

# MANTiS: a program for the analysis of X-ray spectromicroscopy data

Mirna Lerotic,<sup>a\*</sup> Rachel Mak,<sup>b</sup> Sue Wirick,<sup>c</sup> Florian Meirer<sup>d</sup> and Chris Jacobsen<sup>e,b,f</sup>

Received 31 January 2014

Accepted 13 June 2014

<sup>a</sup>2nd Look Consulting, Room 1702, 17/F, Tung Hip Commercial Building, 248 Des Voeux Road, Hong Kong, <sup>b</sup>Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3112, USA, <sup>c</sup>Center for Advanced Radiation Sources, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA, <sup>d</sup>Inorganic Chemistry and Catalysis, Utrecht University, Universiteitsweg 99, 3584 CG Utrecht, The Netherlands, <sup>e</sup>Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439-4837, USA, and <sup>f</sup>Chemistry of Life Processes Institute, Northwestern University, 2170 Campus Drive, Evanston, IL 60208, USA. \*E-mail: mirna@2ndlookconsulting.com

Spectromicroscopy combines spectral data with microscopy, where typical datasets consist of a stack of images taken across a range of energies over a microscopic region of the sample. Manual analysis of these complex datasets can be time-consuming, and can miss the important traits in the data. With this in mind we have developed *MANTiS*, an open-source tool developed in Python for spectromicroscopy data analysis. The backbone of the package involves principal component analysis and cluster analysis, classifying pixels according to spectral similarity. Our goal is to provide a data analysis tool which is comprehensive, yet intuitive and easy to use. *MANTiS* is designed to lead the user through the analysis using story boards that describe each step in detail so that both experienced users and beginners are able to analyze their own data independently. These capabilities are illustrated through analysis of hard X-ray imaging of iron in Roman ceramics, and soft X-ray imaging of a malaria-infected red blood cell.

© 2014 International Union of Crystallography

**Keywords:** X-ray microscopy; X-ray spectromicroscopy; XANES; NEXAFS.

## 1. Introduction

Imaging shows how heterogeneous materials are organized, while spectroscopy reveals the characteristics of the various parts. Like spectrum imaging in electron microscopy (Jeanguillaume & Colliex, 1989; Hunt & Williams, 1991), X-ray spectromicroscopy (Ade *et al.*, 1992; Zhang *et al.*, 1994; De Stasio *et al.*, 1993; Jacobsen *et al.*, 2000) involves the collection of images across an X-ray absorption near-edge structure (XANES) or near-edge X-ray absorption fine-structure (NEXAFS) region in order to see the distribution of different chemical states. X-ray spectromicroscopy thus delivers rich datasets consisting of a separate spectrum at each pixel of an image. Because one can easily obtain  $10^4$ – $10^6$  spectra each consisting of  $\sim 10^2$  points, by-hand examination of all spectra is impractical; instead, it is very helpful to find similarities between spectra and use these to simplify data analysis.

We describe here a computer program *MANTiS* (for Multivariate ANalysis Tool for Spectromicroscopy) for analysis of X-ray spectromicroscopy data. This program draws upon methods previously developed for data classification including principal component analysis (King *et al.*, 1989; Wasserman, 1997; Osanna & Jacobsen, 2000) and cluster analysis (Jacobsen *et al.*, 2003; Lerotic *et al.*, 2004, 2005), and it will soon have an implementation of non-negative matrix approximation analysis (Lee & Seung, 1999; Fleckenstein, 2008). *MANTiS* draws upon earlier programs (*stack\_analyze.pro* and *pca\_gui.pro*) for spectromicroscopy developed by our team when at Stony Brook University (Jacobsen *et al.*, 2000; Lerotic *et al.*, 2004), but

it provides new capabilities and is now implemented as an open-source Python project (see §4) so that the mathematical actions it employs for data analysis can be examined and verified. We note that *MANTiS* takes its place alongside other programs written for soft X-ray spectromicroscopy analysis. The program *aXis2000* from Peter and Adam Hitchcock (<http://unicorn.mcmaster.ca/aXis2000.html>) provides analysis capabilities centered on the use of reference spectra, or user-identified spectral regions; it does not provide cluster analysis except *via* a link to an older compiled version of *pca\_gui.pro*. The Python program *STXMPy* by Haraszti *et al.* (2010) provides some script-based analysis tools, but it has not been updated since 2010. For hard X-ray transmission X-ray microscopy data analysis, the *TXM-XANES Wizard* program within the *TXM-Wizard* package of Liu *et al.* (2012) provides some cluster analysis capabilities; a Matlab-compiled version is available. The program *PyMca* (Solé *et al.*, 2007) provides a wide range of options and analysis approaches for X-ray fluorescence microscopy, plus tools for XANES analysis, with source code available. Amongst these programs, *MANTiS* emphasizes simplicity of program operation for advanced analysis, as an alternative to the rich but complicated set of options available in some of the above programs.

## 2. Analysis workflow with MANTiS

In X-ray spectromicroscopy, one usually acquires a series of transmission images at pixel locations ( $x, y$ ) over a series of photon

energies  $E$ , or  $I(x, y, E)$ . We collect spatial dimensions together into a single index  $p = 1 \dots P$ , which steps through the pixels in  $(x, y)$  and which of course can be extended to handle three-dimensional data; we use  $n = 1 \dots N$  to index the energies. Because X-rays are attenuated as  $I = I_0 \exp(-\mu t)$  where  $\mu$  is a linear absorption coefficient and  $t$  is a thickness, we wish to work with the optical density

$$D = \mu t = -\ln \left[ \frac{I(x, y, E)}{I_0(E)} \right] \quad (1)$$

which is linear in specimen thickness. We can thus convert the set of transmission images  $I(x, y, E)$  into an optical density matrix  $D_{N \times P}$ . Our expectation is that the specimen can be considered to be comprised of a set  $s = 1 \dots S$  of materials with unique spectroscopic responses, or components; we therefore hope to represent the measured data  $D_{N \times P}$  as

$$D_{N \times P} = \mu_{N \times S} \cdot t_{S \times P} \quad (2)$$

where  $D_{N \times P}$  represents the measured optical density as a spectrum per pixel,  $\mu_{N \times S}$  is the set of spectra for each of the  $S$  components, and  $t_{S \times P}$  is the set of thickness or weighting maps for each of the  $S$  components. Our goal in analysis is to find a solution for  $\mu_{N \times S}$  and  $t_{S \times P}$  that adequately represents the measured data  $D_{N \times P}$ . One approach (Lerotic *et al.*, 2004) is to use principal component analysis to find a reduced set of eigenspectra, followed by cluster analysis to group spectroscopically similar pixels together, followed by averaging the spectra of those pixels together to yield a set of spectra  $\mu_{N \times S}$  and then calculate the corresponding thickness maps from

$$t_{S \times P} = \mu_{S \times N}^{-1} \cdot D_{N \times P} \quad (3)$$

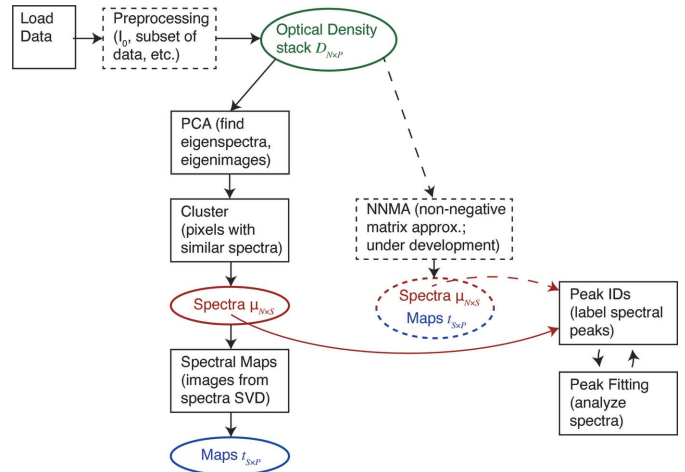
Another approach is to use non-negative matrix approximation methods to jointly deliver solutions for  $\mu_{N \times S}$  and  $t_{S \times P}$  (Lee & Seung, 1999), a capability that is under development for *MANTiS* and will be described in more detail elsewhere.

*MANTiS* is designed around a series of ‘story boards’ that walk a user of the program through a sequence of required analysis steps. The overall workflow is shown schematically in Fig. 1, and we now outline the various steps involved.

### 2.1. Data preprocessing

The first step in data analysis is to preprocess the data to yield an optical density matrix  $D_{N \times P}$ . This is accomplished using a ‘Preprocessing’ story board in *MANTiS* (Fig. 2).

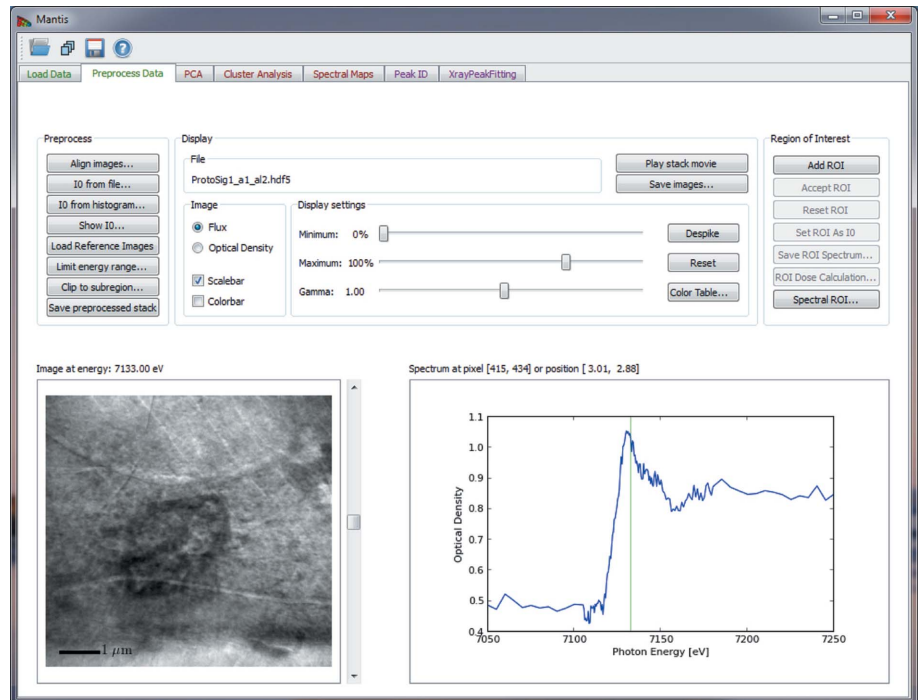
One begins by making sure that all the images are properly registered to each other. When diffractive and refractive X-ray optics (which are chromatic in nature) are used for focusing, the optic-to-specimen distance must be changed as one acquires images over a spectral region. Due to runout errors in mechanical positioning stages (or figure errors or misalignment of the mirrors used in interferometric feedback systems), this often means that the images are shifted relative to each other in  $(x, y)$  and therefore must be registered onto a common axis. While *MANTiS* provides a tool for manual alignment of image shifts by clicking on an



**Figure 1** Schematic of the workflow through which the user is guided in *MANTiS*. Dashed lines show optional tools (Preprocessing; §2.1) and tools under development [non-negative matrix approximation analysis (Lee & Seung, 1999)].

identifiable feature in all images, in most cases the alignment can be done in an automated fashion using cross correlation (Dierksen *et al.*, 1993; Koster *et al.*, 1992; Jacobsen *et al.*, 2000). *MANTiS* allows one to select a particularly distinctive region for cross-correlation analysis, as well as a display of the cross-correlation alignment as it proceeds [replicating the functionality described by Jacobsen *et al.* (2000)]. It also allows one to subsequently reduce the field of view to that present across all images, since otherwise (due to image shifts) some pixels may not have information at certain photon energies.

After the images have been aligned, it is necessary to go from the measured transmission images  $I(x, y, E)$  to the optical density matrix  $D_{N \times P}$  [equation (1)]. This requires division by the incident spectrum



**Figure 2** *MANTiS* program layout. The user is guided through the sequence of operations shown in Fig. 1 using story board tabs. Shown here is the ‘Preprocessing’ tab with access to settings on transmission image normalization and image sequence alignment.

$I_0(E)$ . If there are any fully transmitting regions in the specimen (such as from holes or specimen-free regions), one can recover the incident flux spectrum by plotting a histogram of all intensities, and then selecting the pixels at the maximum intensity end of the histogram.

Alternatively, one can read in an incident flux  $I_0(E)$  spectrum obtained by some other means, in which case *MANTiS* uses cubic interpolation to re-map the spectrum onto the same energy points as the transmission images  $I(x, y, E)$ . The incident flux spectrum  $I_0(E)$  can also be defined by selecting a region on the transmission image using the ‘Region of Interest’ (ROI) tool.

Especially when using automated data collection, one sometimes finds that certain images have artifacts such as a change in flux due to storage-ring orbit errors, or alterations in the flow of helium gas used to minimize air absorption. *MANTiS* thus provides for a way to easily scan through the set of images and remove any artifact which might otherwise confound subsequent analysis steps, as well as limiting the photon energy range to concentrate on the spectroscopically most interesting images.

## 2.2. Elemental maps

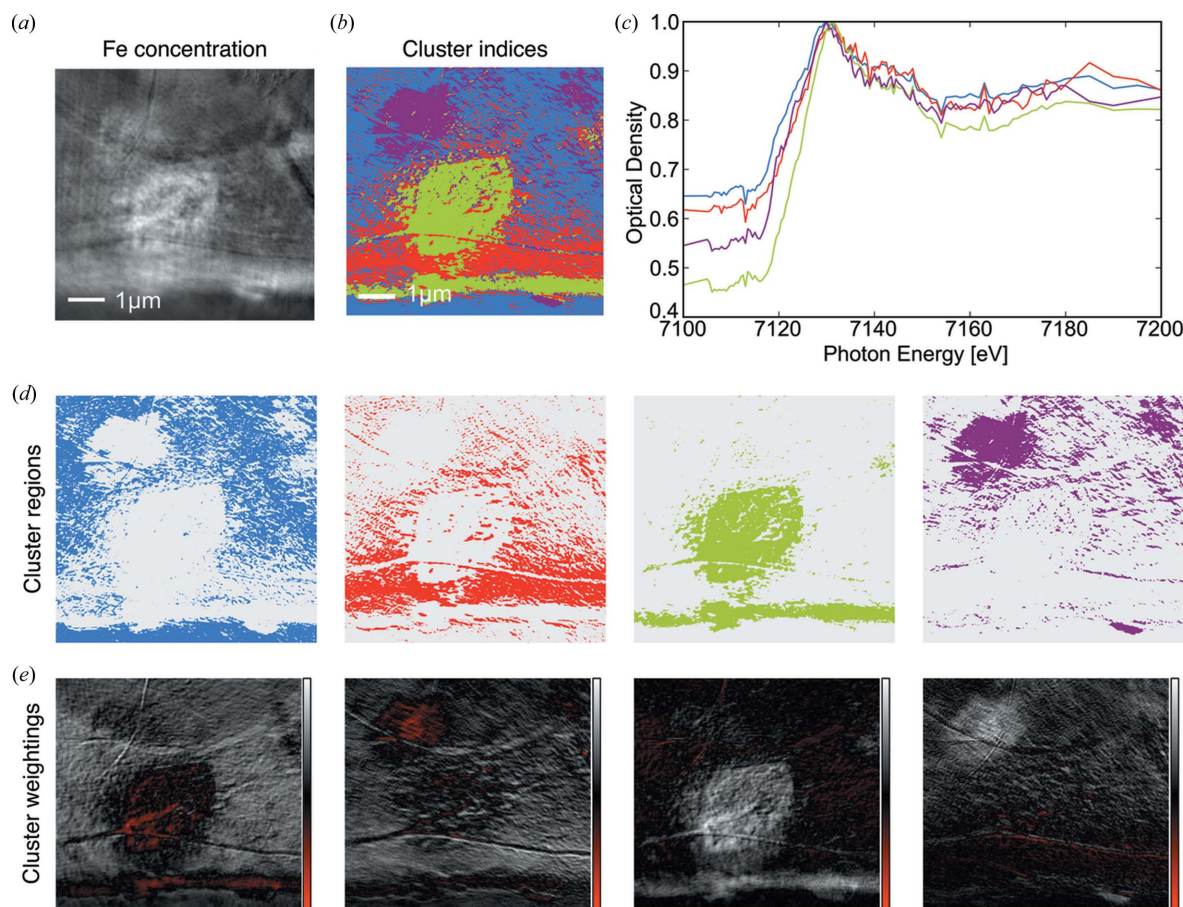
*MANTiS* also provides a way to obtain a concentration map of a certain element by averaging together a set of images above and below an X-ray absorption edge (Engström, 1946). Because many

images are averaged together, the resulting elemental maps often have significantly improved signal to noise relative to maps calculated from just two images. An example of this is shown in Fig. 3(a).

## 2.3. PCA and cluster analysis

One approach to obtaining a set of spectra  $\mu_{N \times S}$  and thickness maps  $t_{S \times P}$  that describe the data  $D_{N \times P}$  is to use cluster analysis (Jacobsen *et al.*, 2003; Lerotic *et al.*, 2004, 2005). The first step in this approach is to use principal component analysis (PCA) to obtain a set of  $S_{\text{abstract}}$  eigenimages and eigenspectra from the data’s covariance matrix, where  $S_{\text{abstract}}$  is the smaller of  $N$  or  $P$  (usually  $N$  in practice). While one can represent any spectrum or image in the data  $D_{N \times P}$  by a linear combination of these eigenvectors, one usually finds that a very small subset  $\bar{S}_{\text{abstract}}$  is sufficient to represent all the non-noise variations in the data (Jolliffe, 2002). *MANTiS* provides a way to view the positive and negative values of the eigenspectra and eigenimages (in the latter case by representing positive values in shades of white and negative values in shades of red) so as to determine the significant number of components  $\bar{S}_{\text{abstract}}$  (hereafter simply referred to as  $S$ ) to use in subsequent analysis, as shown by Lerotic *et al.* (2004).

PCA provides a limited and nicely orthogonalized search space for finding pixels with similar spectral response. This is carried out in the ‘Cluster Analysis’ story board in the analysis sequence. When



**Figure 3**

The ‘Cluster Analysis’ story board was used to examine iron *K*-edge XANES X-ray spectromicroscopy data of a Roman ceramics sample (§3.1) into four clusters. We show the iron concentration map obtained (§2.2) from comparison of the 7100–7160 eV pre-edge and 7155–7175 eV post-edge spectral regions (a), the color-coded cluster index map of which cluster spectrum is predominant at which pixel (b), and the color-coded cluster spectra (c). The cluster spectra were then used to calculate composition maps which give thickness or weighting maps (d) for each of the cluster spectra; red regions represent negative weightings which can occur with cluster analysis, and which are addressed by using non-negative matrix approximation methods that will soon be integrated into *MANTiS*. The cluster thickness maps (e) show a combination of differing iron concentrations (ratios of post-edge to pre-edge absorption), and different iron phases as described in §3.1.



searching for clusters, one is looking for pixels with similar weightings of eigenspectra; this is done by  $k$ -means clustering (MacQueen, 1967) using Euclidian distances in eigenvector space. The number of clusters to be sought is user-adjustable; generally speaking, it is suggested to search for several more clusters than the number  $S$  of significant components. Sensitivity of cluster analysis in finding hidden traits in the data is improved by subtracting offsets from the eigenimages. This feature is an integral part of cluster analysis and it is automatically applied. Furthermore, we introduced a power-law weighting that can be used to give more importance to less significant components which could describe subtle variations in the spectra or represent a feature captured by only a few pixels. The amount of weighting can be adjusted through the 'PC scaling factor' on the 'Cluster Analysis' story board.

In some cases, it is useful to exclude the first component from the cluster search, so as to reduce sensitivity to thickness variations; this is because the first component represents the average absorption spectrum present in the dataset, while subsequent components represent variations from that average in decreasing significance (Lerotic *et al.*, 2004). In addition,  $k$ -means clustering sometimes produces clusters where one or a few pixels with unique spectroscopic signatures are incorrectly grouped with a large number of more self-similar pixels; a 'Divide clusters with large Sigma' option will look for the cluster with the biggest variance of Euclidian distances from the cluster center, and break it into two clusters.

Once clustering has been accomplished, one obtains a set of spectra  $\mu_{N \times S}$  for each cluster by calculating an average of measured spectra of all the pixels assigned to that cluster, and a color-coded map of the clusters by showing which cluster group each pixel belongs to (Fig. 3). One can optionally also view cluster-color-coded scatterplots, which plot pixels by their weightings in any two eigenspectra and which show more explicitly how the data were segmented into different clusters (Lerotic *et al.*, 2004).

#### 2.4. Composition maps from cluster analysis

Once one has a set of representative spectra  $\mu_{N \times S}$  from cluster analysis, one can calculate a pseudo-inverse and multiply it by the optical density data to obtain images of the weightings or thicknesses of these spectra, represented by the  $t_{S \times P}$  matrix as described in equation (2). The pseudo-inverse can be calculated using singular value decomposition (Zhang *et al.*, 1996; Koprinarov *et al.*, 2002) as described in Appendix A of Lerotic *et al.* (2004). Alternatively, if one has obtained an appropriate set of spectra by some other means (*e.g.* spectral standards for samples of known composition, or spectra of hand-selected regions of interest, or non-negative matrix analysis), one can use these spectra to form  $\mu_{N \times S}$ . In *MANTI*S, the  $\mu_{N \times S}$  matrix is referred to as a set of Target Spectra within the 'Spectral Maps' tab, and these Target Spectra can be loaded using either 'Add Cluster Spectra' or by using the 'Load Spectrum' button to load one or more individual spectra obtained by some other means. It should be noted that the absolute calibration of the thicknesses relies on the proper calibration of the linear absorption coefficients in the spectra  $\mu_{N \times S}$ .

Representative thickness maps and spectra can be used to generate a composition map as shown in Fig. 4. Composition

maps are generated by assigning red, green and blue colors to three thickness maps and combining them. These maps aid us in observing how different spectral regions overlap spatially.

#### 2.5. Spectral peak fitting and peak identification

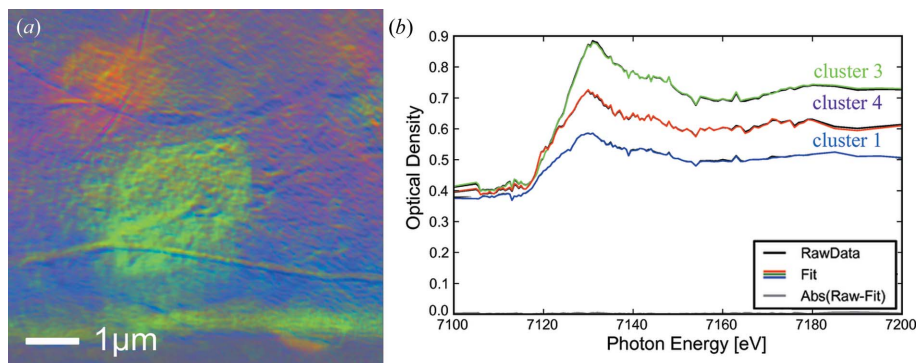
A set of representative spectra, either from cluster analysis or reference spectra loaded from a file, can be further examined using the 'Peak ID' and 'XrayPeakFitting' tabs. In the 'Peak ID' tab, peaks in a spectrum are compared with a set of known peaks positions, and if a match is found that peak can be attributed to the corresponding chemical bond. Precise peak positions in a spectrum can be found using the 'XrayPeakFitting' tab. Spectra are fitted using a number of Gaussian curves and a step function that describes element edges found in X-ray spectra. Examples of fitted spectra are given in Fig. 6.

### 3. Analysis examples

The predecessor programs to *MANTI*S (*stack\_analyze.pro* and *pca\_gui.pro*) have been successfully applied in several studies in soil science (Lehmann *et al.*, 2005, 2007, 2008; Kinyangi *et al.*, 2006) as well as in environmental science (Schäfer *et al.*, 2005, 2009; Bauer *et al.*, 2005; Claret *et al.*, 2005). In a study of free microaggregates by Lehmann *et al.* (2008), cluster analysis was used to segregate regions with distinct organic matter forms; this showed that the regions were significantly different on a very fine spatial scale and that none of these regions was similar to the organic matter of the entire soil. The program *MANTI*S adds additional analysis capabilities and provides an improved user interface as described in §2, so we describe here its application to two separate experimental datasets.

#### 3.1. Roman ceramics

Different ceramic firing conditions used in antiquity can be identified by the presence of different iron oxide minerals. Thin (40–60  $\mu\text{m}$ ) sections of early Roman ceramics were prepared and studied by Meirer *et al.* (2013) using a transmission X-ray microscope at SSRL. One spectromicroscopy data set in that study was challenging in its analysis due to a high degree of uncorrelated noise in the image sequence. *MANTI*S was used to align the original images using both manual and automated cross-correlation approaches. The 'Limit energy range' option was then used to reduce the energy range to a spectroscopically interesting region from 7100 eV to 7200 eV, and PCA was performed. Out of the four most significant principal components, the third one was found to contain mainly noise so it was



**Figure 4**

(a) Color composite view of a Roman ceramics sample (§3.1 and Fig. 3) obtained by showing three different thickness or weighting maps as intensities of red, green and blue colors in a composite display. Spectra in (b) correspond to different iron oxide minerals present in the data.

## computer programs

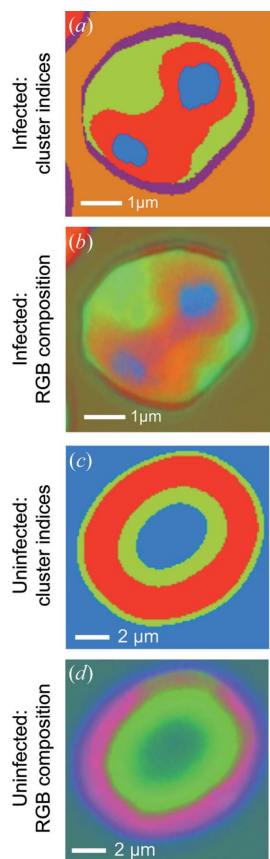
removed from further analysis. Next, cluster analysis was used to classify the pixels according to four distinct spectroscopic signatures [Figs. 3(b) and 3(c)], and the resulting cluster spectra were used to calculate the thickness maps shown in Fig. 3(e) as well as the color composite image shown in Fig. 4. In these maps and composite image, the three spectra correspond to different iron oxide minerals: hematite ( $\alpha$ -Fe<sub>2</sub>O<sub>3</sub>), maghemite ( $\gamma$ -Fe<sub>2</sub>O<sub>3</sub>) and hercynite (FeAl<sub>2</sub>O<sub>4</sub>), respectively. This example shows how *MANTI*S can be used for the analysis of hard X-ray spectromicroscopy data.

### 3.2. Malarial-infected erythrocytes

Malaria is a disease caused by the infection of erythrocytes (red blood cells) by one of several *Plasmodium* parasites carried by the *Anopheles* mosquito. *MANTI*S was used to study human erythrocytes infected with *P. falciparum*, fixed with a solution of 0.1 M sodium cacodylate containing 2.5% glutaraldehyde, and air-dried on silicon nitride windows; these samples were prepared by E. Hanssen alongside those used in a soft X-ray cryo-tomography study (Hanssen *et al.*, 2012). Carbon *K*-edge and iron *L*-edge XANES spectromicroscopy data were acquired using the scanning transmission X-ray microscope formerly located at beamline X1A1 at the National Synchrotron Light Source (NSLS). Data from the energy range 282–295 eV were segmented into five clusters using either the first five PCA components and a PCA scaling factor of 0.4 (infected erythrocyte), or three clusters from the first four PCA components and a PCA scaling factor of 0.2 (uninfected erythrocytes). The scaling factor as given by equation (19) of Lerotic *et al.* (2004) may be used to give greater or lesser weighting to the less significant PCA components that capture increasingly subtle variations in the spectral signatures of the data, or spectral components that contribute only to a few pixels.

Compared with the uninfected cell shown in Fig. 5(c), the infected cell shows the well known departure in shape, and the parasites are identified both by a carbon XANES cluster-analysis-selected region (Fig. 5a) and by a ratioed above-and-below-edge Fe *L*-edge image pair (Fig. 5b). The carbon XANES spectra shown in Fig. 5(e) include a spectrum from the parasite region, and this spectrum is plotted and its primary resonant peaks are fitted (§2.5) and labeled in the top part of Fig. 6(d). For the uninfected cell, the spectrum of the dominant absorptive region [with a cluster index color of red in Fig. 5(f)] is shown along with its fitted primary resonant peaks in the bottom part of Fig. 6(d).

In the uninfected erythrocyte, the 284.98 eV absorption peak is from the C=C bond found in the porphyrin ring of hemoglobin. The absorption peak at 288.0 eV is from the C=O bond that hangs off of the porphyrin ring in hemoglobin. The malaria parasite digests hemoglobin and the end by-product of this digestion in a multiple-step process is hemozoin. The process of digestion changes the valence state of Fe from Fe<sup>2+</sup> to Fe<sup>3+</sup>. The Fe<sup>3+</sup> ion located in the center of the porphyrin ring has a higher



**Figure 5** Cluster analysis maps with color coding of cluster indices (a, c) and three-color composition maps (b, d) for infected (a, b) and uninfected (c, d) erythrocytes with cluster spectra on (e) and (f), respectively. The composite map of the infected cell (b) was calculated using the red, green and blue colored cluster spectra shown in (e). The composite map of the uninfected cell (d) was generated using the spectra in (f).

**Table 1**

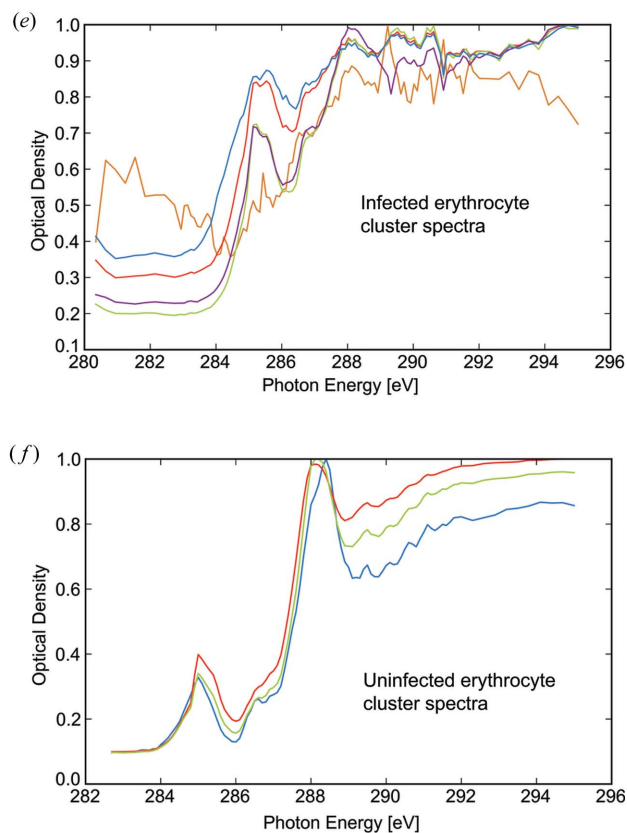
Spectromicroscopy file formats presently supported by *MANTI*S (additional formats can be accommodated upon request).

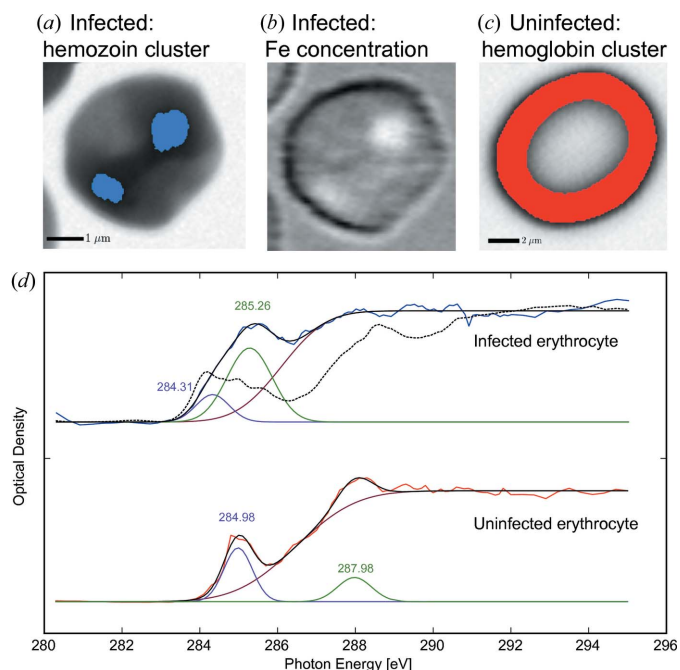
| File format                               | Extension      | Institution  |
|---|----------------|--|
| DataExchange                              | .hdf5 or .h5   | Argonne National Laboratory                                      |
| SDF                                       | .hdr           | Lawrence Berkeley National Laboratory and Paul Scherrer Institut |
| XDR                                       | .stk           | Former Stony Brook STXM at NSLS                                  |
| OLE Compound                              | .xrm and .txrm | Carl Zeiss X-ray Microscopy                                      |
| Multi-Page TIFF, or sequential TIFF files | .tiff          | Adobe Systems  |

local symmetry than that found in hemoglobin and it is this higher symmetry that likely results in multiple absorption peaks at both higher and lower energies around the 285 eV peak similar to the symmetry peaks one finds in molecules such as decoronene and benzopyrene (Butykai *et al.*, 2013; Stöhr, 1992; de Groot, 2005; Tzvetkov *et al.*, 2007; Desiraju & Gavezotti, 1989).

### 4. Programming details and future plans

*MANTI*S is able to read spectromicroscopy data presented in several different file formats listed in Table 1, and additional file formats can be accommodated upon request. The primary output file format is the DataExchange (<http://www.aps.anl.gov/DataExchange/>) schema developed at Argonne National Laboratory for writing data and metadata using the Hierarchical Data Format 5 or HDF5 (The HDF





**Figure 6** Carbon near-edge spectromicroscopy analysis of a normal erythrocyte (c), and one infected with the *P. falciparum* malaria parasite (a, b). A region identified by cluster analysis on the carbon XANES images (a) agrees well with an approximate map of iron concentration taken from the ratio of images acquired above and below the Fe L-edge (b). Spectra of two cluster regions of infected (a) and uninfected (c) cells were fitted using the ‘X-ray Peak Fitting’ tool of *MANTI*s; the energies of the peak centers are indicated on the spectral plots. A map of Fe concentration in the infected cell (b) was generated using the ‘Spectral ROI’ tool to select a 702.23–704.76 eV energy range below the Fe edge and a 709.48–711.98 eV range above the edge; a ratioed image then gives the elemental concentration (Engström, 1946). A spectrum of hematin, a substitute for hemozoin (Walczak *et al.*, 2005), is shown with a dashed spectrum. The peak at 284.31 eV in the infected cell can be attributed to a 284 eV peak in the hematin spectrum. The overlap between areas rich in elemental iron with areas rich in hemozoin confirms the strong concentration of hemozoin precipitated by the parasites.

Group, 2000–2010). *MANTI*s can also work with other file formats as shown in Table 1.

Image stacks can also be built from a set of individual images. The supported single-energy file formats are .sm NetCDF files that follow a format developed by Stony Brook University and formerly operated at beamline X1A at NSLS at Brookhaven, and .xrm OLE Compound single-energy files, in a format developed by Carl Zeiss X-ray Microscopy.

The first story board used to load data can also be used as a data format converter into a readable HDF5 format (DataExchange). The output DataExchange HDF5 files saved by *MANTI*s include several calculational results such as the optical density matrix  $D_{N \times P}$  [equation (1)] calculated using the preprocessing story board. If the ‘PCA’ story board has been invoked, the file contains the eigenvalues, eigenimages and eigenspectra, while, if the ‘Cluster’ story board has been invoked, the per-pixel cluster indices, cluster spectra and thickness maps, as well as a resulting thickness error map are all added to the output file. Besides saving the analysis to an HDF5 file, results from various steps can be saved as bitmap images in the .png format, or as vector graphics files in the .svg or .pdf format (allowing for later editing of text in fonts using programs like Inkscape or Adobe Illustrator).

*MANTI*s is an open-source program written in Python using PyQt for its graphical user interface, yet it can also be run in batch mode with command line options directing the workflow. Working with

Python and Qt ensures cross-platform compatibility. *MANTI*s takes advantage of Python’s capabilities for program modularity. Each analysis step, such as PCA or cluster analysis, is separated from the graphical user interface (which is why the same code can be run in batch mode), and its data input/output routines can be readily adapted for use by another scientific package. Full *MANTI*s code and a ‘Wiki’-style user guide are available on the Google code hosting site at <http://code.google.com/p/spectromicroscopy/>. The code is licensed under GNU General Public License version 3.

To make the program even more available to users, our future plans include Web-based analysis where the server reads the data and performs all calculations. This will allow users with less powerful computers to still analyze large spectromicroscopy data sets, since one could access both the data and the *MANTI*s program on a larger centralized computer and data storage system.

Initial development of *MANTI*s was supported by the US Department of Energy (DOE) Office of Science, under contract DE-AC02-06CH11357 to Argonne National Laboratory. The authors would like to thank Philippe Sciau (University of Toulouse) and Joy C. Andrews (Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory) for providing the Roman ceramics data described in §3.1, as well as Eric Hanssen (University of Melbourne) for the malaria data described in §3.2.

## References

- Ade, H., Zhang, X., Cameron, S., Costello, C., Kirz, J. & Williams, S. (1992). *Science*, **258**, 972–975.
- Bauer, A., Rabung, T., Schäfer, T., Buckau, G. & Fanghänel, T. (2005). *Appl. Clay Sci.* **30**, 1–10.
- Butykai, A., Orbán, A., Kocsis, V., Szaller, D., Bordács, S., Tátrai-Szekeres, E., Kiss, L. F., Bóta, A., Vértessy, B. G., Zelles, T. & Kézsmárki, I. (2013). *Nat. Sci. Rep.* **3**, 1431.
- Claret, F., Schäfer, T., Rabung, T., Wolf, M., Bauer, A. & Buckau, G. (2005). *Appl. Geochem.* **20**, 1158–1168.
- Desiraju, G. R. & Gavezzotti, A. (1989). *Acta Cryst.* **B45**, 473–482.
- De Stasio, G., Hardcastle, S., Koranda, S. F., Tonner, B. P., Mercanti, D., Teresa Ciotti, M., Perfetti, P. & Margaritondo, G. (1993). *Phys. Rev. E*, **47**, 2117–2121.
- Dierksen, K., Typke, D., Hegerl, R. & Baumeister, W. (1993). *Ultramicroscopy*, **49**, 109–120.
- Engström, A. (1946). *Acta Radiol. Suppl.* **63**, 1–106.
- Fleckenstein, H. (2008). PhD thesis, Department of Physics and Astronomy, Stony Brook University, USA.
- Groot, F. de (2005). *Coord. Chem. Rev.* **249**, 31–63.
- Hanssen, E., Knoechel, C., Dearnley, M., Dixon, M. W., Le Gros, M., Larabell, C. & Tilley, L. (2012). *J. Struct. Biol.* **177**, 224–232.
- Haraszti, T., Grunze, M. & Anderson, M. G. (2010). *Chem. Cent. J.* **4**, 1–8.
- Hunt, J. & Williams, D. (1991). *Ultramicroscopy*, **38**, 47–73.
- Jacobsen, C., Feser, M., Lerotic, M., Vogt, S., Maser, J. & Schäfer, T. (2003). *J. Phys. IV (Paris)*, **104**, 623–626.
- Jacobsen, C., Wirick, S., Flynn, G. & Zimba, C. (2000). *J. Microsc.* **197**, 173–184.
- Jeanguillaume, C. & Colliex, C. (1989). *Ultramicroscopy*, **28**, 252–257.
- Jolliffe, I. T. (2002). *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag.
- King, P. L., Browning, R., Pianetta, P., Lindau, I., Keenlyside, M. & Knapp, G. (1989). *J. Vac. Sci. Technol. A*, **7**, 3301–3304.
- Kinyangi, J., Solomon, D., Liang, B., Lerotic, M., Wirick, S. & Lehmann, J. (2006). *Soil Sci. Soc. Am. J.* **70**, 1708–1718.
- Koprinarov, I. N., Hitchcock, A. P., McCrory, C. T. & Childs, R. F. (2002). *J. Phys. Chem. B*, **106**, 5358–5364.
- Koster, A. J., Chen, H., Sedat, J. W. & Agard, D. A. (1992). *Ultramicroscopy*, **46**, 207–227.
- Lee, D. D. & Seung, H. S. (1999). *Nature (London)*, **401**, 788–791.
- Lehmann, J., Kinyangi, J. & Solomon, D. (2007). *Biogeochemistry*, **85**, 45–57.
- Lehmann, J., Liang, B., Solomon, D., Lerotic, M., Luizão, F., Kinyangi, J., Schäfer, T., Wirick, S. & Jacobsen, C. (2005). *Global Biogeochem. Cycles*, **19**, 1013–1025.

- Lehmann, J., Solomon, D., Kinyangi, J., Dathe, L., Wirick, S. & Jacobsen, C. (2008). *Nat. Geosci.* **1**, 238–242.
- Lerotic, M., Jacobsen, C., Gillow, J., Francis, A., Wirick, S., Vogt, S. & Maser, J. (2005). *J. Electron Spectrosc. Relat. Phenom.* **144–147**, 1137–1143.
- Lerotic, M., Jacobsen, C., Schäfer, T. & Vogt, S. (2004). *Ultramicroscopy*, **100**, 35–57.
- Liu, Y., Meirer, F., Williams, P. A., Wang, J., Andrews, J. C. & Pianetta, P. (2012). *J. Synchrotron Rad.* **19**, 281–287.
- MacQueen, J. (1967). *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. M. Le Cam and J. Neyman, Vol. 1, pp. 281–297. University of California Press.
- Meirer, F., Liu, Y., Pouyet, E., Fayard, B., Cotte, M., Sanchez, C., Andrews, J. C., Mehta, A. & Sciau, P. (2013). *J. Anal. At. Spectrom.* **28**, 1870.
- Osanna, A. & Jacobsen, C. (2000). *X-ray Microscopy: Proceedings of the Sixth International Conference*, edited by W. Meyer-Ilse, T. Warwick and D. Attwood, pp. 350–357. Melville: American Institute of Physics.
- Schaefer, T., Michel, R., Claret, F., Beetz, T., Wirick, S. & Jacobsen, C. (2009). *J. Electron Spectrosc. Relat. Phenom.* **170**, 49–56.
- Schäfer, T., Buckau, G., Artinger, R., Kim, J. I., Geyer, S., Wolf, M., Bleam, W. F., Wirick, S. & Jacobsen, C. (2005). *Org. Geochem.* **36**, 567–582.
- Solé, V. A., Papillon, E., Cotte, M., Walter, P. & Susini, J. (2007). *Spectrochim. Acta B*, **62**, 63–68.
- Stöhr, J. (1992). *NEXAFS Spectroscopy*. Berlin: Springer-Verlag.
- The HDF Group (2000–2010). *Hierarchical data format*, Version 5, <http://www.hdfgroup.org/HDF5/>.
- Tzvetkov, G., Schmidt, N., Strunskus, T., Wöll, C. & Fink, R. (2007). *Surf. Sci.* **601**, 2089–2094.
- Walczak, M., Lawniczak-Jablonska, K., Sienkiewicz, A., Demchenko, I. N., Piskorska, E., Chatain, G. & Bohle, D. S. (2005). *Nucl. Instrum. Methods Phys. Res. B*, **238**, 32–38.
- Wasserman, S. R. (1997). *J. Phys. IV (Paris)*, **7(C2)**, 203–205.
- Zhang, X., Ade, H., Jacobsen, C., Kirz, J., Lindaas, S., Williams, S. & Wirick, S. (1994). *Nucl. Instrum. Methods Phys. Res. A*, **347**, 431–435.
- Zhang, X., Balhorn, R., Mazrimas, J. & Kirz, J. (1996). *J. Struct. Biol.* **116**, 335–344.