Robust framework and software implementation for fast speciation mapping

Serge X. Cohen, Samuel Webb, Pierre Gueriau, Emmanuel Curis, Loïc Bertrand

Robust framework and software implementation for fast speciation mapping

Serge X. Cohen,[a]* Samuel M. Webb,[b] Pierre Gueriau,[a,c1]

Emmanuel Curis[d,e] and Loïc Bertrand [a,f]*

[a] *Université Paris-Saclay, CNRS, ministère de la culture, UVSQ, IPANEMA, F-91192 Saint-Aubin, France,* [b] *Stanford Radiation Lightsource (SSRL), SLAC National Accelerator Laboratory, Menlo Park, CA, US,* [c] *Synchrotron SOLEIL, F-91192 Gif-sur-Yvette France,* [d] *Laboratoire de biomathématiques, EA 7537 – BioSTM, Faculté de Pharmacie de Paris – Université Paris Descartes, 4 avenue de l'Observatoire, F-75006 Paris, France,* [e] *Service de biostatistique et d'informatique médicale, hôpital Saint-Louis, Assistance publique – hôpitaux de Paris, 1 avenue Claude-Vellefaux, F-75010 Paris, France, and* [f] *Université Paris-Saclay, F-91190 Saint-Aubin, France. E-mail: serge.cohen@synchrotron-soleil.fr, loic.bertrand@universite-paris-saclay.fr*

Synchrotron Imaging; **Software**; **Chemical speciation**; **X-ray absorption spectroscopy**; **Heterogenous materials**

Abstract

One of the greatest benefits of synchrotron radiation is the ability to perform chemical speciation analysis through X-ray absorption spectroscopies (XAS). XAS imaging of large sample areas can be performed with either full-field or raster-scanning modalities.

[1] Current affiliation: Institute of Earth Sciences, University of Lausanne, CH-1015 Lausanne, Switzerland

A common practice to reduce acquisition time while decreasing dose and/or increasing spatial resolution is to compare X-ray fluorescence images collected at a few diagnostic energies. Several authors have used different multivariate data processing strategies to establish speciation maps. In this manuscript, we focus on the theoretical aspects and assumptions that are often made in the analysis of these datasets. A robust framework is developed to perform speciation mapping in large bulk samples at high spatial resolution by comparison to known references. Two fully operational software implementations are provided: a user-friendly implementation within the `MicroAnalysis Toolkit` software, and a dedicated script developed under the `R` environment. The procedure is exemplified through the study of a cross section of a typical fossil specimen. The algorithm provides accurate speciation and concentration mapping while decreasing the data collection time by typically 2 or 3 orders of magnitude compared to the collection of whole spectra at each pixel. Whereas acquisition of spectral datacubes on large areas leads to very high irradiation times and doses, which can considerably lengthen experiments and generate significant alteration of radiation-sensitive materials, this sparse excitation energy procedure brings the total irradiation dose greatly below radiation damage thresholds identified in previous studies. This approach is particularly adapted to the chemical study of heterogeneous radiation-sensitive samples encountered in environmental, material, and life sciences.

1. Introduction

Microscale chemical speciation is essential to describe a material's local composition, structure and properties in heterogeneous samples. Synchrotron-based X-ray absorption near-edge spectroscopy (XANES) is one of the main analytical methods used for this purpose, and it has been applied in a variety of scientific fields such as environmental sciences, heritage sciences, Earth and meteorite geology, and plant

sciences (Gardea-Torresdey *et al.*, 2005; Linge, 2008; Cotte *et al.*, 2010; Bertrand *et al.*, 2012; Etschmann *et al.*, 2014; Henderson *et al.*, 2014; Gueriau *et al.*, 2016). This excitation spectroscopy probes the energy levels of unoccupied electronic states from core levels upon X-ray photoexcitation. Of particular interest for the characterisation of heterogeneous materials is the use of focusing optics to acquire XANES in a spatially resolved manner. As X-ray optics have continued to improve in terms of reducing the size of focused beams, and synchrotron X-ray sources have become brighter, the ability to collect high-resolution X-ray images has become more and more routine. Two different experimental methodologies have been used to collect speciation images in transmission. The first strategy uses a large beam and a spatially resolved detector to collect 'instantaneous' images of the transmitted X-ray intensity through the sample at successive given excitation energies – effectively taking a radiograph of the sample at each excitation energy. The work of Kinney *et al.* was one of the first studying chemical contrast in imaging using transmitted radiographs above and below the absorption edge of interest to create the chemical contrast (Kinney *et al.*, 1986). These concepts have further been developed with full-field XANES, where the thin section of a sample is positioned in front of an extended X-ray beam; radiographs are collected at successive excitation energy points on an area detector positioned downstream (Ade *et al.*, 1992; De Andrade *et al.*, 2011; Fayard *et al.*, 2013*a*). This approach can be used without additional X-ray optics to collect high-resolution (typically micrometers) images using a luminescent screen and a visible microscope (Fayard *et al.*, 2013*b*; Tack *et al.*, 2014; Rau *et al.*, 2002; Sciau & Wang, 2019), or make use of X-ray condensing optics to illuminate the sample, followed by an X-ray imaging optical element, such as a Fresnel zone plate, to image the transmission onto an area detector as in transmission X-ray microscopy (TXM) with resolutions in the tens of nanometers (Holt *et al.*, 2013; Liu *et al.*, 2012*a*; Nelson *et al.*, 2011; Meirer *et al.*, 2011; Li

et al., 2015). A second strategy uses an X-ray focused microbeam to scan across a sample surface. In scanning transmission X-ray microscopy (STXM), the intensity of a transmitted beam is collected while varying the energy of the incoming beam and raster-scanning the sample with a fixed microscopic or nanoscopic X-ray beam (Zhang *et al.*, 1994; Flynn *et al.*, 1999; Jacobsen *et al.*, 2000). Both strategies allow for the position-sensitive detection of transmitted intensities in the presence and without the sample. The resulting maps allow reconstruction of a XANES spectrum at each pixel. Both are performed by measuring the transmitted X-ray intensities, which requires that samples are thinned down to allow for typically 10–20% of incident photons to cross the sample. Obviously, since the excitation energy is imposed by the absorption edge of the element whose speciation is being studied, it cannot be increased to improve transmission as in hard X-ray radiography or tomography. At low X-ray energies of a few keV or below, this imposes stringent constraints on sample thickness, equivalent to those requested in transmission electron microscopy. At greater energies of 8–22 keV, samples must still be thinned from a few dozen to a few hundred micrometers, which can be very constraining for samples. The transmission method is further constrained by the concentration of the element to be measured, as it needs to show strong enough absorption contrast across the energies to be measured.

When it is not possible or desirable to prepare thin sections, fluorescence detection of XANES signals from the surface of a thick sample is an option. This typically involves scanning the sample as in the second strategy above by detecting the fluorescent signal with an energy-dispersive X-ray fluorescence (XRF) detector generally placed at 90° with respect to the incident X-ray beam. In this geometry, a higher flux must reach the sample in order to obtain a signal-to-noise ratio identical to that of a transmission geometry. For example, the K-shell XRF yield for $3d$ transition metals of the 4th period (Sc–Zn) ranges from 16% to 46%. While XRF emission is isotropic, the solid

angles of most fluorescence detectors are limited, and only a small fraction of the total fluorescence signal can be measured. In addition, matrix effects (reabsorption of fluoresced photons) can significantly impact the collected signal. XRF detection is usually strongly count-rate limited, further decreasing yields of detection and signal-to-noise ratios. This methodology has been applied to a wide host of samples, including geological systems to examine transition metal speciation (Sutton *et al.*, 1995; Sutton *et al.*, 2002) and more dilute biological systems (Pushie *et al.*, 2014; Pickering *et al.*, 2000; Oram *et al.*, 2010).

While the collection of transmission and fluorescence images as a function of incident X-ray energy is a relatively common technique in the literature (Kaulich *et al.*, 2011; Lombi *et al.*, 2011; Wang *et al.*, 2013; Collins & Ade, 2012; Liu *et al.*, 2012*b*; Pickering *et al.*, 2000), the treatment and analysis of the datasets is not always straightforward. Several hundred images at different energies are often collected in full-field imaging, aiming to reconstruct a 'complete' XANES spectrum in each pixel. However, in dilute fluorescence experiments where raster-scan XRF imaging must be performed, a full complement of energies may take a prohibitive amount of experimental time. A selection of a limited number of energies must therefore be made using the chemical knowledge of the system, in order to contrast different chemical species in the specimen. Several review articles have pioneered the idea of mapping the chemical nature of the sample by collecting XRF data at several varied excitation energies (Bertsch & Hunter, 2001; Marcus, 2010). This can be done by exploiting the contrast between different species at, for example different resonances in the S K-edge XAS spectra for organic and inorganic forms in fossilised leaf materials (Edwards *et al.*, 2014). Lerotic *et al.* (2004) used cluster analysis for the analysis of multiple energy datasets. Other authors have proposed statistical treatments like principal component analysis (PCA) while the data collection is on-going to help identify regions of interest in the sample

in both transition metal systems (Mayhew *et al.*, 2011; Etschmann *et al.*, 2014) and complex sulfur speciation systems (Farfan *et al.*, 2018).

In this manuscript, we focus on the theoretical aspects and assumptions that are often made in the analysis of these datasets. A robust framework is developed to perform speciation mapping in large thick samples at high spatial resolution by comparison to known references. We show that, with the proper experimental setup and the measurement of images over a selected number of energy points, the linear decomposition of the XANES signal yields the relative concentration of individual chemical species and the total concentration of the element. While using a linear decomposition, this method takes advantage of the increased number of effective data measurements at each pixel in the image, which results in improved overall counting statistics. The method minimizes beam exposure to the sample, which is particularly important in the case of samples that may be sensitive to radiation damage, such as the real paleontological sample on which the effectiveness of this approach is tested. The low computational cost of the involved multivariate algorithm combined with the usefulness of the produced results make the proposed method suitable to inline use while collecting X-ray data and furthermore provides useful hints to the experimental design. To facilitate this aspect, the proposed approach has been integrated and deployed into a freely available software package, the `MicroAnalysis Toolkit` (Webb, 2011), and a dedicated package developed under the `R` environment. This approach is of interest for the study of a wide range of materials containing different chemical forms of an element, as frequently encountered in many fields of materials science.

2. Methods – Sparse excitation energy XAS

The following approximations have been made considering that the XAS signal is detected in fluorescence, on a thick sample (an infinitely thin sample would give similar

133 approximations), regarding a target element diluted in a matrix presenting a significant
134 regularity.

135 *2.1. Notations*

136 Let I_0 be the incident beam intensity, and θ and ϕ the angles between the sample
137 surface, and the incident and detected beam, respectively (Fig. 1). The experimental
138 frame is set such that the incident beam propagates along the $\mathbf{x}$ axis, the $\mathbf{z}$ axis
139 is vertical pointing upward and $\mathbf{y}$ is such that $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ forms a direct orthonormal
140 frame. The fluorescence is measured at a right angle from the incident beam such
141 that $\theta + \phi = \frac{\pi}{2}$. We consider that the measured fluoresced signal propagates along $\mathbf{y}$.
142 Note that although these constraints correspond to usual geometries and simplify the
143 remaining computation, they have no direct consequence on the overall result.

144 [Table 1 about here.]

145 We assume that we collect the experimental fluoresced intensity I_f^{Exp} from a rect-
146 angular parallelepiped of thickness d. The thickness in the beam direction is then
147 $e = d / \sin\theta$. We attach to the sample a frame with the $\mathbf{u}$ axis orthogonal to
148 the entry face and pointing inward the sample, $\mathbf{z}$ the unitary vertical vector pointing
149 upward and $\mathbf{v}$ parallel to the entry face such that the $(\mathbf{u}, \mathbf{v}, \mathbf{z})$ forms a direct orthonor-
150 mal basis. Both orthonormal frames have their origin on the sample surface such that
151 voxels on the surface are characterised by $x = y = 0$ or $u = 0$ according to the set
152 of coordinates used. We will also use $\mathbf{r}$ to point to the position of a voxel irrespective
153 of the coordinate system used. The pixel coordinates on the sample surface are noted
154 (v, z) with implied $u = 0$.

155 The incident and fluoresced energies are denoted E and E_f respectively. We included
156 a possible indirect dependency of I_0 in (E, v, z) as fluctuations of I_0 occurs in the
157 scanning timescale which we express as $I_0(t(E, v, z))$.

158 [Figure 1 about here.]

159 *2.2. Algorithmic development*

XRF from an elementary volume: general case. The linear absorption coefficient for the voxel at $\mathbf{r}$ is noted $\mu(\mathrm{E}, \mathbf{r})$, and can be decomposed as:

$$\mu(\mathrm{E}, \mathbf{r}) = \mu_{\mathrm{m}}(\mathrm{E}, \mathbf{r}) + \sum_{k=1}^{K} \mathcal{C}_k(\mathbf{r})\ \mu_k(\mathrm{E})$$

160 where $\mu_{\mathrm{m}}(\mathrm{E}, \mathbf{r})$ is the matrix absorption coefficient that varies slowly with energy E
161 and $\mu_k(\mathrm{E})$ is the absorption coefficient of species k around the 'target' elemental edge
162 considered — the linear combination standing if the atom probed exists as K distinct
163 chemical forms. In the energy range which is usually considered for XAS studies,
164 *photo-absorption* is by far the main photon–matter interaction mechanism leading to
165 fluorescence. Hence we will consider all along our computation that global absorption,
166 be it by the matrix or the target element, is photo-absorption related. Furthermore,
167 since the incident energy E is close to the excitation energy of the target element, it
168 is unlikely that the fluoresced signal produces any secondary sources of fluorescence
169 from this element and hence we consider only primary fluorescence signal.

Let's first consider the general case by computing the contribution of a given voxel $\mathbf{r}_0 = (x_0, y_0, z_0)$:

$$\mathrm{I}_{\mathrm{f}}^{\mathrm{Exp}}(\mathrm{E}, \mathbf{r}_0) = \mathrm{I}_0(\mathrm{t}(\mathrm{E}, v, z))\ \frac{\varepsilon_{\mathrm{f}}\,\Omega}{4\,\pi} \tag{1}$$

$$\cdot \exp\left(-\int_{x=0}^{x_0} \mu(\mathrm{E}, x, y_0, z_0)\mathrm{d}\,x\right) \tag{2}$$

$$\cdot \left(\sum_{k=1}^{K} \gamma_{k,\mathrm{E_f}}\,\mathcal{C}_k(\mathbf{r}_0)\,\mu_k(\mathrm{E})\right) \tag{3}$$

$$\cdot \exp\left(-\int_{y=y_0}^{y_0+x_0\tan\theta} \mu(\mathrm{E_f}, x_0, y, z_0)\mathrm{d}\,y\right) + \epsilon(\mathrm{E}, v, z) \tag{4}$$

170 Line 1 of this equation corresponds to factors independent from (x_0, y_0, z_0): the
171 incident beam flux, $\mathrm{I}_0(\mathrm{t}(\mathrm{E}, v, z))$, and the detector characteristics, $\frac{\varepsilon_{\mathrm{f}}\,\Omega}{4\,\pi}$ that accounts

for its quantum efficiency (ε_{f}) and the solid angle it covers (Ω). Note that considering that Ω is constant for all values of x in the sample assumes that the detector is sufficiently far away from the voxel of interest to consider the incoming fluorescent beam as a parallel beam coming from the infinite.

Line 2 corresponds to the decrease of the incident flux due to absorption while the incident beam penetrates in the $\mathbf{x}$ direction to the voxel of interest.

Line 3 accounts for the quantity of fluorescence emitted by the elementary volume irradiated by the incident microbeam. It takes into account the fluorescence yield $(\gamma_{k,\mathrm{E_f}})_{k\in[\![1,K]\!]}$ of the K species of the target element which edges are probed and their amount in the given voxel. The latter will be called the *fluorescence source term*.

Finally, line 4 of this equation considers that the fluorescence emission at $\mathrm{E_f}$ is attenuated due to reabsorption in the sample on the outgoing path expressed in its integral form arising from the Beer–Lambert law. Finally we have the error term $\epsilon(\mathrm{E}, v, z)$ that we will omit from now on to simplify notations.

Target element diluted within the matrix. Assuming that the K species of the target element are very diluted in the matrix, the absorption coefficient, $\mu(\mathrm{E}, \mathbf{r})$ is largely dominated by its matrix-related contribution:

$$\mu(\mathrm{E}, \mathbf{r}) \approx \mu_{\mathrm{m}}(\mathrm{E}, \mathbf{r}) \tag{5}$$

Under this hypothesis, the amplitude of the attenuation, both on the incident and fluoresced beam, is made independent of the concentration of the species of the target element. In this approximation, the absorption of the target element affects only the *fluorescence source term*.

Matrix and target element of homogeneous composition at the scale of the probed volume. We consider that both the composition of the matrix and the

concentrations of the target element in its various species, $(\mathcal{C}_k)_{k\in[\![1,K]\!]}$, vary sufficiently slowly so that they can be considered as constant on the neighborhood corresponding to the measure in each value of (v,z). More precisely, if we consider the $\mathbf{r}_0 = (x_0 = 0, y_0, z_0) = (u_0 = 0, v_0, z_0)$ voxel, we assume that there is a *homogeneity radius*, noted l, greater than the radius of the volume probed by the beam when the sample is fixed and such that:

$$\forall \mathbf{r}, \|\mathbf{r}-\mathbf{r}_0\| < l \Rightarrow (\mu_{\mathrm{m}}(\mathrm{E},\mathbf{r}) \approx \mu_{\mathrm{m}}(\mathrm{E},\mathbf{r}_0)) \text{ and } (\forall k \in [\![1,K]\!], \mathcal{C}_k(\mathbf{r}) \approx \mathcal{C}_k(\mathbf{r}_0)) \tag{6}$$

Combining these hypotheses with the one developed in the previous paragraph we obtain an analytical expression of the measured fluorescence originating from a voxel centred on $\mathbf{r}$, showing a simple dependency on the concentrations $\mathcal{C}_k(\mathbf{r})$:

$$\begin{aligned} \mathrm{I}_{\mathrm{f}}^{\mathrm{Exp}}(\mathrm{E},\mathbf{r}) = \mathrm{I}_0(\mathrm{t}(\mathrm{E},v,z))\,\frac{\varepsilon_{\mathrm{f}}\,\Omega}{4\,\pi} \\ \cdot \exp\left(-\mu_{\mathrm{m}}(\mathrm{E},\mathbf{r})x_0\right)\left(\sum_{k=1}^{K}\gamma_{k,\mathrm{E_f}}\,\mathcal{C}_k(\mathbf{r})\,\mu_k(\mathrm{E})\right) \\ \cdot \exp\left(-\mu_{\mathrm{m}}(\mathrm{E_f},\mathbf{r})x_0\tan\theta\right) \end{aligned} \tag{7}$$

Characteristic matrix absorption and thickness of the sample. In the previous equation, terms in the exponential can be grouped and the $-x_0$ term can be factorised and we can define $\mu_{\mathrm{cm}}(\mathrm{E},\mathrm{E_f},v,z) = \mu_m(\mathrm{E},v,z) + \tan\theta\mu_{\mathrm{m}}(\mathrm{E_f},v,z)$, such that the expression of the contribution of a voxel becomes:

$$\mathrm{I}_{\mathrm{f}}^{\mathrm{Exp}}(\mathrm{E},\mathbf{r}) = \mathrm{I}_0(\mathrm{t}(\mathrm{E},v,z))\,\frac{\varepsilon_{\mathrm{f}}\,\Omega}{4\,\pi}\;\cdot\;\exp\left(-\mu_{\mathrm{cm}}(\mathrm{E},\mathrm{E_f},v,z)x_0\right)\;\cdot\;\left(\sum_{k=1}^{K}\gamma_{k,\mathrm{E_f}}\,\mathcal{C}_k(v,z)\,\mu_k(\mathrm{E})\right) \tag{8}$$

$l_c(\mathrm{E_f},v,z) = \max_{\mathrm{E}}(1/\mu_{\mathrm{cm}}(\mathrm{E},\mathrm{E_f},v,z))$ can be interpreted as the characteristic attenuation length of the experimental setup, and we need to have $l_c(\mathrm{E_f},v,z) \ll l$ to satisfy the hypothesis of an homogeneous matrix and target element species composition. Furthermore if d, the thickness of sample, satisfies $d \gg l_c(\mathrm{E_f},v,z)$ we can treat the sample as infinitely thick. This condition, together with the hypothesis expressed in

the previous section, basically states that the matrix has a *homogeneity radius* which is both greater than the beam radius and the penetration depth of the incident beam interaction.

Summation along the incident beam. Eq. 8 is implicitly an infinitesimal form, corresponding to

$$d\,\mathrm{I}_{\mathrm{f}}^{\mathrm{Exp}} = \mathrm{I}_{\mathrm{f}}^{\mathrm{Exp}}\,(\mathrm{E}, \mathbf{r}) d\mathbf{r} \tag{9}$$

This can now be summed over the path of the incident beam (in x_0), considering a semi-infinite sample, leading to the intensity of the fluorescence signal measured when the pixel (v, z) is illuminated by the probe:

$$d\,\mathrm{I}_{\mathrm{f}}^{\mathrm{Exp}}\,(\mathrm{E}, v, z) = \mathrm{I}_0(\mathrm{t}(\mathrm{E}, v, z))\,\frac{\varepsilon_{\mathrm{f}}\,\Omega}{4\,\pi}\;\cdot\;\frac{1}{\mu_{\mathrm{cm}}(\mathrm{E}, \mathrm{E_f}, v, z)}\;\cdot\;\left(\sum_{k=1}^{K}\gamma_{k,\mathrm{E_f}}\,\mathcal{C}_k(v, z)\,\mu_k(\mathrm{E})\right) dvdz \tag{10}$$

which is a linear equation from the viewpoint of the $(\mathcal{C}_k)_{k\in[\![1,K]\!]}$. We can then estimate the coefficients of these linear equations by measuring K spectra of references that are linearly independent. The simplest way would be to measure K spectra $(\mathcal{S}_i)_{i\in[\![1,K]\!]}$ such that $\forall(i,k) \in [\![1,K]\!]^2$ the *a priori* known concentrations are $(\mathcal{C}_k)_i = \delta(i,k)$. In other words, K reference spectra corresponding to a single species in a unique matrix.

Reference and normalised spectra. In the general case corresponding to the approximations (Eq. 10), the reference spectra allowing to estimate the parameters $(\varepsilon_{\mathrm{f}}, \mu_{\mathrm{cm}}(\mathrm{E}, \mathrm{E_f}, v, z), (\gamma_{k,\mathrm{E_f}})_{k\in[\![1,K]\!]})$ through the knowledge of $\mu_k(\mathrm{E}))_{k\in[\![1,K]\!]})$ must be measured under the same conditions, the same matrix, the same experimental setup as the measurements made during the sparse excitation energy XAS experiment. This is in the best cases unpractical and in most cases unfeasible.

However, the X-ray absorption spectra on an energy domain allowing to probe an electron edge bear specificity that will allow us to relax these constraints.

Approximate affine behaviour of the matrix. First considering the matrix alone in the typical energy domain probed during a XAS experiment, we approximate the absorption coefficient using a first order Taylor expansion leading to an affine function. This holds as long as no constituent of the matrix has a fluorescence edge in the domain:

$$\exists(\mathcal{A}_m, \mathcal{B}_m) \in \mathbb{R}^2 / \mu_m(\mathrm{E}) \approx \mathcal{B}_m + \mathcal{A}_m \times \delta\mathrm{E}, \text{ for } \delta\mathrm{E} = \mathrm{E} - \mathrm{E}_f \tag{11}$$

$$\text{with } |\mathcal{A}_m \times \delta\mathrm{E}| \ll \mathcal{B}_m \tag{12}$$

In this setting we can approximate the value of $\mu_{\mathrm{cm}}(\mathrm{E}, \mathrm{E_f}, v, z)$:

$$\begin{aligned}
\mu_{\mathrm{cm}}(\mathrm{E}, \mathrm{E_f}, v, z) &= \mu_m(\mathrm{E}, v, z) + \tan\theta \mu_{\mathrm{m}}(\mathrm{E_f}, v, z) \\
&\approx \mathcal{B}_m(v, z) + \mathcal{A}_m(v, z) \times \delta\mathrm{E} + \tan\theta \mathcal{B}_m(v, z), \text{ as } \delta\mathrm{E}_f = 0 \text{ by definition} \\
&\approx (1 + \tan\theta)\, \mathcal{B}_m(v, z) + \mathcal{A}_m(v, z) \times \delta\mathrm{E} \\
&\approx (1 + \tan\theta)\, \mathcal{B}_m(v, z) \text{ since } |\mathcal{A}_m \times \delta\mathrm{E}| \ll \mathcal{B}_m
\end{aligned} \tag{13}$$

basically stating that $\mu_{\mathrm{cm}}(\mathrm{E}, \mathrm{E_f}, v, z)$ has very small variation when E varies within a small energy domain.

Linear composition of near-edge absorption spectra. The contribution of the different electronic levels of an atom to the absorption signals that correspond to different absorption edges is composed by simple addition of the effects at each of the edges taken individually. Thus the approximation described above is valid and usable for the contribution of the absorption of unprobed edges, and the absorption coefficient of the K species probed can be written as the sum of an affine contribution and a contribution representing the very edge probed for each species, measured through fluorescence signal, $\gamma_{k,\mathrm{E_f}}\, \mu_k^{\mathrm{edge}}(\mathrm{E})$:

$$\forall \mathrm{E}, \gamma_{k,\mathrm{E_f}}\, \mu_k(\mathrm{E}) \approx \mathcal{B}^{\mathrm{edge}} + \mathcal{A}^{\mathrm{edge}} \times \delta\mathrm{E} + \gamma_{\mathrm{E_f}}\, \mu_k^{\mathrm{edge}}(\mathrm{E}) \tag{14}$$

with

$$\forall k \in [\![1, K]\!], \mathrm{E} \ll \mathrm{E}^{\mathrm{edge}} \Rightarrow \mu_k^{\mathrm{edge}}(\mathrm{E}) = 0$$
$$\mathrm{E} \gg \mathrm{E}^{\mathrm{edge}} \Rightarrow \mu_k^{\mathrm{edge}}(\mathrm{E}) = \mu^{\mathrm{edge}}(\mathrm{E}) \approx b^{\mathrm{edge}} + a^{\mathrm{edge}} \times \delta\mathrm{E} \qquad (15)$$

In Eq. 15, the term $a^{\mathrm{edge}} \times \delta\mathrm{E}$ corresponds to the out-of-edge change in slope, *i.e.* the slope correction applicable to Eq. 14 for the part of the spectrum collected at $\mathrm{E} > \mathrm{E}^{\mathrm{edge}}$. If $|a^{\mathrm{edge}}| \ll |\mathcal{A}^{\mathrm{edge}}|$, one can, on a small domain that does not extend too much after the edge energy $\mathrm{E}^{\mathrm{edge}}$, neglect a^{edge} and therefore write:

$$\forall \mathrm{E}, \gamma_{k,\mathrm{E_f}}\, \mu_k(\mathrm{E}) \approx \mathcal{B}^{\mathrm{edge}} + \mathcal{A}^{\mathrm{edge}} \times \delta\mathrm{E} + \gamma_{\mathrm{E_f}}\, b^{\mathrm{edge}}\, \widehat{\mu_k}(\mathrm{E}) \qquad (16)$$

with

$$\forall k \in [\![1, K]\!], \mathrm{E} \ll \mathrm{E}^{\mathrm{edge}} \Rightarrow \widehat{\mu_k}(\mathrm{E}) = 0$$
$$\mathrm{E} \gg \mathrm{E}^{\mathrm{edge}} \Rightarrow \widehat{\mu_k}(\mathrm{E}) = 1$$

Eq. 16 defines the $(\widehat{\mu_k}(\mathrm{E}))_{k\in[\![1,K]\!]}$, the normalised spectra of the K reference species at the probed edge, which under the current approximations are unaffected by the matrix used to measure the reference spectra, as long they are measured in fluorescence mode at energy $\mathrm{E_f}$. Indeed under the current approximations the dependence of the measurement to the matrix is encoded within the $\mathcal{A}^{\mathrm{edge}}$ and $\mathcal{B}^{\mathrm{edge}}$ which in turn are not affected by the species of the probed element, whilst b^{edge} depends only on the edge probed. In other words, Eq. 16 allows us to model the reference spectra using species-independent terms (parameters: $\mathcal{B}^{\mathrm{edge}}$, $\mathcal{A}^{\mathrm{edge}}$), a matrix-independent term ($(\widehat{\mu_k}(\mathrm{E}))_{k\in[\![1,K]\!]}$), and a term independent of both ($\gamma_{\mathrm{E_f}} b^{\mathrm{edge}}$). In this context, the $(\widehat{\mu_k}(\mathrm{E}))_{k\in[\![1,K]\!]}$ can be obtained by measuring the reference species in any matrix and at any concentration provided that the approximation of the diluted sample holds, as usually done.

Note finally that the affine absorption behaviour of the matrix, characterized in Eq. 11 by A_m and B_m, is accounted for with the parameters $\mathcal{A}^{\mathrm{edge}}$ and $\mathcal{B}^{\mathrm{edge}}$.

Linear combination of reference spectra. We may now rewrite Eq. 10 using the approximation from Eq. 13 for $\mu_{\rm cm}(\mathrm{E},\mathrm{E_f},v,z)$ and Eq. 16 for $(\mu_k(\mathrm{E}))_{k\in[\![1,K]\!]}$:

$$\frac{\mathrm{I_f^{Exp}}\,(\mathrm{E},v,z)}{\mathrm{I_0(t(E},v,z))} = \frac{\varepsilon_\mathrm{f}\,\Omega}{4\,\pi}\;\cdot\;\frac{1}{(1+\tan\theta)\,\mathcal{B}_m(v,z)} \left[\left(\mathcal{B}^{\rm edge}+\mathcal{A}^{\rm edge}\times\delta\mathrm{E}\right)\cdot\left(\sum_{k=1}^{K}\mathcal{C}_k(v,z)\right)+\gamma_{\mathrm{E_f}}\,b^{\rm edge}\left(\sum_{k=1}^{K}\mathcal{C}_k(v,z)\,\widehat{\mu_k}(\mathrm{E})\right)\right] \tag{17}$$

Which states that for a given (v,z) position the measured signal, $\frac{\mathrm{I_f^{Exp}}\,(\mathrm{E},v,z)}{\mathrm{I_0(t(E},v,z))}$ is an affine function of the $(\widehat{\mu_k}(\mathrm{E}))_{k\in[\![1,K]\!]}$ and $\delta\mathrm{E}$ with coefficients:

- $\mathrm{C_0}(v,z)=\mathrm{D_m}(\mathrm{E_f},v,z)\,\mathcal{B}^{\rm edge}\left(\sum_{k=1}^{K}\mathcal{C}_k(v,z)\right)$ for the intercept;
- $\mathrm{C_{\delta E}}(v,z)=\mathrm{D_m}(\mathrm{E_f},v,z)\,\mathcal{A}^{\rm edge}\left(\sum_{k=1}^{K}\mathcal{C}_k(v,z)\right)$ for the coefficient of $\delta\mathrm{E}$;
- $\mathrm{C_k}(v,z)=(\mathrm{D_m}(\mathrm{E_f},v,z)\,\gamma_{\mathrm{E_f}}\,b^{\rm edge}\,\mathcal{C}_k(v,z))_{k\in[\![1,K]\!]}$ for the coefficient of each of the $(\widehat{\mu_k}(\mathrm{E}))_{k\in[\![1,K]\!]}$,

where $\mathrm{D_m}(\mathrm{E_f},v,z)=\frac{\varepsilon_\mathrm{f}\,\Omega}{4\,\pi}\;\cdot\;\frac{1}{(1+\tan\theta)\,\mathcal{B}_m(v,z)}$ is independent of the incident energy E.

For any given coordinate set (v,z) on the sample, one can estimate the above expressed coefficient, $(\mathrm{C_0}(v,z),\mathrm{C_{\delta E}}(v,z),(\mathrm{C_k}(v,z))_{k\in[\![1,K]\!]})$, by performing a linear least square fitting of $(N\geq K+2)$ measurements corresponding to energies $(\mathrm{E}_n)_{n\in[\![1,N]\!]}$ such that the vector $(\widehat{\mu_k}(\mathrm{E}_1),..,\widehat{\mu_k}(\mathrm{E}_N))_{k\in[\![1,K]\!]}$ and $(\delta\mathrm{E}_1,..,\delta\mathrm{E}_N)$ are linearly independent. The estimated coefficients $(\mathrm{C_k}(v,z))_{k\in[\![1,K]\!]}$ correspond indeed to the relative concentration of each of the K species that could be present in the samples at coordinates (v,z) hence providing a relative estimation of the $(\mathcal{C}_k(v,z))_{k\in[\![1,K]\!]}$.

Further notational simplification. Since the absolute scale of Eq. 17 can not be easily computed from information one usually have *a priori* on the system, there is no incentive to keep the full notations. We can simplify the notation by integrating the scaling factors within the parameters without altering the relative values of $\mathrm{C_k}(v,z)$ taken from diffrent pixels. For pixel (v,z) the scale is directly integrated in the $\mathrm{C_k}(v,z)$, while the slope and intersect of the affine baseline will be noted as $\mathrm{A}(v,z)$ and $\mathrm{B}(v,z)$

which leads to Eq. 20 in the main text.

Validation of the model and estimation of the relative accuracy. Under the model developed above, as long as the matrix of the sample is homogeneous, the estimated coefficients $\mathrm{C}_0(v,z)$ and $\mathrm{C}_{\delta\mathrm{E}}(v,z)$ should be strictly correlated, within measurement noise error, as they both should be correlated to the estimated sum of relative concentrations, $\mathrm{C}^{\mathrm{sum}}(v,z) = \sum_{k=1}^{K} \mathrm{C}_{\mathrm{k}}(v,z)$. Indeed such correlations are very unlikely to be present by chance and the quality of the correlations can be used as a marker of the validity of the proposed model. Furthermore, as the proposed method is used to perform speciation imaging, hence collecting data on a large set of (v,z) coordinates, these correlations can be verified on the basis of a significant number of data points.

Note, however, that if the hypothesis of the homogeneity of the matrix is not met anymore only the $\mathrm{C}_0(v,z)$ and $\mathrm{C}_{\delta\mathrm{E}}(v,z)$ estimation are affected since only $\mathcal{A}^{\mathrm{edge}}$ and $\mathcal{B}^{\mathrm{edge}}$ depend on the matrix. In these circumstances $\mathcal{A}^{\mathrm{edge}}$ and $\mathcal{B}^{\mathrm{edge}}$ become functions of the probed coordinates, (v,z), and the correlations that are described above no longer stand. Then $\mathrm{C}^{\mathrm{sum}}(v,z)$ is the only available estimate of the relative concentration of the probed element at the different measured coordinates.

Finally, within the framework of linear least square, one can compute the condition number, κ, of the design matrix as soon as the $(\widehat{\mu_k}(\mathrm{E}_1),..,\widehat{\mu_k}(\mathrm{E}_N))_{k\in[\![1,K]\!]}$ and $(\delta\mathrm{E}_1,..,\delta\mathrm{E}_N)$ are known, and before performing any measure on the actual sample. With κ known it is easy to obtain an upper bound on relative errors in relative concentrations, $\epsilon_{\mathcal{C}_k}$, given a level of relative errors, also termed signal to noise ratio, in fluorescence signal measurements:

$$\epsilon_{\mathcal{C}_k} \leq \kappa\, \epsilon_{\mathrm{fluorescence}} \tag{18}$$

Through this simple inequality which can be computed very early in the experiment

design, one can set the appropriate exposure time to reach the fluorescence signal to noise ratio, $\epsilon_{\text{fluorescence}}$, so as to reach a predefined requirement in term of accuracy of the relative species concentrations.

Under such conditions, the experimental XAS spectrum at any given pixel of coordinates (v, z) results from a linear combination of the K reference spectra $(\mu_k(\mathrm{E}))_{k\in[\![1,K]\!]}$ and of an affine baseline:

$$\frac{\mathrm{I_f^{Exp}}(\mathrm{E},v,z)}{\mathrm{I_0}(\mathrm{t}(\mathrm{E},v,z))} \approx \frac{\Gamma}{\mu_{\mathrm{cm}}(v,z)}\left[\left(\sum_{k=1}^{K}\mathcal{C}_k(v,z)\mathrm{dV}\,\mu_k(\mathrm{E})\right)+\mathcal{A}(v,z)\times\mathrm{E}+\mathcal{B}(v,z)\right] + \epsilon(\mathrm{E},v,z) \tag{19}$$

where $\mathrm{I_f^{Exp}}(\mathrm{E},v,z)$ is the fluorescence signal measured when illuminating pixel (v, z) at an excitation energy E and $\mathrm{I_0}(\mathrm{t}(\mathrm{E},v,z))$ is the incident beam intensity at the moment $\mathrm{t}(\mathrm{E},v,z)$, when $\mathrm{I_f^{Exp}}(\mathrm{E},v,z)$ is measured. $\mathcal{A}(v,z)\times\mathrm{E}+\mathcal{B}(v,z)$ models the summed up slow varying background from XAS spectra of the sample and of the references and $\mathcal{C}_k(v,z)\mathrm{dV}$ is the quantity of the species k in the elementary volume dV probed. Γ is a proportionality constant depending on the XRF yield of the probed element and on the geometry of the setup (solid angle of the detector, etc.), $\mu_{\mathrm{cm}}(v,z)$ is the apparent absorption coefficient of the matrix, accounting for both incident beam and fluoresced signal attenuation, at pixel (v, z) and $\epsilon(\mathrm{E},v,z)$ is a zero mean stochastic error term which is independent from one measurement to the next.

The parameters $(\mathcal{C}_k(v,z))_{k\in[\![1,K]\!]}$, $\mathcal{A}(v,z)$ and $\mathcal{B}(v,z)$ can be estimated using the linear least square fit of the equation system to the measurements. With the previous assumption of a zero-mean $\epsilon(\mathrm{E},v,z)$, these estimates are unbiased; with the additional assumption of independence and constant variance, they correspond to maximum likelihood estimation. From now, we will simplify the notations by keeping the error term out of all expressions. Note that by normalising the reference spectra, $(\widehat{\mu_k}(\mathrm{E}))_{k\in[\![1,K]\!]}$, we bring $\mathcal{C}_k$ on an arbitrary unit and as such there is not much interest in obtaining $\mathcal{C}_k$ rather than $\mathrm{C_k} \propto \frac{\Gamma}{\mu_{\mathrm{cm}}(v,z)}\mathcal{C}_k\mathrm{dV}$. Finally the very limited dependence that $\mu_{\mathrm{cm}}(v,z)$

has in term of E can be accounted for in the affine baseline (Eq. 13), thus we can simplify our equation to:

$$\frac{\mathrm{I_f^{Exp}}(\mathrm{E},v,z)}{\mathrm{I_0}(\mathrm{t}(\mathrm{E},v,z))} \approx \left(\sum_{k=1}^{K} \mathrm{C_k}(v,z)\widehat{\mu_k}(\mathrm{E})\right) + \mathrm{A}(v,z)\times\mathrm{E} + \mathrm{B}(v,z) \tag{20}$$

For a given (v,z) pixel position and using N probing energies $(\mathrm{E}_n)_{n\in[\![1,N]\!]}$ equation (20) produces a linear system of N equations and $K+2$ unknowns ($\mathrm{C_k}(v,z)$, $\mathrm{A}(v,z)$ and $\mathrm{B}(v,z)$). The linear least square solution requires that the number of equations is at least equal to the number of unknowns, $N \geq K+2$. Interestingly when all needed references are identified, $\sum_{k=1}^{K}\mathrm{C_k}(v,z) = \mathrm{C^{sum}}(v,z)$, is proportional to the absorbing element concentration. This provides an elemental *relative concentration* map as a side result of the full procedure.

3. Open source software implementation

The XAS procedure described here has been integrated into the most recent versions of the `MicroAnalysis Toolkit` package (Webb, 2011) under the "XANES Fitting" analysis option. The latter package utilizes a combination of self-written routines in Python (VanRossum & Drake, 2010) with the `SciPy` (Jones *et al.*, 2001–) package used for minimization. Compiled versions and source code for the software updated with these routines is made available as of version 1.5 under an open source license in parallel to this publication (Webb, 2010–). File formats from many synchrotron sources worldwide are able to load, and the list of compatible data formats is expanding upon request. The program is freely available and can be run on multiple operating system platforms.

In addition, a specific script was developed under the `R` statistical environment (R Development Core Team, 2011) and is made freely available with this publication. The script is annotated so as to provide a step by step description of the data processing,

in such a way that it is easy for a user to adapt it to its own data. The script starts from raw XRF data and covers all steps up to decomposition and statistical analysis. In particular several diagnosis steps are present in this script, as a contribution to data validation, while not all described in this manuscript. The script and all datasets needed to reproduce the presented results can be found at `https://doi.org/10.5281/zenodo.3748752`.

4. Application to a fossil sample

The proposed procedure was tested in the study of the distribution of the oxidation states of cerium in a paleontological sample at micrometric scales, as typical heterogeneous sample containing an element with two distinct valencies. An exception among rare earth elements, the element cerium exists in both the +III and +IV oxidation states. The resulting sensitivity to redox conditions makes the oxidation state of cerium a frequently used proxy of the chemistry of past environments (German & Elderfield, 1990). Synchrotron XAS mapping allows studying directly the speciation of cerium within ancient samples (such as sediment or bone; Gueriau *et al.*, 2015). Yet, a significant difficulty is that a high radiation dose can result in the photo-oxidation of Ce(III) to Ce(IV) during collection of entire XANES spectra, thereby biasing the measured data (Gueriau *et al.*, 2015).

We studied a loose fragment (not a cross-section) of well preserved soft tissues (mineralized muscles) around a bone of a yet undescribed ?gonorynchiform fish (where "?" indicates that the supposed identification has not yet been corroborated by a complete anatomical description) coming from the Late Cretaceous levels of the Djebel Oum Tkout locality in Morocco (ca. 95 million years old, Fig. 2a,b). We produced rapid maps of cerium speciation of the sample at the Lucia beamline of the SOLEIL synchrotron source (Vantelon *et al.*, 2016) over a 6466 pixel area ($120\times210\,\mu m^2$ with

a scan step of $2\times2\,\mu m^2$; beam size $3.0\times2.5\,\mu m^2$ (H×V); flux: 2.8×10^{10} photons/s in the focused spot) using the sparse excitation energy procedure presented above. Spectra of cerium III and IV reference compounds were collected using the procedure indicated in Gueriau *et al.* (2015).

Gueriau *et al.* (2015) studied a series of cross sections from well-preserved fossil teleost fishes and crustaceans from the same locality (see Fig. 4 in Gueriau *et al.*, 2015) using a limited number of excitation energy points to limit photo-oxidation changes they observed within a few tens seconds during XAS acquisition. They selected six excitation energies: below the L_3 absorption edge of cerium (5680 eV), in the post-edge region (5765 and 5900 eV) and at energies corresponding to three main spectral features that allow distinguishing Ce(III) from Ce(IV) in cerium-containing compounds — a white line at 5726 eV resulting from the $2p \rightarrow (4f^1)5d^*$ electron transition and two peaks that partly overlap with the former attributed to the $2p_{3/2} \rightarrow (4f^1\underline{L})5d^*$ and the $2p_{3/2} \rightarrow (4f^0)5d^*$ transitions, where $\underline{L}$ denotes a hole in ligand orbitals (at 5733 and 5736 eV respectively) (Takahashi *et al.*, 2000).

[Figure 2 about here.]

In the loose soft tissue fragment of the ?gonorynchiform fish, the procedure developed here (using the same excitation energies) reveals the presence of a fraction of cerium in the redox +IV state ([Ce(IV)]/[Ce(tot)] = $25.6 \pm 8.3\%$) over the whole mapped area, suggesting slightly oxidative local conditions of burial (Fig. 2c–f). This value agrees well with the indirect data obtained from micro- and macroscale elemental concentration profiles measured at the site (so-called 'REE patterns'), which also display a slight negative cerium anomaly attributed to oxidative burial conditions (Gueriau *et al.*, 2015). The total cerium concentration, [Ce(tot)], was mapped in the sample using two different approaches. The cerium distribution shown in Fig. 2d is the result of the usual processing of an XRF dataset collected after the cerium

edge (5900 eV), while Fig. 2e shows the [Ce(tot)] map derived from the proposed XAS procedure. Since the latter takes advantage of the total collection time of the several XRF maps at different energies, it led to a far higher signal-to-noise ratio than the single-energy XRF map, without the need to switch to an additional incident energy significantly above the absorption edge associated with the element. While the total concentration of cerium differs significantly between bone and surrounding mineralized muscles, the distribution of the Ce(IV) fraction relative to total cerium appears very stable and homogeneous between bone ($27.2 \pm 2.9\%$) and the surrounding mineralized muscles ($27.7 \pm 5.2\%$, Fig. 2g), which corroborates the validity of the redox proxy at different length scales in the presence of varying cerium ion concentrations.

A dwell time of 1 s per pixel was used. The dead times associated with the movements of the scanning motors led to a total collection time of about 165 min per excitation energy, i.e. about 16.5 h for the full dataset (collection at six excitation energies). The typical acquisition time for a single full XAS spectrum was 243 s, dead times included; the collection of the same dataset with a complete XAS spectrum per pixel would therefore have taken 436 h, i.e. about 26 times longer. While the collection of a full XAS spectrum under these conditions would have resulted in a *surface* irradiation dose (Howells *et al.*, 2009; Bertrand *et al.*, 2015) of 5.3 GGy, this value was reduced by a similar factor of 26 to about 200 MGy per pixel, with a counting time of 9.2 s per pixel, significantly less than the dose leading to a change in speciation, as shown by the photo-oxidation kinetics of cerium in comparable specimens (Fig. 3C Gueriau *et al.*, 2015).

The sparse excitation energy XAS approach produced maps of cerium speciation and total concentration allowing comparison of different physiological zones and showing homogeneous speciation, even if the concentration of the trace element varies considerably on the map.

5. Conclusion

X-ray absorption imaging can be used to establish chemical speciation at sub-micrometric to centimetric spatial scales over large sample areas. However, as observed in many samples, including the example discussed here of cerium speciation in fossil paleontological materials, the traditional way of collecting XAS spectra, which requires the accumulation of data for sometimes several minutes per point, can create changes in the redox state of the sample at the points of analysis. With the increased brightness of synchrotron X-ray sources, this approach leads to the deposition of large radiation doses in small sample volumes. Such alterations can occur in many samples and present a significant risk to the quality of the measurement. The procedure presented here with the implementation of the provided software allows to produce speciation and total elemental composition maps by collecting several individual XRF maps across the absorption edge of the element of interest at relatively low total irradiation doses. Although similar procedures have been reported many times in the past, we show with a rigorous framework that the determination of proportions of individual species can be solved accurately, in particular by making explicit the typical assumptions and conditions used when collecting XRF data on thick samples. A particular advantage of this low-dose XAS imaging approach, that we propose to denominate as see-XAS (sparse excitation energy XAS), is that not only can speciation of the sample be obtained at thousands or millions of micro-scale locations to reconstruct high-resolution images of the speciation, but also that the actual dose in a given pixel is much lower (up to several orders of magnitude) owing to the selective nature of the number of energies collected. This constitutes a true risk mitigation procedure that can be invaluable for radiation-sensitive samples in the fields of cultural heritage, environment, life and materials science.

Acknowledgements

SW acknowledges support from the Jean d'Alembert fellowship program at the Université Paris–Saclay. PG thanks D.B. Dutheil, S. Charbonnier, G. Clément, N.-E. Jalil, H. Bourget, B. Khalloufi (MNHN), A. Tourani, F. Khaldoune (Cadi Ayyad University, Marrakesh) and P.M. Brito (Rio de Janeiro State University) for collecting the fossils. E. Janots (Université Grenoble Alpes, Grenoble) is warmly acknowledged for providing us Ce(III) and Ce(IV) standard XANES data. We thank D. Vantelon and N. Trcera (SOLEIL) for their support at the Lucia beamline. Study of heritage materials at IPANEMA and the SOLEIL synchrotron was supported by the Research Infrastructures activity IPERION CH of the Horizon2020 program of the European Commission (Grant Agreement No. 654028). Beamtime was funded by the SOLEIL synchrotron under project no. 20141143. Contribution of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. Fast X-ray imaging research at IPANEMA is supported by Agence nationale de la recherche within the PATRIMEX EquipEx (ANR-11-EQPX-0034).

References

Ade, H., Zhang, X., Cameron, S., Costello, C., Kirz, J. & Williams, S. (1992). *Science*, **258**(5084), 972–975.

Bertrand, L., Cotte, M., Stampanoni, M., Thoury, M., Marone, F. & Schöder, S. (2012). *Physics Reports*, **519**(2), 51–96.

Bertrand, L., Schöder, S., Anglos, D., Breese, M. B. H., Janssens, K., Moini, M. & Simon, A. (2015). *Trends Anal. Chem.* **66**, 128–145.

Bertsch, P. M. & Hunter, D. B. (2001). *Chemical Reviews*, **101**(6), 1809–1842.

Collins, B. A. & Ade, H. (2012). *Journal of Electron Spectroscopy and Related Phenomena*, **185**(5-7), 119–128.

Cotte, M., Susini, J., Dik, J. & Janssens, K. (2010). *Accounts of Chemical Research*, **43**(6), 705–714.

De Andrade, V., Susini, J., Salomé, M., Beraldin, O., Rigault, C., Heymes, T., Lewin, E. & Vidal, O. (2011). *Analytical Chemistry*, **83**(11), 4220–4227.

Edwards, N., Manning, P., Bergmann, U., Larson, P., Van Dongen, B., Sellers, W., Webb, S., Sokaras, D., Alonso-Mori, R., Ignatyev, K. *et al.* (2014). *Metallomics*, **6**(4), 774–782.

Etschmann, B. E., Donner, E., Brugger, J., Howard, D. L., de Jonge, M. D., Paterson, D., Naidu, R., Scheckel, K. G., Ryan, C. G. & Lombi, E. (2014). *Environmental Chemistry*, **11**(3), 341–350.

Farfan, G., Apprill, A., Webb, S. M. & Hansel, C. M. (2018). *Analytical Chemistry*, **90**(21), 12559–12566.

Fayard, B., Pouyet, E., Berruyer, G., Bugnazet, D., Cornu, C., Cotte, M., De Andrade, V., Di Chiaro, F., Hignette, O., Kieffer, J., Martin, T., Papillon, E., Salomé, M. & Solé, V. A. (2013*a*). *Journal of Physics: Conference Series*, **425**(19), 192001.

Fayard, B., Pouyet, E., Berruyer, G., Bugnazet, D., Cornu, C., Cotte, M., De Andrade, V., Di Chiaro, F., Hignette, O., Kieffer, J. *et al.* (2013*b*). *Journal of Physics: Conference Series*, **425**(19), 192001.

Flynn, G., Keller, L., Jacobsen, C. & Wirick, S. (1999). *Meteoritics and Planetary Science Supplement*, **34**, A36.

Gardea-Torresdey, J., Peralta-Videa, J., de la Rosa, G. & Parsons, J. (2005). *Coordination Chemistry Reviews*, **249**(17–18), 1797–1810.

German, C. R. & Elderfield, H. (1990). *Paleoceanography*, **5**(5), 823–833.

Gueriau, P., Bernard, S. & Bertrand, L. (2016). *Elements*, **12**(1), 45–50.

Gueriau, P., Mocuta, C. & Bertrand, L. (2015). *Analytical Chemistry*, **87**(17), 8827–8836.

Henderson, G. S., de Groot, F. M. & Moulton, B. J. (2014). *Reviews in Mineralogy and Geochemistry*, **78**(1), 75–138.

Holt, M., Harder, R., Winarski, R. & Rose, V. (2013). *Annual Review of Materials Research*, **43**, 183–211.

Howells, M., Beetz, T., Chapman, H., Cui, C., Holton, J., Jacobsen, C., Kirz, J., Lima, E., Marchesini, S., Miao, H., Sayre, D., Shapiro, D., Spence, J. & Starodub, D. (2009). *Journal of Electron Spectroscopy and Related Phenomena*, **170**(1–3), 4–12.

Jacobsen, C., Wirick, S., Flynn, G. & Zimba, C. (2000). *Journal of Microscopy*, **197**(2), 173–184.

Jones, E., Oliphant, T., Peterson, P. *et al.*, (2001–). SciPy: Open source scientific tools for Python. `http://www.scipy.org/`.

Kaulich, B., Thibault, P., Gianoncelli, A. & Kiskinova, M. (2011). *Journal of Physics: Condensed Matter*, **23**(8), 083002.

Kinney, J., Johnson, Q., Nichols, M. C., Bonse, U. & Nusshardt, R. (1986). *Applied Optics*, **25**(24), 4583–4585.

Lerotic, M., Jacobsen, C., Schäfer, T. & Vogt, S. (2004). *Ultramicroscopy*, **100**(1-2), 35–57.

Li, L., Chen-Wiegart, Y.-c. K., Wang, J., Gao, P., Ding, Q., Yu, Y.-S., Wang, F., Cabana, J., Wang, J. & Jin, S. (2015). *Nature communications*, **6**, 6883.

Linge, K. L. (2008). *Critical Reviews in Environmental Science and Technology*, **38**(3), 165–196.

Liu, Y., Meirer, F., Wang, J., Requena, G., Williams, P., Nelson, J., Mehta, A., Andrews, J. C. & Pianetta, P. (2012*a*). *Analytical and Bioanalytical Chemistry*, **404**(5), 1297–1301.

Liu, Y., Meirer, F., Williams, P. A., Wang, J., Andrews, J. C. & Pianetta, P. (2012*b*). *Journal of Synchrotron Radiation*, **19**(2), 281–287.

Lombi, E., de Jonge, M. D., Donner, E., Ryan, C. G. & Paterson, D. (2011). *Analytical and Bioanalytical Chemistry*, **400**(6), 1637–44.

Marcus, M. A. (2010). *TrAC Trends in Analytical Chemistry*, **29**(6), 508–517.

Mayhew, L., Webb, S. & Templeton, A. (2011). *Environmental Science and Technology*, **45**(10), 4468–4474.

Meirer, F., Cabana, J., Liu, Y., Mehta, A., Andrews, J. C. & Pianetta, P. (2011). *Journal of Synchrotron Radiation*, **18**(5), 773–781.

Nelson, G. J., Harris, W. M., Izzo Jr, J. R., Grew, K. N., Chiu, W. K., Chu, Y. S., Yi, J., Andrews, J. C., Liu, Y. & Pianetta, P. (2011). *Applied Physics Letters*, **98**(17), 173109.

Oram, L. L., Strawn, D. G. & Möller, G. (2010). *Environmental Science and Technology*, **45**(3), 870–875.

Pickering, I. J., Prince, R. C., Salt, D. E. & George, G. N. (2000). *Proceedings of the National Academy of Sciences*, **97**(20), 10717–10722.

Pushie, M. J., Pickering, I. J., Korbas, M., Hackett, M. J. & George, G. N. (2014). *Chemical Reviews*, **114**(17), 8499–8541.

R Development Core Team, (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rau, C., Somogyi, A., Bytchkov, A. & Simionovici, A. S. (2002). In *Developments in X-ray tomography III*, vol. 4503, pp. 249–255. International Society for Optics and Photonics.

Sciau, P. & Wang, T. (2019). In *Synchrotron Radiation*, chap. 3. IntechOpen.

Sutton, S., Bajt, S., Delaney, J., Schulze, D. & Tokunaga, T. (1995). *Review of Scientific Instruments*, **66**(2), 1464–1467.

Sutton, S. R., Bertsch, P. M., Newville, M., Rivers, M., Lanzirotti, A. & Eng, P. (2002). *Reviews in Mineralogy and Geochemistry*, **49**(1), 429–483.

Tack, P., Garrevoet, J., Bauters, S., Vekemans, B., Laforce, B., Van Ranst, E., Banerjee, D., Longo, A., Bras, W. & Vincze, L. (2014). *Analytical Chemistry*, **86**(17), 8791–8797.

Takahashi, Y., Shimizu, H., Kagi, H., Yoshida, H., Usui, A. & Nomura, M. (2000). *Earth and Planetary Science Letters*, **182**(3-4), 201–207.

VanRossum, G. & Drake, F. L. (2010). *The Python language reference*. Python software foundation Amsterdam, The Netherlands.

Vantelon, D., Trcera, N., Roy, D., Moreno, T., Mailly, D., Guilet, S., Metchalkov, E., Delmotte, F., Lassalle, B., Lagarde, P. & Flank, A.-M. (2016). *Journal of Synchrotron radiation*, **23**(2), 635–640.

Wang, J., Chen-Wiegart, Y.-c. K. & Wang, J. (2013). *Chemical Communications*, **49**(58), 6480–6482.

Webb, S., (2010–). Sam's X-rays – Applications for Synchrotrons. `https://www.sams-xrays.com/`.

Webb, S. (2011). *AIP Conference Proceedings*, **1365**(1), 196–199.

Zhang, X., Ade, H., Jacobsen, C., Kirz, J., Lindaas, S., Williams, S. & Wirick, S. (1994). *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **347**(1-3), 431–435.

Synopsis

A formal framework and two software packages are provided for rapid mapping of speciation using synchrotron X-ray absorption spectroscopy.

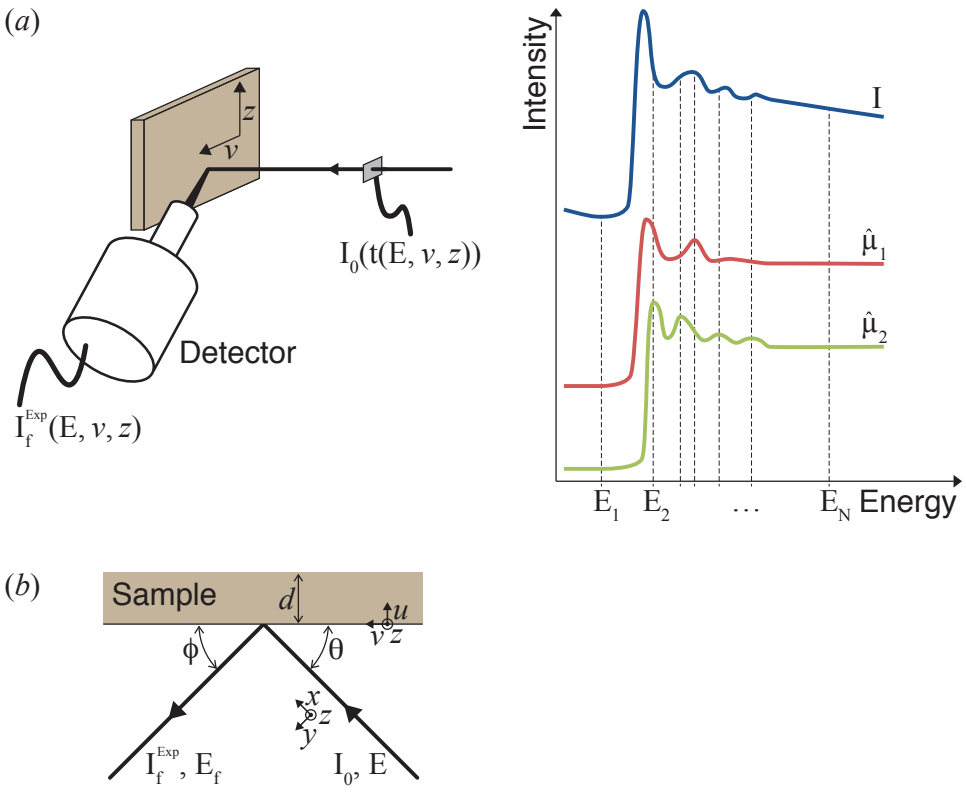

Fig. 1. (a) Graphical illustration of the acquisition process. (b) Schematic of the geometry and of the geometrical notations used.

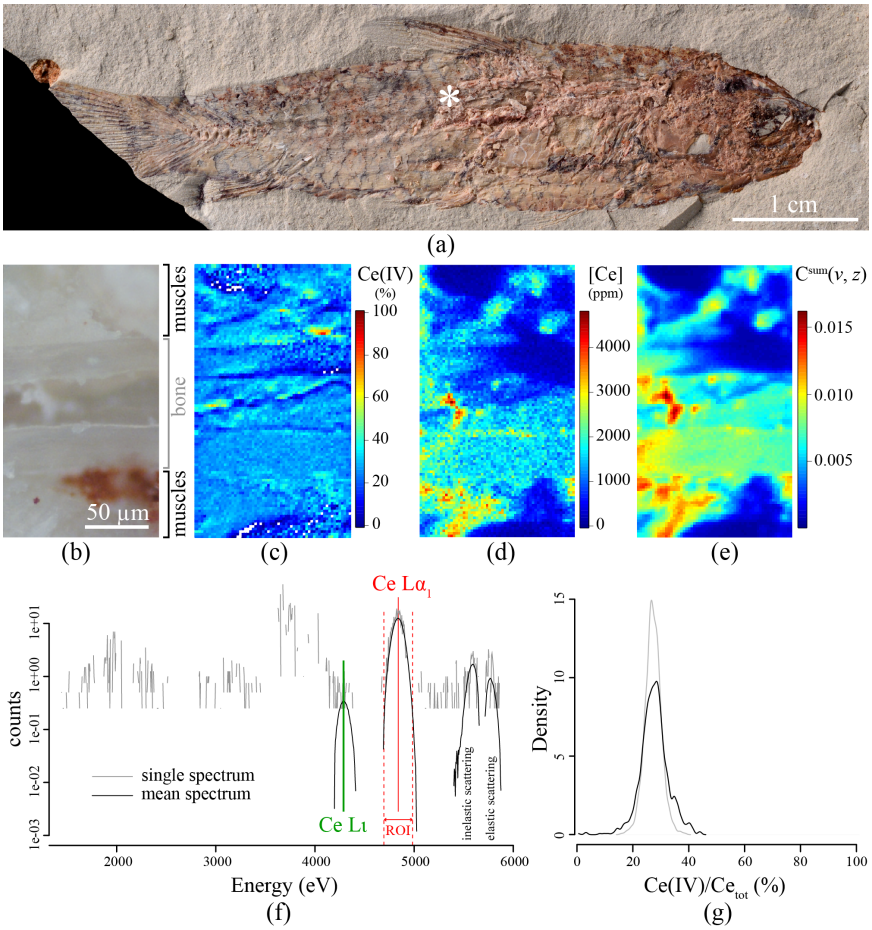

Fig. 2. Mapping of the two speciation states of cerium in a well-preserved Cretaceous teleost ?gonorynchiform fossil fish from Morocco (MHNM-KK-OT 10, ca. 100 Myr old). (a) Photograph of the fossil. (b) Light microscopy image of the spot denoted by a star in (a). (c) Corresponding [Ce(IV)]/[Ce(tot)] ratio (scan step: $2\times2\,\mu m^2$, 6466 pixels). (d) Total cerium content [Ce(tot)] determined from post-edge spectral decomposition (5900 eV). (e) $C^{sum}(v, z)$ map. Note the similarity and improved signal-to-noise with respect to the cerium map in (d). (f) µXRF spectrum from a single spot and mean spectrum from the full map, calculated from subtraction of the spectra collected at 5726 eV to that collected before the edge at 5680 eV. The difference spectrum reveals the specific contribution of the excited fluorescence lines from cerium. (g) Kernel density estimates of the [Ce(IV)]/[Ce(tot)] distribution from areas corresponding respectively to fossilised muscles (black curve; 915 pixels, bandwidth: 0.009) and to bone (grey curve; 915 pixels, bandwidth: 0.006). Data collected at the Lucia beamline of the SOLEIL synchrotron.

Table 1. *Notations used in the article.*

K	Number of species present (indexed by $k \in [\![1, K]\!]$)
N	Number of energies at which measurement is performed (indexed by $n \in [\![1, N]\!]$)
$\mathbf{r}$, dV	Position and volume, respectively, of an elementary volume (*voxel*)
I_0	Incident beam intensity
$\mathrm{I}_\mathrm{f}^\mathrm{Exp}$	Experimentally measured intensity of the fluorescence signal
μ_m	Mass attenuation coefficient of the matrix
l	The homogeneity length of the matrix composition, or a lower bound thereof, used to express the hypotheses of matrix homogeneity both transversely and longitudinally
μ_cm	Apparent *complete mass attenuation coefficient* of the matrix (accounting for the attenuation of both the incident and the fluoresced beams)
l_c	Maximum apparent attenuation length of the matrix, defined as $l_c(\mathrm{E_f}, v, z) = \max_\mathrm{E}(1/\mu_\mathrm{cm}(\mathrm{E}, \mathrm{E_f}, v, z))$
d	Sample thickness
μ_k	Mass attenuation coefficient of the species k of the probed element
$\widehat{\mu_k}$	Normalized reference spectra for the species k
$\mathcal{C}_k$	Concentration of species k
E, E^edge, δE	Incident energy, reference energy and deviation from reference: $\mathrm{E} = \mathrm{E}^\mathrm{edge} + \delta\mathrm{E}$
$\mathrm{E_f}$	Energy of the fluoresced signal from the excited element, energy at which the detection is performed
μ_k^edge	Mass attenuation contribution of excited electron in species k
μ^edge	Mass attenuation contribution of excited electron at energy higher than the edge
ε_f, Ω	Quantum efficiency of the fluorescence detector, and solid angle of detection
$\gamma_{k,\mathrm{E_f}}$, $\gamma_{\mathrm{E_f}}$	Yield of species k, and excited element (irrelevant of the species considered) to produce fluorescence at energy $\mathrm{E_f}$
$\mathcal{A}_m$, $\mathcal{B}_m$	Coefficients in the affine model of μ_m
$\mathcal{A}^\mathrm{edge}$, $\mathcal{B}^\mathrm{edge}$	Coefficients in the affine model of *out of edge* mass attenuation of the excited element, that is contribution to the mass attenuation of non-excited electrons
a^edge, b^edge	Coefficients in the affine model for μ^edge
A, B	Affine model of the XAS baseline on *relative scale*, when using normalized reference spectra
C_0, $\mathrm{C}_{\delta\mathrm{E}}$	Linear regression coefficients, respectively, intercept and δE coefficient
$\mathrm{C_k}$	Linear regression coefficients, relative to each of the K tested species
C^sum	Sum of the $\mathrm{C_k}$; $\mathrm{C}^\mathrm{sum} = \sum_{k\in[\![1,K]\!]} \mathrm{C_k}$
$\mathrm{D_m}$	A proportionality factor depending on the experimental setup, matrix composition and $\mathrm{E_f}$ used to measure the fluorescence signal.