to GeoPIXE and PyMCA constantly.

All three available software packages discussed (AXIL, PyMCA and GeoPIXE) could provide largely artefact-free elemental distribution images of comparable quality. As all of them have their individual advantages and drawbacks, datamuncher was developed at the University of Antwerp to exploit the advantages while avoiding the drawbacks, *e.g.* scheduling of batch jobs for the processing of sub-parts of the dataset by AXIL and PyMCA and the use of dynamic analysis with variable background shapes.

|| <http://sourceforge.net/projects/datamuncher/>

The authors have largely stopped using AXIL, as its unique features (fast NL-LS and Siegbahn grouped lines) in general do not justify the long processing times. Given the file format employed by AXIL it also benefits less from parallel processing on modern systems than other software packages. However, recent development in multi-channel analyser (MCA) technology has resulted in new instruments that can handle up to several millions counts per second, but with a count rate dependent energy resolution, such as the Xspress3 (Quantum Detectors, Harwell Oxford, UK) or FalconX (XIA LLC, Hayward CA, USA). As discussed, L-LS is sufficient for the processing of imaging data. However, as energy resolution is one of the non-linear parameters that is kept constant during L-LS, NL-LS might be necessary to correctly process the data acquired with these MCAs, provided it features sufficient statistics. As AXIL features the fastest NL-LS routines for the deconvolution of XRF spectra known to the authors, it might be a suitable tool to treat such data.

The findings and conclusions presented here on different approaches and software packages are inevitably shaped by the complex nature of historical paintings. While they will be of interest to other scientists investigating, e.g., biological or geological samples by XRF imaging, one has to be aware that they might not be directly transferable and that of the software packages discussed only those features of use to the investigation of historical paintings were discussed.

## Acknowledgements

The authors would like to thank W. de Nolf, A. Rothkirch, C. Ryan, A. Solé, B. Vekemans, P. van Espen and L. Vincze for their fruitful discussions over the years. Furthermore, the authors thank D. Swetozoff for his support. M. Alfeld was from 2009 to 2013 the recipient of a Ph.D. fellowship of the Research Foundation-Flanders (FWO, Brussels).

## References

- 1 M. Haschke, U. Rossek, R. Tagle and U. Waldschläger, *Adv. X-Ray Anal.*, 2012, **55**, 286–298.
- 2 S. Chen, J. Deng, Y. Yuan, C. Flachenecker, R. Mak, B. Hornberger, Q. Jin, D. Shu, B. Lai, J. Maser, C. Roehrig, T. Paunesku, S. C. Gleber, D. J. Vine, L. Finney, J. VonOsinski, M. Bolbat, I. Spink, Z. Chen, J. Steele, D. Trapp, J. Irwin, M. Feser, E. Snyder, K. Brister, C. Jacobsen, G. Woloschak and S. Vogt, *J. Synchrotron Radiat.*, 2014, **21**, 66–75.
- 3 C. Ryan, D. Siddons, R. Kirkham, P. Duun, A. Kuczewski, G. Moorhead, G. De Geronimo, D. Paterson, M. de Jonge, R. Hough, M. Lintern, D. Howard, P. Kappen and J. Cleverley, *AIP Conf. Proc.*, 2010, **1221**, 9–17.
- 4 R. Kirkham, P. Dunn, A. Kuczewski, D. Siddons, R. Dodanwela, G. Moorhead, C. Ryan, G. De Geronimo, R. Beuttenmuller, D. Pinelli, M. Pfeffer, P. Davey, M. Jensen, D. Paterson, M. de Jonge, D. Howard, M. Küsel and J. McKinlay, *AIP Conf. Proc.*, 2010, **1234**, 240–243.
- 5 C. G. Ryan, D. P. Siddons, R. Kirkham, Z. Y. Li, M. D. de Jonge, D. J. Paterson, A. Kuczewski, D. L. Howard, P. A. Dunn, G. Falkenberg, U. Boesenberg, G. De Geronimo, L. A. Fisher, A. Halfpenny, M. J. Lintern, E. Lombi, K. A. Dyl, M. Jensen, G. F. Moorhead, J. S. Cleverley, R. M. Hough, B. Godel, S. J. Barnes, S. A. James, K. M. Spiers, M. Alfeld, G. Wellenreuther, Z. Vukmanovic and S. Borg, *J. Phys.: Conf. Ser.*, 2014, **499**, 012002.
- 6 O. Scharf, S. Ihle, I. Ordavo, V. Arkadiev, A. Bjeoumikhov, S. Bjeoumikhova, G. Buzanich, R. Gubzhokov, A. Guenther, R. Hartmann, M. Kuehbacher, M. Lang, N. Langhoff, A. Liebel, M. Radtke, U. Reinholz, H. Riesemeier, H. Soltau, L. Strueder, A. F. Thuenemann and R. Wedell, *Anal. Chem.*, 2011, **83**, 2532–2538.
- 7 J. Dik, K. Janssens, G. Van der Snickt, L. van der Loeff, K. Rickers and M. Cotte, *Anal. Chem.*, 2008, **80**, 6436–6442.
- 8 M. Alfeld, K. Janssens, K. Appel, B. Thijssse, J. Blaas and J. Dik, *Z. Kunsttechnol. Konserv.*, 2011, **25**, 157–163.
- 9 M. Alfeld, G. Van der Snickt, F. Vanmeert, K. Janssens, J. Dik, K. Appel, L. van der Loeff, M. Chavannes, T. Meedendorp and E. Hendriks, *Appl. Phys. A: Mater. Sci. Process.*, 2013, **111**, 165–175.
- 10 M. Alfeld, D. Siddons, K. Janssens, J. Dik, A. Woll, R. Kirkham and E. van de Wetering, *Appl. Phys. A: Mater. Sci. Process.*, 2013, **111**, 157–164.
- 11 M. Alfeld, K. Janssens, J. Dik, W. De Nolf and G. Van der Snickt, *J. Anal. At. Spectrom.*, 2011, **26**, 899–909.
- 12 M. Alfeld, J. V. Pedroso, M. van Eikema Hommes, G. Van der Snickt, G. Tauber, J. Blaas, M. Haschke, K. Erler, J. Dik and K. Janssens, *J. Anal. At. Spectrom.*, 2013, **28**, 760–767.
- 13 B. Vekemans, K. Janssens, L. Vincze, F. Adams and P. Van Espen, *X-Ray Spectrom.*, 1994, **23**, 278–285.
- 14 P. Van Espen, H. Nullens and F. Adams, *Nucl. Instrum. Methods*, 1977, **142**, 243–250.
- 15 B. Vekemans, K. Janssens, L. Vincze, F. Adams and P. van Espen, *Spectrochim. Acta, Part B*, 1995, **50**, 149–169.
- 16 V. A. Solé, E. Papillon, M. Cotte, P. Walter and J. Susini, *Spectrochim. Acta, Part B*, 2007, **62**, 63–68.
- 17 C. Ryan and D. Jamieson, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 1993, **77**, 203–214.
- 18 C. Ryan, D. Jamieson, C. Churms and J. Pilcher, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 1995, **104**, 157–165.
- 19 C. Ryan, *Int. J. Imag. Syst. Tech.*, 2000, **11**, 219–230.
- 20 C. Ryan, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2001, **181**, 170–179.
- 21 C. Ryan, E. van Achterbergh, C. Yeats, T. Win and G. Cripps, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2002, **189**, 400–407.
- 22 C. Ryan, E. van Achterbergh, C. Yeats, S. Drieberg, G. Mark, B. McInnes, T. Win, G. Cripps and G. Suter, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2002, **188**, 18–27.
- 23 C. Ryan, B. Etschmann, S. Vogt, J. Maser, C. Harland, E. van Achterbergh and D. Legnini, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2005, **231**, 183–188.
- 24 S. Vogt, *J. Phys. IV*, 2003, **104**, 635–638.
- 25 N. Eastaugh, V. Walsh, T. Chaplin and R. Siddall, *The pigment compendium*, Elsevier, Amsterdam, 2004.

- 26 L. Robinet, M. Spring, S. Pages-Camagna, D. Vantelon and N. Trcera, *Anal. Chem.*, 2011, **83**, 5145–5152.
- 27 P. L. Manning, N. P. Edwards, R. A. Wogelius, U. Bergmann, H. E. Barden, P. L. Larson, D. Schwarz-Wings, V. M. Egerton, D. Sokaras, R. A. Mori and W. I. Sellers, *J. Anal. At. Spectrom.*, 2013, **28**, 1024–1030.
- 28 P. Van Espen, *Handbook of X-Ray Spectrometry*, ed. R. Van Grieken and A. Markowicz, Marcel Dekker, Inc., New York, 2nd edn, 2002, pp. 239–340.
- 29 E. Clayton and C. Ryan, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 1990, **49**, 161–165.
- 30 C. Ryan, E. Clayton, W. Griffin, S. Sie and D. Cousins, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 1988, **34**, 396–402.
- 31 V. A. Solé, *Understanding and using the strip background*, <http://pymca.sourceforge.net/stripbackground.html>, last accessed, 01 August 2014.
- 32 D. S. Bright and D. E. Newbury, *J. Microsc.*, 2004, **216**, 186–193.
- 33 C. Ryan, E. van Achterbergh and D. Jamieson, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2005, **231**, 162–169.
- 34 A. Anitha, A. Brasoveanu, M. Duarte, S. Hughes, I. Daubechies, J. Dik, K. Janssens and M. Alfeld, *Signal Process.*, 2013, **93**, 592–604.
- 35 P. Noble, A. van Loon, M. Alfeld, K. Janssens and J. Dik, *Techè*, 2012, **35**, 36–45.
- 36 L. de Vignerie, P. Walter, E. Laval, B. Mottin and V. Solé, *Angew. Chem., Int. Ed.*, 2010, **49**, 6125–6128.