



Simplex Volume Maximization (SiVM): A matrix factorization algorithm with non-negative constraints and low computing demands for the interpretation of full spectral X-ray fluorescence imaging data



Matthias Alfeld^{a,h,*}, Mirwaes Wahabzada^b, Christian Bauckhage^c, Kristian Kersting^d, Geert van der Snickt^e, Petria Noble^{f,1}, Koen Janssens^e, Gerd Wellenreuther^g, Gerald Falkenberg^a

^a DESY, Hamburg, Germany

^b INRES-Phytomedicine, University of Bonn, Germany

^c Fraunhofer IAIS, Sankt Augustin, Germany

^d Computer Science Department, TU Dortmund, Germany

^e Department of Chemistry, University of Antwerp, Antwerp, Belgium

^f Mauritshuis, The Hague, Netherlands

^g European XFEL GmbH, Hamburg, Germany

^h Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 8220, Laboratoire d'archéologie moléculaire et structurale (LAMS), 4 place Jussieu, 75005 Paris, France

ARTICLE INFO

Article history:

Received 11 October 2016

Received in revised form 1 February 2017

Accepted 1 February 2017

Available online 3 February 2017

ABSTRACT

Technological progress allows for an ever-faster acquisition of hyperspectral data, challenging the users to keep up with interpreting the recorded data. Matrix factorization, the representation of data sets by bases (or loads) and coefficient (or score) images is long used to support the interpretation of complex data sets. We propose in this publication Simplex Volume Maximization (SiVM) for the analysis of X-ray fluorescence (XRF) imaging data sets. SiVM selects archetypical data points that represents the data set and thus provides easily understandable bases, preserves the non-negative character of XRF data sets and has low demands concerning computing resources. We apply SiVM on an XRF data set of Hans Memling's *Portrait of a man from the Lespinette family* from the collection of the Mauritshuis (The Hague, NL) and discuss capabilities and shortcomings of SiVM.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Most spectroscopic techniques were developed for the acquisition of individual or small series of spectra. Technological progress has allowed to extend many of these techniques to spectroscopic imaging, which yields information on the heterogeneity of samples and provides more representative results [1].

Spectroscopic imaging creates its own challenges, namely that the data evaluation needs to be automatized, as a manual inspection of more than a few tens of spectra is not realistic. It is also desirable to have the results quickly after an experiment, as it allows to take the results of the previous one into account while planning the next one. This is especially the case if the time for experiments is limited, e.g. during a synchrotron beam time or an in-situ measurement for the investigation of cultural heritage objects.

X-ray fluorescence (XRF) imaging has in this respect an advantage over many other techniques, as XRF spectra can be described by a straightforward set of equations. These equations allow the definition of an analytical model that can be fitted to the acquired spectra. The state of the art XRF fitting routines have been recently summarized [2].

However, even the elemental distribution images obtained in this way can still contain significant amount of redundancies and non-obvious correlations. Several approaches have been made to reduce the first and highlight the latter. Further, when utilizing such an approach one has to be aware that understanding the data is the first step. In the second step the results need to be presented and communicated to other scientists, which is supported by easily understandable images.

It is simple to design and measure or simulate a test data set that verifies the validity of a novel approach. However, these data sets do not represent the complexity commonly encountered in real samples and the noise level can be adjusted to produce convincing results. Historical paintings are among the most complex samples that can be investigated by XRF imaging, as they have a broad range of elements present in complex mixtures and layers in superposition. For this reason we decided to use a data set of Hans Memling's *Portrait of a man from the Lespinette family* (see Fig. 4) from the collection of the Mauritshuis (The Hague,

* Corresponding author at: Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 8220, Laboratoire d'archéologie moléculaire et structurale (LAMS), 4 place Jussieu, 75005 Paris, France.

E-mail address: matthias.alfeld@upmc.fr (M. Alfeld).

¹ Current address: Painting conservation, Rijksmuseum, Amsterdam, Netherlands.

the Netherlands) to demonstrate the value of Simplex Volume Maximization (SiVM), a novel approach for factorizing complex data sets with non-negative constraints. The painting was investigated by XRF to build a database for comparison with other paintings by Memling and his contemporaries.

2. Matrix factorization

2.1. Established approaches

The manual inspection of scatter plots of the intensities of fluorescence lines for manual clustering is an easy, interactive way to obtain attractive images and identify correlations between elements [3–5]. However, it requires a high degree of attention, it is easily biased by assumptions about the object investigated and the number of scatter plots that can be visually inspected is ultimately limited.

An (semi-)automatic alternative is Factor Analysis (see Eq. (1)). Here the data matrix V is decomposed into two matrices W and H . The matrix V has columns (n) corresponding to pixels represented by the recorded intensity of d spectral channels. The basis vectors W describe the loading of the data set's variables on k bases, while the coefficients in H represent the lateral distribution of the degree that bases describe the data matrix V . The difference between V and its representation by W and H is in general expressed in the residue R .

$$V^{d \times n} = W^{d \times k} H^{k \times n} + R^{d \times n} \quad (1)$$

The data matrix V is obtained by reshaping the hyperspectral data block with two lateral and one spectral dimension into one lateral (n) and one spectral dimension (d), while the coefficient images shown below are obtained by reshaping H for each component k back into a 2-dimensional image.

Factor Analysis in the form of Principal Component Analysis (PCA) followed by K-means clustering is long known in XRF imaging [6]. It is limited by the fact that K-means clusters are binary, i.e. a pixel in the image belongs to one cluster or the other, never to a mixture. PCA yields a good representation of the data set, but its interpretation and presentation is at times challenging, as it does not preserve the non-negative nature of XRF data sets.

Factor Analysis with non-negative constraints to W and H and the later not restricted to binary values allows to obtain easily understandable and presentable results. Many routines used today are based on Non-negative Matrix Factorization (NMF), as described by Lee and Seung [7]. While similar algorithms have been used for a while in other forms of spectroscopic imaging [8] its application to XRF imaging data has been limited so far. It has been used during studies of the metal uptake of argillaceous rocks by micro-XRF [9,10]. The authors of this paper have used it to distinguish between components containing different ratios of Co and Ni in a painting by Rembrandt [11] and for the near-real time detection of self-absorption artifacts in alloy samples [12].

The studies mentioned above made use of elemental distribution images yielded by fitting of the raw spectral data. It is obvious that by using raw spectral data for the Factor Analysis one could simplify the process of data evaluation. Furthermore, the correct calculation of elemental distribution images is not trivial, so that another source of errors could be avoided. Santos et al. did full spectral factorization with the NMF function of PyMCA [13] to distinct between authentic and fake historical enamels [14]. Multivariate Curve Resolution (by) Alternating Least Square (MCR-ALS) was used on full spectral XRF data acquired on Piet Mondrian's *Broadway Boogie Woogie* and Jackson Pollock's *Number 1A, 1948* [15,16]. As in NMF in MCR-ALS W and H are determined iteratively but meaningful initial estimates of W are used.

However, NMF is not optimal for the analysis of full spectral data sets as it is an iterative process that requires at each iteration a matrix inversion with the whole process taking several tens to hundreds of

iterations. As discussed below, the processing time of a data set is linearly depending on d and n .

Further, Factor Analysis often allows new insights into the data set, but one has to be aware of its ambiguity. By the insertion of a neutral element (a non-zero matrix and its inverse) in Eq. (1) an infinite number of solutions can be found. This is the rotation ambiguity (Eq. (2)) [17].

$$V^{d \times n} \approx (W^{d \times k} M^{-1}) (M H^{k \times n}) = W_{\text{new}}^{d \times k} H_{\text{new}}^{k \times n} \quad (2)$$

2.2. Simplex Volume Maximization

In this work we use a factorization, where the basis vectors W correspond to the extreme, archetypal data points that are located on the convex hull of the d dimensional point cloud of the data set [18]. For that we employ Simplex Volume Maximization (SiVM) [19], a recently developed approximation that represents the data as mixture of extreme data points. It is based on an efficient and linear time greedy approach that iteratively determines suitable basis vectors by using the notion of distance geometry. As the bases are real data points, they are easily interpretable and uncover the variations existing in the data.

The principle of SiVM is illustrated in Fig. 1 on a simple test data set ($d = 2$). The data set features three components with a common non-zero origin and added random noise.

With SiVM a k -dimensional simplex is calculated which engulfs the data set and each vertex (corner) of the simplex is an actual data point and used as a base of W . In Fig. 1a and b the case with $k = 3$ is shown, which results in a triangle engulfing the data set. In Fig. 1b the degree to that these three bases represent the data is shown by the intensity of the colors red (base 0), green (base 1) and blue (base 2). Extreme points that are outside of the simplex are for the calculation of H projected on its surface, so that H features values between 0 and 1.

It is obvious that this is not enough to fully explain the data set as only two of the three components are represented by bases. Enhancing k to 4 would not yield a point that does represent the third component, as SiVM starts with an extreme point as first base and is then moving on to the most isolated point in the hull of the data set as next base, thus maximizing the volume of the simplex. However, if the number of bases is enhanced the 8th base represents the third component (s. Fig. 1c), but several other bases are yielded that are similar not in value, but in the ratio of the two variables. Thus, they do not provide much additional value in terms of interpretable results and contain ultimately redundant information. The data set was deliberately chosen to show the limitations of SiVM and why in the real data redundant bases are obtained. If k is smaller than d the (relative) number of redundant bases is considerably reduced.

The main advantage of SiVM is the way it computes the factorization. First, it determines a matrix of basis vectors W consisting of extreme, archetypal columns selected from input data matrix V . Once the bases are determined, the coefficients in H can be computed in a single pass over the entire dataset. This allows for a fast computation of the decomposition compared to batch methods, such as NMF that consider all data points in each iteration in order to perform an update on both matrices W and H . This is illustrated in Fig. 2, where duration of the factorization of random data sets with variable dimensions by SiVM and NMF is shown. Although SiVM is rather fast and its processing speed not strongly dependent of d , one has to be aware that in terms of factorization time and completeness of representation an optimized PCA algorithm outperforms both approaches on the data sets discussed, but yields less easily interpretable results, as mentioned above. We use the PCA, NMF and SiVM functions of the Python Matrix Factorization module (pymf) [20].

Another advantage of SiVM is that the components are represented by actual data points and are easily interpretable, as they do not contain any artifacts. The disadvantages of SiVM are that the bases obtained are

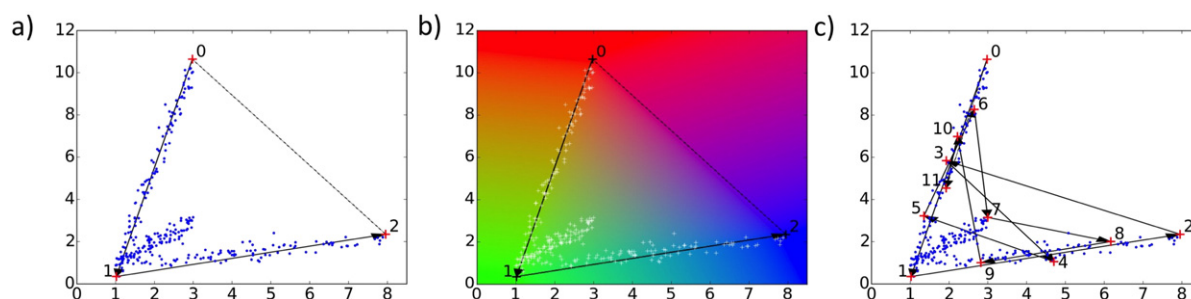


Fig. 1. Principle of SiVM on a test data set. a) bases 0 to 2 with $k = 3$ b) value of H for the three bases 0 (red) 1 (green) and 2 (blue) c) bases 0 to 11 with $k = 12$.

limited to statistics of individual actual data points, that components that do not reach to the data sets external hull might be missed and if a high value of k is used redundant bases will be obtained.

Using archetypical points of the data set reduces the rotation ambiguity, as W has to be selected from a limited range of spectra. This loss of freedom might, on the other hand, result in a less complete description of the data set.

3. Experimental

The *Portrait of a man from the Lespinette family* was investigated in the conservation studio of the Mauritshuis making use of a commercial, mobile macro-XRF scanner, the Bruker M6 Jetstream. The instrument features a 30 W Rh-anode X-ray tube with a polycapillary optic that allows for a minimal beam size of 50 μm . The emitted fluorescence radiation is recorded with a SD-detector. Both devices are mounted with two cameras on a motorized stage ($80 \times 60 \text{ cm}^2$, $h \times v$). The painting was scanned with a step size of 75 μm and a dwell time of 0.01 s, resulting in an acquisition time of 45 min.

The X-ray tube was operated at 50 kV and 0.6 mA. The polycapillary used as primary optic attenuated the high-energy part of the bremsstrahlung continuum (above approx. 23 keV) exciting the painting. At the same time the air between instrument and painting surface acted as a high pass filter, as a safety distance of 2 cm was kept. With the given dwell time the instrument was capable of detecting elements from K to Mo by their K-lines and elements heavier than Cd by their L-lines. A full description and characterization of the instrument can be found in [21].

The data was processed in two manners. First the established approach, the calculation of elemental distribution images, was done, using PyMCA [13] and datamuncher [2]. In parallel, the same raw data set was factorized by SiVM in the p06pymfgui [12] to yield coefficient images. All elemental distribution images are shown in grey scale with a brighter pixel indicating a stronger signal, all coefficient images H are shown in temperature scale. In all images the color scale was adjusted for enhanced readability by gamma correction.

4. Results and discussion

In Fig. 3 sum and maximum pixel spectrum of the raw spectral data are shown. Only primary peaks are indicated and weak signals of omnipresent elements that give noisy elemental distribution images are omitted. In the sum spectrum the sum of all spectra in the data set is shown. In the maximum pixel spectrum the most intense value any channel has in the entire data set is shown [22]. So, the maximum pixel spectrum gives an impression of the most intense signal of each element in the data set and allows to identify elements that contribute only to a few pixels, such as Co and Zn in this case. Especially the Co signal in the maximum pixel spectrum is close to its limit of detection.

In Fig. 4 the visible light image and the elemental distribution images of the *Portrait of a man from the Lespinette family* are shown. The first information that these elemental distribution images provide is an indication of the pigments used: Gypsum (CaSO_4) or Chalk (CaCO_3) and Bone black (a black carbon based pigment, containing (CaCO_3) and $(\text{Ca}_3(\text{PO}_4)_2)$ as remnants of the bones burned during its production). Earth pigments with Fe oxides as main coloring agents and a blue pigment, either azurite or its synthetic analogue blue verditer (both: $\text{Cu}_3(\text{CO}_3)_2(\text{OH})_2$). Finally, the red vermilion (HgS) and lead white (PbCO_3 and $2\text{PbCO}_3 \cdot \text{Pb}(\text{OH})_2$). Further the elements Zn and Co were found present in a few isolated spots, which can be attributed to the use of the anachronistic pigments cobalt blue (CoAl_2O_4) and zinc white (ZnO) in limited, restoration treatments. At the lower right edge restoration treatments are clearly discernible in the Ca, Fe and Co distribution images. The distribution of elements shows that the artists mixed pigments to achieve different effects. For example vermilion is present in many areas that are not explicitly red. However, no significant differences between visible impression of the image and elemental distribution images are obvious.

In Fig. 5 coefficient images H and the associated bases of W are shown. A cross indicates the position of the base in the coefficient image. For this factorization 1200 ($=d$) energy channels between 1 and 23 keV were used with 16 ($=k$) bases. To select k , we processed a sub set of V ($\sim 10\%$ of the original size) with variable k . We used the smallest value after which no significant reduction of the residue R

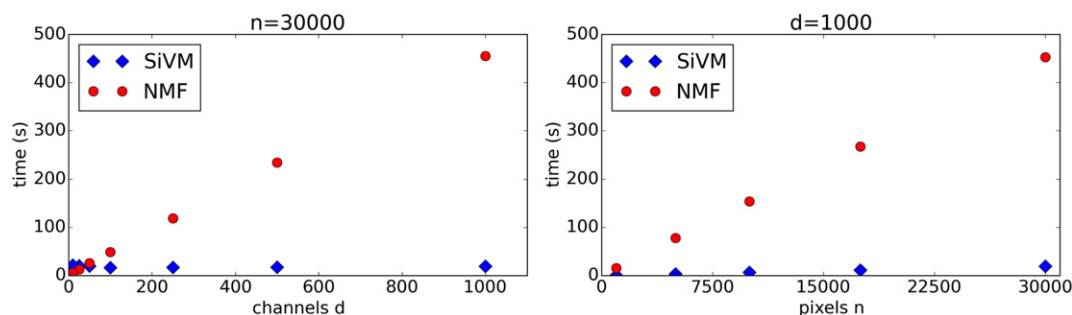


Fig. 2. Factorization speeds of NMF (500 iterations) and SiVM on a random data set with $n = 30,000$ and variable d (left) and $d = 1000$ with variable n . For both methods $k = 8$ bases were used.

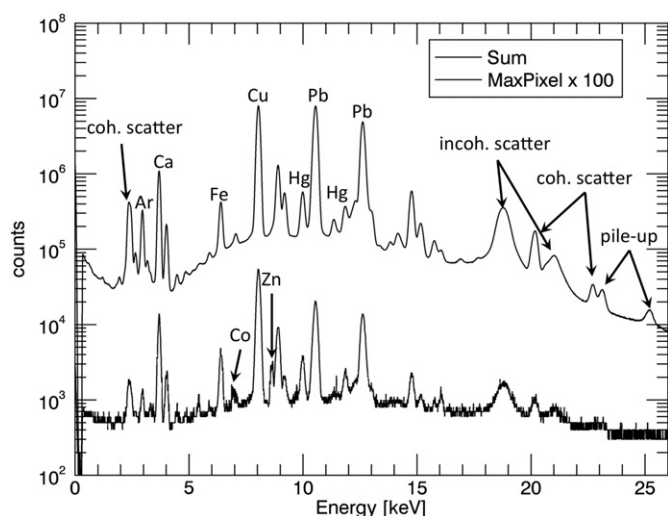


Fig. 3. Sum spectrum and maximum pixel spectrum of the raw spectral data.

was observed to achieve a minimal degree of redundancy among the data with a maximal degree of representation.

To obtain bases representative for the painting and not the restored edges at each side of the scanned area stripes of 5 pixels width were excluded when calculating W . Further, to enhance the statistical weight of signals of lesser intensity the square root of V was used to find the bases constituting W . For the calculation of H the complete data set V without any additional normalization was used. These parameters were chosen as they yield the clearest and easiest readable results but comparable results were achieved with different parameters. The SiVM factorization took 230 s on a normal office PC.

As discussed above, SiVM factorizations have a certain degree of redundancy. So Pb is represented by two bases (0 and 6) from the same area of the painting. The bases were summed in Fig. 5 for enhanced

readability, with the coefficient image well agreeing with the Pb elemental distribution image.

The coefficient image of the bases 2 and 8 agrees well with the Ca distribution image, but next to Ca the spectrum also contains signals by elements such as Fe, Cu and Pb. This is due to the fact that no pixel featuring only Ca is present in the data set V .

The strong Cu signal results in the Cu distribution being represented by in total seven bases, which feature different combinations of Cu with other elements. In practice, these bases can be considered largely redundant as the statistics of the data set do not allow for an informative separation of different components, so that also these bases are summed up here for enhanced readability.

Base 4 features strong Hg and Pb signals. As discussed above, vermilion (HgS) is in general used in a mixture with other pigments and the coefficient image highlights the areas in that vermilion and lead white were used together and is a good approximation of the Hg distribution image with one notable exception: vermilion is used together with earth pigments to paint the rope holding the sitters pendant, but this line does not show up in the coefficient image of Base 4. Instead it is clearly visible in Base 14.

Base 14 has the strongest Fe signal of all bases, so that its coefficient image is similar to the Fe distribution image. As also Hg is present the coefficient image also explains a part of this element's distribution, especially in the rope of the pendant. However, the coefficient image is also a bit misleading as areas that feature a strong Fe signal, but no Hg signal, have high values in the coefficient image, such as the landscape painted above the sitter's left shoulder.

Base 12 is the only base to feature Zn and thus highlights the presence of this element. However, its coefficient image does not resemble the Zn distribution image, as the other pixels containing strong Zn signals feature different combinations of elements.

Base 7 features the elements Ca, Fe, Cu and Pb, which are also present in other bases. The coefficient image is of great interest, as it features as "halo" around the sitter's head. This highlights the area in that the artist painted outside of the prepared reserve and the Ca containing paint

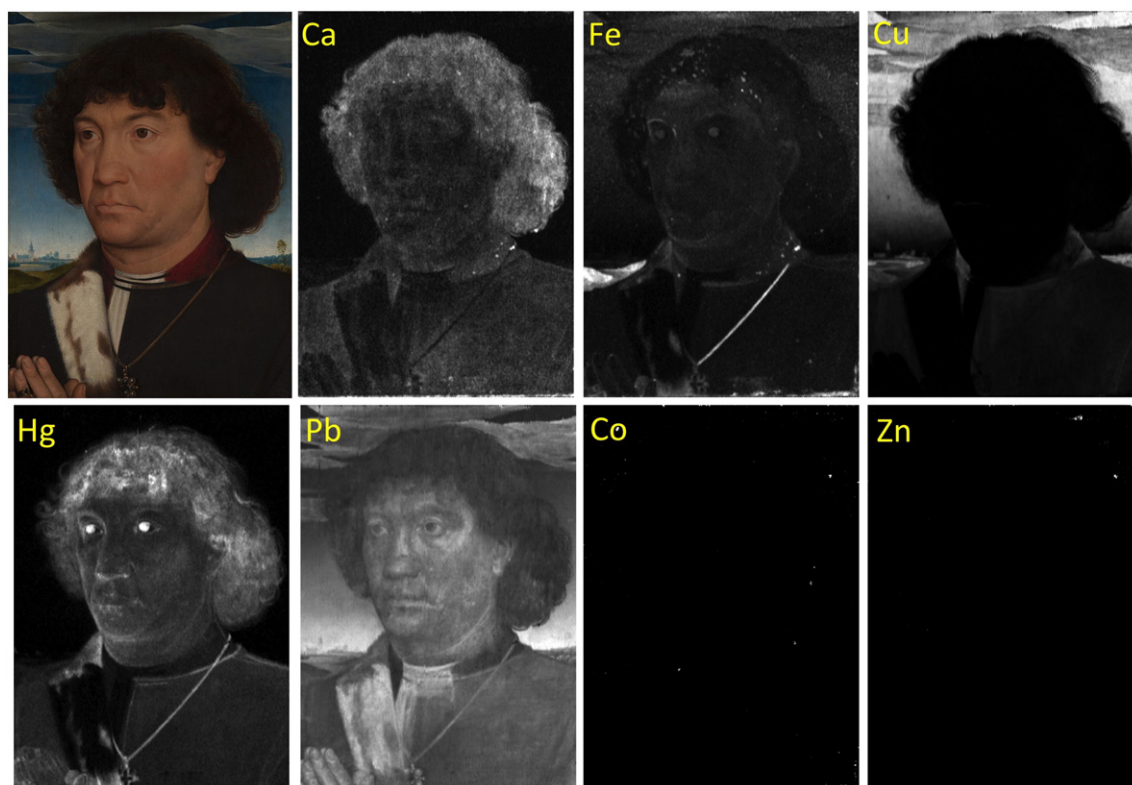


Fig. 4. Portrait of a man from the Lespinette family, Hans Memling, 1485–1490, Mauritshuis, The Hague, Oil on panel, 30.1 × 22.3 cm² as photograph and as elemental distribution images.

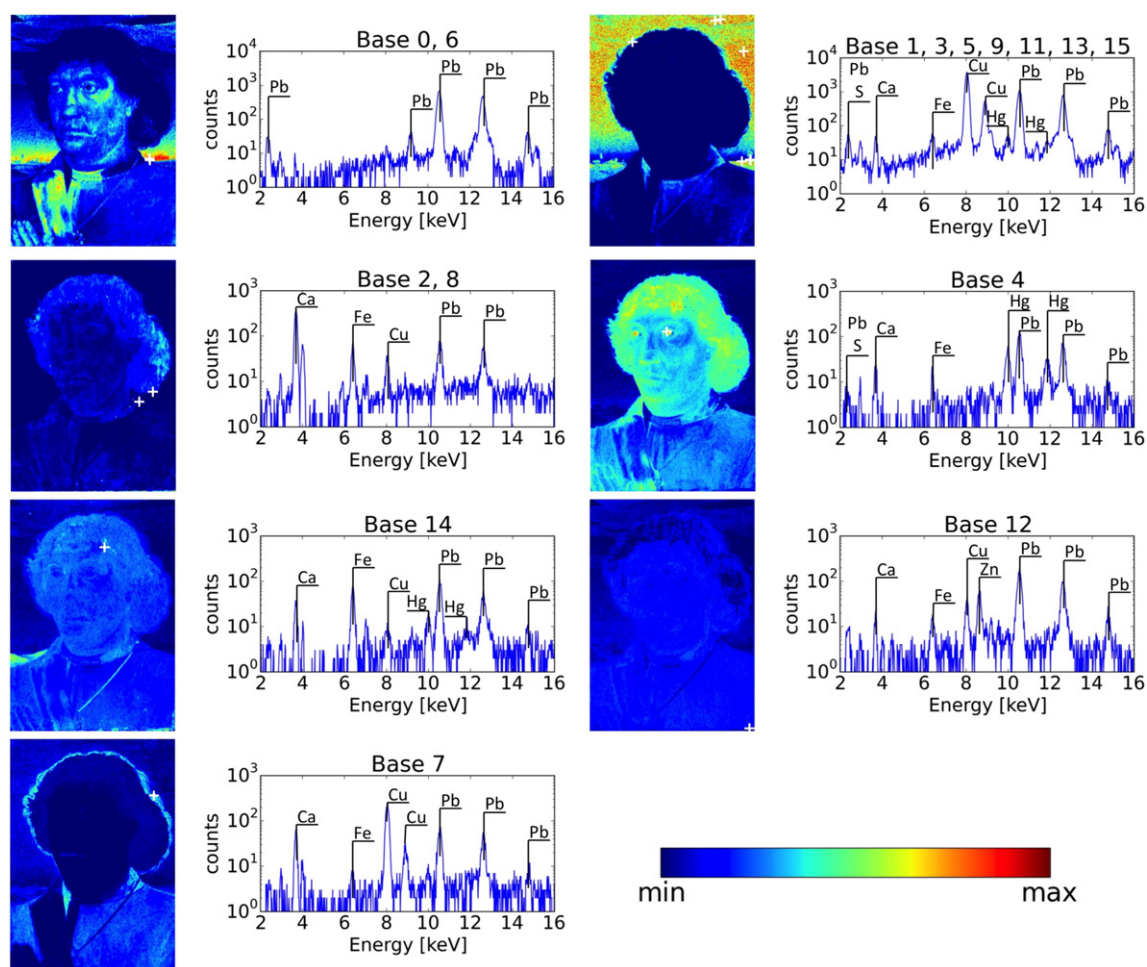


Fig. 5. Representation of the raw spectral data of the *Portrait of a man from the Lespinette family* obtained by Simplex Volume Maximization in 230 s. The crosses indicate the lateral position of the bases found.

of the hair is present on the Cu and Pb containing paint of the sky. This feature was not obvious to the authors, when they initially interpreted the elemental distribution images.

5. Conclusions

We have shown that SiVM allows for the factorization of full spectral XRF data in a short time, compared to the acquisition time. The results obtained are due to the non-negative nature of the coefficient images and the real data points used as bases easily interpretable. The results are not identical, but similar to elemental distribution images and highlight correlations otherwise not obvious. SiVM is thus suitable to support the near real time data evaluation with a minimum of user intervention and due to its low computing demands allows to process data sets of near arbitrary size. That no information of Co is obtained by SiVM and the distribution of Zn could not be shown is regrettable but not surprising as both elements are close to their respective limit of detection. Also, using real data points as bases limits the statistics of the individual base and the redundancy of bases requires a small amount of user intervention to obtain easily readable results.

In other data sets spectral artifacts (e.g. from oversaturating the detector) or artifacts in the sample (like a single Fe containing dust particle on a thin biological sample) were represented by several bases and complicated the interpretation, which could be avoided by masking these pixels and not taking them into account for the calculation of W . Ultimately, one has to be aware that the development of factorization approaches is still ongoing and that improved versions of SiVM are expected to provide more rapidly even easier interpretable results.

Acknowledgements

The German Federal Ministry of Education and Research (BMBF) is acknowledged for the financial support (Verbundprojekt 05K2012 – POISSON: Fortschrittliche Faktorenanalyse für Poisson-verteilte Daten).

References

- [1] F. Adams, Spectroscopic imaging: a spatial Odyssey, *J. Anal. At. Spectrom.* 29 (2014) 1197–1205, <http://dx.doi.org/10.1039/c4ja00050a>.
- [2] M. Alfeld, K. Janssens, Strategies for processing mega-pixel X-ray fluorescence hyperspectral data: a case study on a version of Caravaggio's painting *Supper at Emmaus*, *J. Anal. At. Spectrom.* 30 (2015) 777–789, <http://dx.doi.org/10.1039/C4JA00387J>.
- [3] P. Noble, A. van Loon, M. Alfeld, K. Janssens, J. Dik, Rembrandt and/or studio, *Saul and David*, c. 1655: visualising the curtain using cross-section analyses and X-ray fluorescence imaging, *Technè* 35 (2012) 36–45.
- [4] K. Janssens, G. Van Der Snickt, M. Alfeld, P. Noble, A. van Loon, J.K. Delaney, et al., Rembrandt's "Saul and David" (c. 1652): use of multiple types of smalt evidenced by means of non-destructive imaging, *Microchem. J.* 126 (2016) 515–523, <http://dx.doi.org/10.1016/j.microc.2016.01.013>.
- [5] G. Van der Snickt, S. Legrand, J. Caen, F. Vanmeert, M. Alfeld, K. Janssens, Chemical imaging of stained-glass windows by means of macro X-ray fluorescence (MA-XRF) scanning, *Microchem. J.* 124 (2016) 615–622, <http://dx.doi.org/10.1016/j.microc.2015.10.010>.
- [6] B. Vekemans, K. Janssens, L. Vincze, A. Aerts, F. Adams, J. Hertogen, Automated segmentation of μ -XRF image sets, *X-Ray Spectrom.* 26 (1997) 333–346, [http://dx.doi.org/10.1002/\(SICI\)1097-4539\(199711/12\)26:6<333::AID-XRS231>3.0.CO;2-D](http://dx.doi.org/10.1002/(SICI)1097-4539(199711/12)26:6<333::AID-XRS231>3.0.CO;2-D).
- [7] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* 401 (1999) 788–791, <http://dx.doi.org/10.1038/44565>.
- [8] P.G. Kotula, M.R. Keenan, J.R. Michael, Automated analysis of SEM X-ray spectral images: a powerful new microanalysis tool, *Microsc. Microanal.* 9 (2003) 1–17, <http://dx.doi.org/10.1017/S1431927603030058>.

- [9] A. Kéri, J. Osán, M. Fábrián, R. Dähn, S. Török, Combined X-ray microanalytical study of the Nd uptake capability of argillaceous rocks, *X-Ray Spectrom.* 45 (2016) 54–62, <http://dx.doi.org/10.1002/xrs.2656>.
- [10] J. Osán, A. Kéri, D. Breitner, M. Fábrián, R. Dähn, R. Simon, et al., Microscale analysis of metal uptake by argillaceous rocks using positive matrix factorization of microscopic X-ray fluorescence elemental maps, *Spectrochim. Acta Part B At. Spectrosc.* 91 (2014) 12–23, <http://dx.doi.org/10.1016/j.sab.2013.11.002>.
- [11] M. Alfeld, M. Wahabzada, C. Bauckhage, K. Kersting, G. Wellenreuther, G. Falkenberg, Non-negative factor analysis supporting the interpretation of elemental distribution images acquired by XRF, *J. Phys. Conf. Ser.* 499 (2014) 12013, <http://dx.doi.org/10.1088/1742-6596/499/1/012013>.
- [12] M. Alfeld, M. Wahabzada, C. Bauckhage, K. Kersting, G. Wellenreuther, P. Barriobero-Vila, et al., Non-negative matrix factorization for the near real-time interpretation of absorption effects in elemental distribution images acquired by X-ray fluorescence imaging, *J. Synchrotron Radiat.* 23 (2016) 579–589, <http://dx.doi.org/10.1107/S1600577515023528>.
- [13] V.A. Solé, E. Papillon, M. Cotte, P. Walter, J. Susini, A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra, *Spectrochim. Acta B At. Spectrosc.* 62 (2007) 63–68, <http://dx.doi.org/10.1016/j.sab.2006.12.002>.
- [14] H.C. Santos, C. Caliri, L. Pappalardo, R. Catalano, A. Orlando, F. Rizzo, et al., Identification of forgeries in historical enamels by combining the non-destructive scanning XRF imaging and alpha-PIXE portable techniques, *Microchem. J.* 124 (2016) 241–246, <http://dx.doi.org/10.1016/j.microc.2015.08.025>.
- [15] A. Martins, C. Albertson, C. McGlinchey, J. Dik, Piet Mondrian's Broadway Boogie Woogie: non-invasive analysis using macro X-ray fluorescence mapping (MA-XRF) and multivariate curve resolution-alternating least square (MCR-ALS), *Herit. Sci.* 4 (2016) 22, <http://dx.doi.org/10.1186/s40494-016-0091-4>.
- [16] A. Martins, J. Coddington, G. Van der Snickt, B. van Driel, C. McGlinchey, D. Dahlberg, et al., Jackson Pollock's Number 1A, 1948: a non-invasive study using macro-x-ray fluorescence mapping (MA-XRF) and multivariate curve resolution-alternating least squares (MCR-ALS) analysis, *Herit. Sci.* 4 (2016) 33, <http://dx.doi.org/10.1186/s40494-016-0105-2>.
- [17] A. Malik, A. de Juan, R. Tauler, Multivariate curve resolution: a different way to examine chemical data, 40 Years Chemom. – From Bruce Kowalski to Futur 2015, pp. 95–128, <http://dx.doi.org/10.1021/bk-2015-1199.ch005>.
- [18] A. Cutler, L. Breiman, Archetypal analysis, *Technometrics* 36 (1994) 338–347, <http://dx.doi.org/10.1080/00401706.1994.10485840>.
- [19] C. Thureau, K. Kersting, M. Wahabzada, C. Bauckhage, Descriptive matrix factorization for sustainability adopting the principle of opposites, *Data Min. Knowl. Disc.* 24 (2012) 325–354, <http://dx.doi.org/10.1007/s10618-011-0216-z>.
- [20] code.google.com/p/pymf/
- [21] M. Alfeld, J.V. Pedroso, M. van Eikema Hommes, G. der Snickt, G. Tauber, J. Blaas, et al., A mobile instrument for in situ scanning macro-XRF investigation of historical paintings, *J. Anal. At. Spectrom.* 28 (2013) 760–767, <http://dx.doi.org/10.1039/C3JA30341A>.
- [22] D.S. Bright, D.E. Newbury, Maximum pixel spectrum: a new tool for detecting and recovering rare, unanticipated features from spectrum image data cubes, *J. Microsc.* 216 (2004) 186–193, <http://dx.doi.org/10.1111/j.0022-2720.2004.01412.x>.