



HAL
open science

Robust framework and software implementation for fast speciation mapping

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

► **To cite this version:**

Serge X. Cohen, Samuel Webb, Pierre Gueriau, Emmanuel Curis, Loïc Bertrand. Robust framework and software implementation for fast speciation mapping. *Journal of Synchrotron Radiation*, 2020, 27 (4), pp.1049-1058. 10.1107/S1600577520005822. hal-02915778

HAL Id: hal-02915778

<https://hal.science/hal-02915778v1>

Submitted on 8 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

1

2 One of the greatest benefits of synchrotron radiation is the ability to perform chemical
3 speciation analysis through X-ray absorption spectroscopies (XAS). XAS imaging of
4 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

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

24 **1. Introduction**

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

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

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

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

82 angles of most fluorescence detectors are limited, and only a small fraction of the
83 total fluorescence signal can be measured. In addition, matrix effects (reabsorption
84 of fluoresced photons) can significantly impact the collected signal. XRF detection is
85 usually strongly count-rate limited, further decreasing yields of detection and signal-
86 to-noise ratios. This methodology has been applied to a wide host of samples, including
87 geological systems to examine transition metal speciation (Sutton *et al.*, 1995; Sutton
88 *et al.*, 2002) and more dilute biological systems (Pushie *et al.*, 2014; Pickering *et al.*,
89 2000; Oram *et al.*, 2010).

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

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

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

130 2. Methods – Sparse excitation energy XAS

131 The following approximations have been made considering that the XAS signal is
132 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 $\mu(\mathbf{E}, \mathbf{r})$, and can be decomposed as:

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

160 where $\mu_m(\mathbf{E}, \mathbf{r})$ is the matrix absorption coefficient that varies slowly with energy \mathbf{E}
 161 and $\mu_k(\mathbf{E})$ 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 \mathbf{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
 $\mathbf{r}_0 = (x_0, y_0, z_0)$:

$$I_f^{\text{Exp}}(\mathbf{E}, \mathbf{r}_0) = I_0(t(\mathbf{E}, v, z)) \frac{\varepsilon_f \Omega}{4\pi} \quad (1)$$

$$\cdot \exp\left(-\int_{x=0}^{x_0} \mu(\mathbf{E}, x, y_0, z_0) dx\right) \quad (2)$$

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

$$\cdot \exp\left(-\int_{y=y_0}^{y_0+x_0 \tan \theta} \mu(\mathbf{E}_f, x_0, y, z_0) dy\right) + \epsilon(\mathbf{E}, v, z) \quad (4)$$

170 Line 1 of this equation corresponds to factors independent from (x_0, y_0, z_0) : the
 171 incident beam flux, $I_0(t(\mathbf{E}, v, z))$, and the detector characteristics, $\frac{\varepsilon_f \Omega}{4\pi}$ that accounts

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

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

178 Line 3 accounts for the quantity of fluorescence emitted by the elementary volume
 179 irradiated by the incident microbeam. It takes into account the fluorescence yield
 180 $(\gamma_{k,E_f})_{k \in \llbracket 1, K \rrbracket}$ of the K species of the target element which edges are probed and their
 181 amount in the given voxel. The latter will be called the *fluorescence source term*.

182 Finally, line 4 of this equation considers that the fluorescence emission at E_f is
 183 attenuated due to reabsorption in the sample on the outgoing path expressed in its
 184 integral form arising from the Beer–Lambert law. Finally we have the error term
 185 $\epsilon(\mathbf{E}, v, z)$ 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, $\mu(\mathbf{E}, \mathbf{r})$ is
 largely dominated by its matrix-related contribution:

$$\mu(\mathbf{E}, \mathbf{r}) \approx \mu_m(\mathbf{E}, \mathbf{r}) \quad (5)$$

186 Under this hypothesis, the amplitude of the attenuation, both on the incident and
 187 fluoresced beam, is made independent of the concentration of the species of the target
 188 element. In this approximation, the absorption of the target element affects only the
 189 *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, $(\mathcal{C}_k)_{k \in \llbracket 1, K \rrbracket}$, 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:

$$\forall \mathbf{r}, \|\mathbf{r} - \mathbf{r}_0\| < l \Rightarrow (\mu_m(\mathbf{E}, \mathbf{r}) \approx \mu_m(\mathbf{E}, \mathbf{r}_0)) \text{ and } (\forall k \in \llbracket 1, K \rrbracket, \mathcal{C}_k(\mathbf{r}) \approx \mathcal{C}_k(\mathbf{r}_0)) \quad (6)$$

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 $\mathcal{C}_k(\mathbf{r})$:

$$\begin{aligned} I_f^{\text{Exp}}(\mathbf{E}, \mathbf{r}) = I_0(t(\mathbf{E}, v, z)) \frac{\varepsilon_f \Omega}{4 \pi} \\ \cdot \exp(-\mu_m(\mathbf{E}, \mathbf{r})x_0) \left(\sum_{k=1}^K \gamma_{k, \mathbf{E}_f} \mathcal{C}_k(\mathbf{r}) \mu_k(\mathbf{E}) \right) \\ \cdot \exp(-\mu_m(\mathbf{E}_f, \mathbf{r})x_0 \tan \theta) \end{aligned} \quad (7)$$

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 $\mu_{\text{cm}}(\mathbf{E}, \mathbf{E}_f, v, z) = \mu_m(\mathbf{E}, v, z) + \tan \theta \mu_m(\mathbf{E}_f, v, z)$, such that the expression of the contribution of a voxel becomes:

$$I_f^{\text{Exp}}(\mathbf{E}, \mathbf{r}) = I_0(t(\mathbf{E}, v, z)) \frac{\varepsilon_f \Omega}{4 \pi} \cdot \exp(-\mu_{\text{cm}}(\mathbf{E}, \mathbf{E}_f, v, z)x_0) \cdot \left(\sum_{k=1}^K \gamma_{k, \mathbf{E}_f} \mathcal{C}_k(v, z) \mu_k(\mathbf{E}) \right) \quad (8)$$

190 $l_c(\mathbf{E}_f, v, z) = \max_{\mathbf{E}}(1/\mu_{\text{cm}}(\mathbf{E}, \mathbf{E}_f, v, z))$ can be interpreted as the characteristic atten-
 191 uation length of the experimental setup, and we need to have $l_c(\mathbf{E}_f, v, z) \ll l$ to satisfy
 192 the hypothesis of an homogeneous matrix and target element species composition.
 193 Furthermore if d , the thickness of sample, satisfies $d \gg l_c(\mathbf{E}_f, v, z)$ we can treat the
 194 sample as infinitely thick. This condition, together with the hypothesis expressed in

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

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

$$dI_f^{\text{Exp}} = I_f^{\text{Exp}}(\mathbf{E}, \mathbf{r})d\mathbf{r} \quad (9)$$

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:

$$dI_f^{\text{Exp}}(\mathbf{E}, v, z) = I_0(t(\mathbf{E}, v, z)) \frac{\varepsilon_f \Omega}{4\pi} \cdot \frac{1}{\mu_{\text{cm}}(\mathbf{E}, \mathbf{E}_f, v, z)} \cdot \left(\sum_{k=1}^K \gamma_{k, \mathbf{E}_f} \mathcal{C}_k(v, z) \mu_k(\mathbf{E}) \right) dv dz \quad (10)$$

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

203 **Reference and normalised spectra.** In the general case corresponding to the
 204 approximations (Eq. 10), the reference spectra allowing to estimate the parameters
 205 $(\varepsilon_f, \mu_{\text{cm}}(\mathbf{E}, \mathbf{E}_f, v, z), (\gamma_{k, \mathbf{E}_f})_{k \in \llbracket 1, K \rrbracket})$ through the knowledge of $(\mu_k(\mathbf{E}))_{k \in \llbracket 1, K \rrbracket}$ must be
 206 measured under the same conditions, the same matrix, the same experimental setup
 207 as the measurements made during the sparse excitation energy XAS experiment. This
 208 is in the best cases unpractical and in most cases unfeasible.

209 However, the X-ray absorption spectra on an energy domain allowing to probe an
 210 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:

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

$$\text{with } |\mathcal{A}_m \times \delta\mathbf{E}| \ll \mathcal{B}_m \quad (12)$$

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

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

211 basically stating that $\mu_{\text{cm}}(\mathbf{E}, \mathbf{E}_f, v, z)$ has very small variation when \mathbf{E} varies within a
212 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, $\gamma_{k, \mathbf{E}_f} \mu_k^{\text{edge}}(\mathbf{E})$:

$$\forall \mathbf{E}, \gamma_{k, \mathbf{E}_f} \mu_k(\mathbf{E}) \approx \mathcal{B}^{\text{edge}} + \mathcal{A}^{\text{edge}} \times \delta\mathbf{E} + \gamma_{\mathbf{E}_f} \mu_k^{\text{edge}}(\mathbf{E}) \quad (14)$$

with

$$\begin{aligned} \forall k \in \llbracket 1, K \rrbracket, E \ll E^{\text{edge}} &\Rightarrow \mu_k^{\text{edge}}(E) = 0 \\ E \gg E^{\text{edge}} &\Rightarrow \mu_k^{\text{edge}}(E) = \mu^{\text{edge}}(E) \approx b^{\text{edge}} + a^{\text{edge}} \times \delta E \end{aligned} \quad (15)$$

In Eq. 15, the term $a^{\text{edge}} \times \delta E$ 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 $E > E^{\text{edge}}$. If $|a^{\text{edge}}| \ll |\mathcal{A}^{\text{edge}}|$, one can, on a small domain that does not extend too much after the edge energy E^{edge} , neglect a^{edge} and therefore write:

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

with

$$\begin{aligned} \forall k \in \llbracket 1, K \rrbracket, E \ll E^{\text{edge}} &\Rightarrow \widehat{\mu}_k(E) = 0 \\ E \gg E^{\text{edge}} &\Rightarrow \widehat{\mu}_k(E) = 1 \end{aligned}$$

213 Eq. 16 defines the $(\widehat{\mu}_k(E))_{k \in \llbracket 1, K \rrbracket}$, the normalised spectra of the K reference species
 214 at the probed edge, which under the current approximations are unaffected by the
 215 matrix used to measure the reference spectra, as long they are measured in fluores-
 216 cence mode at energy E_f . Indeed under the current approximations the dependence
 217 of the measurement to the matrix is encoded within the $\mathcal{A}^{\text{edge}}$ and $\mathcal{B}^{\text{edge}}$ which in
 218 turn are not affected by the species of the probed element, whilst b^{edge} depends only
 219 on the edge probed. In other words, Eq. 16 allows us to model the reference spec-
 220 tra using species-independent terms (parameters: $\mathcal{B}^{\text{edge}}$, $\mathcal{A}^{\text{edge}}$), a matrix-independent
 221 term $((\widehat{\mu}_k(E))_{k \in \llbracket 1, K \rrbracket})$, and a term independent of both $(\gamma_{E_f} b^{\text{edge}})$. In this context, the
 222 $(\widehat{\mu}_k(E))_{k \in \llbracket 1, K \rrbracket}$ can be obtained by measuring the reference species in any matrix and
 223 at any concentration provided that the approximation of the diluted sample holds, as
 224 usually done.

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

Linear combination of reference spectra. We may now rewrite Eq. 10 using the approximation from Eq. 13 for $\mu_{\text{cm}}(\text{E}, \text{E}_f, v, z)$ and Eq. 16 for $(\mu_k(\text{E}))_{k \in \llbracket 1, K \rrbracket}$:

$$\frac{I_f^{\text{Exp}}(\text{E}, v, z)}{I_0(t(\text{E}, v, z))} = \frac{\varepsilon_f \Omega}{4\pi} \cdot \frac{1}{(1 + \tan \theta) \mathcal{B}_m(v, z)} \left[\left(\mathcal{B}^{\text{edge}} + \mathcal{A}^{\text{edge}} \times \delta \text{E} \right) \cdot \left(\sum_{k=1}^K \mathcal{C}_k(v, z) \right) + \gamma_{\text{E}_f} b^{\text{edge}} \left(\sum_{k=1}^K \mathcal{C}_k(v, z) \widehat{\mu}_k(\text{E}) \right) \right] \quad (17)$$

227 Which states that for a given (v, z) position the measured signal, $\frac{I_f^{\text{Exp}}(\text{E}, v, z)}{I_0(t(\text{E}, v, z))}$ is an
228 affine function of the $(\widehat{\mu}_k(\text{E}))_{k \in \llbracket 1, K \rrbracket}$ and δE with coefficients:

- 229 • $\text{C}_0(v, z) = \text{D}_m(\text{E}_f, v, z) \mathcal{B}^{\text{edge}} \left(\sum_{k=1}^K \mathcal{C}_k(v, z) \right)$ for the intercept;
- 230 • $\text{C}_{\delta \text{E}}(v, z) = \text{D}_m(\text{E}_f, v, z) \mathcal{A}^{\text{edge}} \left(\sum_{k=1}^K \mathcal{C}_k(v, z) \right)$ for the coefficient of δE ;
- 231 • $\text{C}_k(v, z) = (\text{D}_m(\text{E}_f, v, z) \gamma_{\text{E}_f} b^{\text{edge}} \mathcal{C}_k(v, z))_{k \in \llbracket 1, K \rrbracket}$ for the coefficient of each of the
232 $(\widehat{\mu}_k(\text{E}))_{k \in \llbracket 1, K \rrbracket}$,

233 where $\text{D}_m(\text{E}_f, v, z) = \frac{\varepsilon_f \Omega}{4\pi} \cdot \frac{1}{(1 + \tan \theta) \mathcal{B}_m(v, z)}$ is independent of the incident energy E .

234 For any given coordinate set (v, z) on the sample, one can estimate the above
235 expressed coefficient, $(\text{C}_0(v, z), \text{C}_{\delta \text{E}}(v, z), (\text{C}_k(v, z))_{k \in \llbracket 1, K \rrbracket})$, by performing a linear
236 least square fitting of $(N \geq K + 2)$ measurements corresponding to energies $(\text{E}_n)_{n \in \llbracket 1, N \rrbracket}$
237 such that the vector $(\widehat{\mu}_k(\text{E}_1), \dots, \widehat{\mu}_k(\text{E}_N))_{k \in \llbracket 1, K \rrbracket}$ and $(\delta \text{E}_1, \dots, \delta \text{E}_N)$ are linearly indepen-
238 dent. The estimated coefficients $(\text{C}_k(v, z))_{k \in \llbracket 1, K \rrbracket}$ correspond indeed to the relative
239 concentration of each of the K species that could be present in the samples at coordi-
240 nates (v, z) hence providing a relative estimation of the $(\mathcal{C}_k(v, z))_{k \in \llbracket 1, K \rrbracket}$.

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

247 which leads to Eq. 20 in the main text.

248 **Validation of the model and estimation of the relative accuracy.** Under
 249 the model developed above, as long as the matrix of the sample is homogeneous,
 250 the estimated coefficients $C_0(v, z)$ and $C_{\delta E}(v, z)$ should be strictly correlated, within
 251 measurement noise error, as they both should be correlated to the estimated sum of
 252 relative concentrations, $C^{\text{sum}}(v, z) = \sum_{k=1}^K C_k(v, z)$. Indeed such correlations are very
 253 unlikely to be present by chance and the quality of the correlations can be used as a
 254 marker of the validity of the proposed model. Furthermore, as the proposed method
 255 is used to perform speciation imaging, hence collecting data on a large set of (v, z)
 256 coordinates, these correlations can be verified on the basis of a significant number of
 257 data points.

258 Note, however, that if the hypothesis of the homogeneity of the matrix is not met
 259 anymore only the $C_0(v, z)$ and $C_{\delta E}(v, z)$ estimation are affected since only $\mathcal{A}^{\text{edge}}$ and
 260 $\mathcal{B}^{\text{edge}}$ depend on the matrix. In these circumstances $\mathcal{A}^{\text{edge}}$ and $\mathcal{B}^{\text{edge}}$ become functions
 261 of the probed coordinates, (v, z) , and the correlations that are described above no
 262 longer stand. Then $C^{\text{sum}}(v, z)$ is the only available estimate of the relative concentra-
 263 tion of the probed element at the different measured coordinates.

Finally, within the framework of linear least square, one can compute the condi-
 tion number, κ , of the design matrix as soon as the $(\widehat{\mu}_k(\mathbf{E}_1), \dots, \widehat{\mu}_k(\mathbf{E}_N))_{k \in [1, K]}$ and
 $(\delta \mathbf{E}_1, \dots, \delta \mathbf{E}_N)$ 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 con-
 centrations, ϵ_{C_k} , given a level of relative errors, also termed signal to noise ratio, in
 fluorescence signal measurements:

$$\epsilon_{C_k} \leq \kappa \epsilon_{\text{fluorescence}} \quad (18)$$

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

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

Under such conditions, the experimental XAS spectrum at any given pixel of coordi-
 nates (v, z) results from a linear combination of the K reference spectra $(\mu_k(\mathbf{E}))_{k \in \llbracket 1, K \rrbracket}$
 and of an affine baseline:

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

268 where $I_f^{\text{Exp}}(\mathbf{E}, v, z)$ is the fluorescence signal measured when illuminating pixel (v, z) at
 269 an excitation energy \mathbf{E} and $I_0(t(\mathbf{E}, v, z))$ is the incident beam intensity at the moment
 270 $t(\mathbf{E}, v, z)$, when $I_f^{\text{Exp}}(\mathbf{E}, v, z)$ is measured. $\mathcal{A}(v, z) \times \mathbf{E} + \mathcal{B}(v, z)$ models the summed up
 271 slow varying background from XAS spectra of the sample and of the references and
 272 $\mathcal{C}_k(v, z) dV$ is the quantity of the species k in the elementary volume dV probed. Γ is
 273 a proportionality constant depending on the XRF yield of the probed element and on
 274 the geometry of the setup (solid angle of the detector, etc.), $\mu_{\text{cm}}(v, z)$ is the apparent
 275 absorption coefficient of the matrix, accounting for both incident beam and fluoresced
 276 signal attenuation, at pixel (v, z) and $\epsilon(\mathbf{E}, v, z)$ is a zero mean stochastic error term
 277 which is independent from one measurement to the next.

The parameters $(\mathcal{C}_k(v, z))_{k \in \llbracket 1, K \rrbracket}$, $\mathcal{A}(v, z)$ 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 $\epsilon(\mathbf{E}, v, z)$, these estimates are unbiased; with the additional
 assumption of independence and constant variance, they correspond to maximum like-
 lihood estimation. From now, we will simplify the notations by keeping the error term
 out of all expressions. Note that by normalising the reference spectra, $(\widehat{\mu}_k(\mathbf{E}))_{k \in \llbracket 1, K \rrbracket}$,
 we bring \mathcal{C}_k on an arbitrary unit and as such there is not much interest in obtaining
 \mathcal{C}_k rather than $\mathcal{C}_k \propto \frac{\Gamma}{\mu_{\text{cm}}(v, z)} \mathcal{C}_k dV$. Finally the very limited dependence that $\mu_{\text{cm}}(v, z)$

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

$$\frac{I_f^{\text{Exp}}(\mathbf{E}, v, z)}{I_0(t(\mathbf{E}, v, z))} \approx \left(\sum_{k=1}^K C_k(v, z) \widehat{\mu}_k(\mathbf{E}) \right) + A(v, z) \times E + B(v, z) \quad (20)$$

For a given (v, z) pixel position and using N probing energies $(E_n)_{n \in \llbracket 1, N \rrbracket}$ equation (20) produces a linear system of N equations and $K + 2$ unknowns ($C_k(v, z)$, $A(v, z)$ and $B(v, z)$). 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, $\sum_{k=1}^K C_k(v, z) = C^{\text{sum}}(v, z)$, 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,

299 in such a way that it is easy for a user to adapt it to its own data. The script starts
300 from raw XRF data and covers all steps up to decomposition and statistical analysis.
301 In particular several diagnosis steps are present in this script, as a contribution to
302 data validation, while not all described in this manuscript. The script and all datasets
303 needed to reproduce the presented results can be found at [https://doi.org/10.](https://doi.org/10.5281/zenodo.3748752)
304 [5281/zenodo.3748752](https://doi.org/10.5281/zenodo.3748752).

305 **4. Application to a fossil sample**

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

317 We studied a loose fragment (not a cross-section) of well preserved soft tissues
318 (mineralized muscles) around a bone of a yet undescribed ?gonorynchiform fish (where
319 "?" indicates that the supposed identification has not yet been corroborated by a
320 complete anatomical description) coming from the Late Cretaceous levels of the Djebel
321 Oum Tkout locality in Morocco (ca. 95 million years old, Fig. 2a,b). We produced
322 rapid maps of cerium speciation of the sample at the Lucia beamline of the SOLEIL
323 synchrotron source (Vantelon *et al.*, 2016) over a 6466 pixel area ($120 \times 210 \mu\text{m}^2$ with

324 a scan step of $2 \times 2 \mu\text{m}^2$; beam size $3.0 \times 2.5 \mu\text{m}^2$ (H×V); flux: 2.8×10^{10} photons/s
325 in the focused spot) using the sparse excitation energy procedure presented above.
326 Spectra of cerium III and IV reference compounds were collected using the procedure
327 indicated in Gueriau *et al.* (2015).

328 Gueriau *et al.* (2015) studied a series of cross sections from well-preserved fossil
329 teleost fishes and crustaceans from the same locality (see Fig. 4 in Gueriau *et al.*, 2015)
330 using a limited number of excitation energy points to limit photo-oxidation changes
331 they observed within a few tens seconds during XAS acquisition. They selected six
332 excitation energies: below the L_3 absorption edge of cerium (5680 eV), in the post-edge
333 region (5765 and 5900 eV) and at energies corresponding to three main spectral fea-
334 tures that allow distinguishing Ce(III) from Ce(IV) in cerium-containing compounds
335 — a white line at 5726 eV resulting from the $2p \rightarrow (4f^1)5d^*$ electron transition and
336 two peaks that partly overlap with the former attributed to the $2p_{3/2} \rightarrow (4f^1L)5d^*$
337 and the $2p_{3/2} \rightarrow (4f^0)5d^*$ transitions, where L denotes a hole in ligand orbitals (at
338 5733 and 5736 eV respectively) (Takahashi *et al.*, 2000).

339 [Figure 2 about here.]

340 In the loose soft tissue fragment of the ?gonorynchiform fish, the procedure devel-
341 oped here (using the same excitation energies) reveals the presence of a fraction of
342 cerium in the redox +IV state ($[\text{Ce(IV)}]/[\text{Ce(tot)}] = 25.6 \pm 8.3\%$) over the whole
343 mapped area, suggesting slightly oxidative local conditions of burial (Fig. 2c–f). This
344 value agrees well with the indirect data obtained from micro- and macroscale ele-
345 mental concentration profiles measured at the site (so-called ‘REE patterns’), which
346 also display a slight negative cerium anomaly attributed to oxidative burial condi-
347 tions (Gueriau *et al.*, 2015). The total cerium concentration, $[\text{Ce(tot)}]$, was mapped
348 in the sample using two different approaches. The cerium distribution shown in Fig.
349 2d is the result of the usual processing of an XRF dataset collected after the cerium

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

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

372 The sparse excitation energy XAS approach produced maps of cerium speciation and
373 total concentration allowing comparison of different physiological zones and showing
374 homogeneous speciation, even if the concentration of the trace element varies consid-
375 erably on the map.

5. Conclusion

376

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

401 **Acknowledgements**

402 SW acknowledges support from the Jean d’Alembert fellowship program at the
403 Université Paris–Saclay. PG thanks D.B. Dutheil, S. Charbonnier, G. Clément, N.-E.
404 Jalil, H. Bourget, B. Khalloufi (MNHN), A. Tourani, F. Khaldoune (Cadi Ayyad Uni-
405 versity, Marrakesh) and P.M. Brito (Rio de Janeiro State University) for collecting
406 the fossils. E. Janots (Université Grenoble Alpes, Grenoble) is warmly acknowledged
407 for providing us Ce(III) and Ce(IV) standard XANES data. We thank D. Vantelon
408 and N. Trcera (SOLEIL) for their support at the Lucia beamline. Study of heritage
409 materials at IPANEMA and the SOLEIL synchrotron was supported by the Research
410 Infrastructures activity IPERION CH of the Horizon2020 program of the European
411 Commission (Grant Agreement No. 654028). Beamtime was funded by the SOLEIL
412 synchrotron under project no. 20141143. Contribution of the Stanford Synchrotron
413 Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the
414 U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under
415 Contract No. DE-AC02-76SF00515. Fast X-ray imaging research at IPANEMA is sup-
416 ported by Agence nationale de la recherche within the PATRIMEX EquipEx (ANR-
417 11-EQPX-0034).

418 **References**

- 419 Ade, H., Zhang, X., Cameron, S., Costello, C., Kirz, J. & Williams, S. (1992). *Science*,
420 **258**(5084), 972–975.
- 421 Bertrand, L., Cotte, M., Stampanoni, M., Thoury, M., Marone, F. & Schöder, S. (2012).
422 *Physics Reports*, **519**(2), 51–96.
- 423 Bertrand, L., Schöder, S., Anglos, D., Breese, M. B. H., Janssens, K., Moini, M. & Simon, A.
424 (2015). *Trends Anal. Chem.* **66**, 128–145.
- 425 Bertsch, P. M. & Hunter, D. B. (2001). *Chemical Reviews*, **101**(6), 1809–1842.
- 426 Collins, B. A. & Ade, H. (2012). *Journal of Electron Spectroscopy and Related Phenomena*,
427 **185**(5-7), 119–128.
- 428 Cotte, M., Susini, J., Dik, J. & Janssens, K. (2010). *Accounts of Chemical Research*, **43**(6),
429 705–714.
- 430 De Andrade, V., Susini, J., Salomé, M., Beraldin, O., Rigault, C., Heymes, T., Lewin, E. &
431 Vidal, O. (2011). *Analytical Chemistry*, **83**(11), 4220–4227.
- 432 Edwards, N., Manning, P., Bergmann, U., Larson, P., Van Dongen, B., Sellers, W., Webb, S.,
433 Sokaras, D., Alonso-Mori, R., Ignatyev, K. *et al.* (2014). *Metallomics*, **6**(4), 774–782.

- 434 Etschmann, B. E., Donner, E., Brugger, J., Howard, D. L., de Jonge, M. D., Paterson, D.,
435 Naidu, R., Scheckel, K. G., Ryan, C. G. & Lombi, E. (2014). *Environmental Chemistry*,
436 **11**(3), 341–350.
- 437 Farfan, G., Apprill, A., Webb, S. M. & Hansel, C. M. (2018). *Analytical Chemistry*, **90**(21),
438 12559–12566.
- 439 Fayard, B., Pouyet, E., Berruyer, G., Bugnazet, D., Cornu, C., Cotte, M., De Andrade, V.,
440 Di Chiaro, F., Hignette, O., Kieffer, J., Martin, T., Papillon, E., Salomé, M. & Solé, V. A.
441 (2013a). *Journal of Physics: Conference Series*, **425**(19), 192001.
- 442 Fayard, B., Pouyet, E., Berruyer, G., Bugnazet, D., Cornu, C., Cotte, M., De Andrade, V.,
443 Di Chiaro, F., Hignette, O., Kieffer, J. *et al.* (2013b). *Journal of Physics: Conference*
444 *Series*, **425**(19), 192001.
- 445 Flynn, G., Keller, L., Jacobsen, C. & Wirick, S. (1999). *Meteoritics and Planetary Science*
446 *Supplement*, **34**, A36.
- 447 Gardea-Torresdey, J., Peralta-Videa, J., de la Rosa, G. & Parsons, J. (2005). *Coordination*
448 *Chemistry Reviews*, **249**(17–18), 1797–1810.
- 449 German, C. R. & Elderfield, H. (1990). *Paleoceanography*, **5**(5), 823–833.
- 450 Gueriau, P., Bernard, S. & Bertrand, L. (2016). *Elements*, **12**(1), 45–50.
- 451 Gueriau, P., Mocuta, C. & Bertrand, L. (2015). *Analytical Chemistry*, **87**(17), 8827–8836.
- 452 Henderson, G. S., de Groot, F. M. & Moulton, B. J. (2014). *Reviews in Mineralogy and*
453 *Geochemistry*, **78**(1), 75–138.
- 454 Holt, M., Harder, R., Winarski, R. & Rose, V. (2013). *Annual Review of Materials Research*,
455 **43**, 183–211.
- 456 Howells, M., Beetz, T., Chapman, H., Cui, C., Holton, J., Jacobsen, C., Kirz, J., Lima, E.,
457 Marchesini, S., Miao, H., Sayre, D., Shapiro, D., Spence, J. & Starodub, D. (2009). *Journal*
458 *of Electron Spectroscopy and Related Phenomena*, **170**(1–3), 4–12.
- 459 Jacobsen, C., Wirick, S., Flynn, G. & Zimba, C. (2000). *Journal of Microscopy*, **197**(2),
460 173–184.
- 461 Jones, E., Oliphant, T., Peterson, P. *et al.*, (2001–). SciPy: Open source scientific tools for
462 Python. <http://www.scipy.org/>.
- 463 Kaulich, B., Thibault, P., Gianoncelli, A. & Kiskinova, M. (2011). *Journal of Physics: Con-*
464 *densed Matter*, **23**(8), 083002.
- 465 Kinney, J., Johnson, Q., Nichols, M. C., Bonse, U. & Nusshardt, R. (1986). *Applied Optics*,
466 **25**(24), 4583–4585.
- 467 Lerotic, M., Jacobsen, C., Schäfer, T. & Vogt, S. (2004). *Ultramicroscopy*, **100**(1-2), 35–57.
- 468 Li, L., Chen-Wiegart, Y.-c. K., Wang, J., Gao, P., Ding, Q., Yu, Y.-S., Wang, F., Cabana, J.,
469 Wang, J. & Jin, S. (2015). *Nature communications*, **6**, 6883.
- 470 Linge, K. L. (2008). *Critical Reviews in Environmental Science and Technology*, **38**(3), 165–
471 196.
- 472 Liu, Y., Meirer, F., Wang, J., Requena, G., Williams, P., Nelson, J., Mehta, A., Andrews, J. C.
473 & Pianetta, P. (2012a). *Analytical and Bioanalytical Chemistry*, **404**(5), 1297–1301.
- 474 Liu, Y., Meirer, F., Williams, P. A., Wang, J., Andrews, J. C. & Pianetta, P. (2012b). *Journal*
475 *of Synchrotron Radiation*, **19**(2), 281–287.
- 476 Lombi, E., de Jonge, M. D., Donner, E., Ryan, C. G. & Paterson, D. (2011). *Analytical and*
477 *Bioanalytical Chemistry*, **400**(6), 1637–44.
- 478 Marcus, M. A. (2010). *TrAC Trends in Analytical Chemistry*, **29**(6), 508–517.
- 479 Mayhew, L., Webb, S. & Templeton, A. (2011). *Environmental Science and Technology*,
480 **45**(10), 4468–4474.
- 481 Meirer, F., Cabana, J., Liu, Y., Mehta, A., Andrews, J. C. & Pianetta, P. (2011). *Journal of*
482 *Synchrotron Radiation*, **18**(5), 773–781.
- 483 Nelson, G. J., Harris, W. M., Izzo Jr, J. R., Grew, K. N., Chiu, W. K., Chu, Y. S., Yi, J.,
484 Andrews, J. C., Liu, Y. & Pianetta, P. (2011). *Applied Physics Letters*, **98**(17), 173109.

- 485 Oram, L. L., Strawn, D. G. & Möller, G. (2010). *Environmental Science and Technology*,
486 **45**(3), 870–875.
- 487 Pickering, I. J., Prince, R. C., Salt, D. E. & George, G. N. (2000). *Proceedings of the National*
488 *Academy of Sciences*, **97**(20), 10717–10722.
- 489 Pushie, M. J., Pickering, I. J., Korbas, M., Hackett, M. J. & George, G. N. (2014). *Chemical*
490 *Reviews*, **114**(17), 8499–8541.
- 491 R Development Core Team, (2011). *R: A Language and Environment for Statistical Comput-*
492 *ing*. R Foundation for Statistical Computing, Vienna, Austria.
- 493 Rau, C., Somogyi, A., Bytchkov, A. & Simionovici, A. S. (2002). In *Developments in X-ray*
494 *tomography III*, vol. 4503, pp. 249–255. International Society for Optics and Photonics.
- 495 Sciau, P. & Wang, T. (2019). In *Synchrotron Radiation*, chap. 3. IntechOpen.
- 496 Sutton, S., Bajt, S., Delaney, J., Schulze, D. & Tokunaga, T. (1995). *Review of Scientific*
497 *Instruments*, **66**(2), 1464–1467.
- 498 Sutton, S. R., Bertsch, P. M., Newville, M., Rivers, M., Lanzirrotti, A. & Eng, P. (2002).
499 *Reviews in Mineralogy and Geochemistry*, **49**(1), 429–483.
- 500 Tack, P., Garrevoet, J., Bauters, S., Vekemans, B., Laforce, B., Van Ranst, E., Banerjee, D.,
501 Longo, A., Bras, W. & Vincze, L. (2014). *Analytical Chemistry*, **86**(17), 8791–8797.
- 502 Takahashi, Y., Shimizu, H., Kagi, H., Yoshida, H., Usui, A. & Nomura, M. (2000). *Earth and*
503 *Planetary Science Letters*, **182**(3-4), 201–207.
- 504 VanRossum, G. & Drake, F. L. (2010). *The Python language reference*. Python software foun-
505 dation Amsterdam, The Netherlands.
- 506 Vantelon, D., Trcera, N., Roy, D., Moreno, T., Mailly, D., Guilet, S., Metchalkov, E., Delmotte,
507 F., Lassalle, B., Lagarde, P. & Flank, A.-M. (2016). *Journal of Synchrotron radiation*,
508 **23**(2), 635–640.
- 509 Wang, J., Chen-Wiegart, Y.-c. K. & Wang, J. (2013). *Chemical Communications*, **49**(58),
510 6480–6482.
- 511 Webb, S., (2010–). Sam’s X-rays – Applications for Synchrotrons. <https://www.sams-xrays.com/>.
- 512
- 513 Webb, S. (2011). *AIP Conference Proceedings*, **1365**(1), 196–199.
- 514 Zhang, X., Ade, H., Jacobsen, C., Kirz, J., Lindaas, S., Williams, S. & Wirick, S. (1994).
515 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spec-*
516 *trometers, Detectors and Associated Equipment*, **347**(1-3), 431–435.

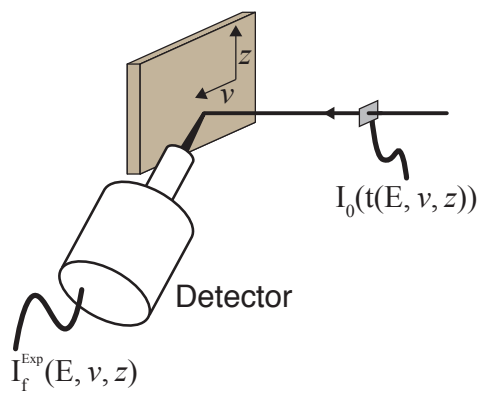
Synopsis

517

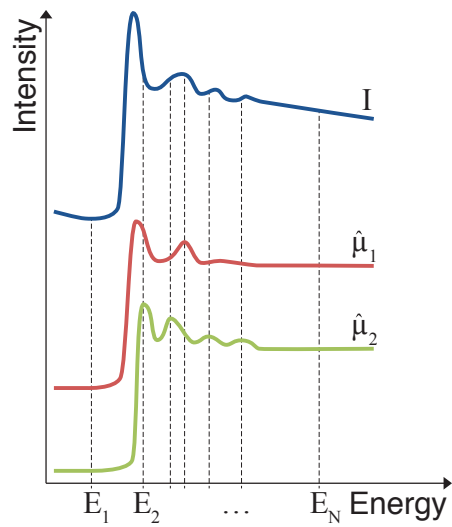
A formal framework and two software packages are provided for rapid mapping of speciation using synchrotron X-ray absorption spectroscopy.

518

(a)



519



(b)

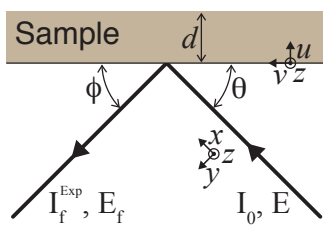
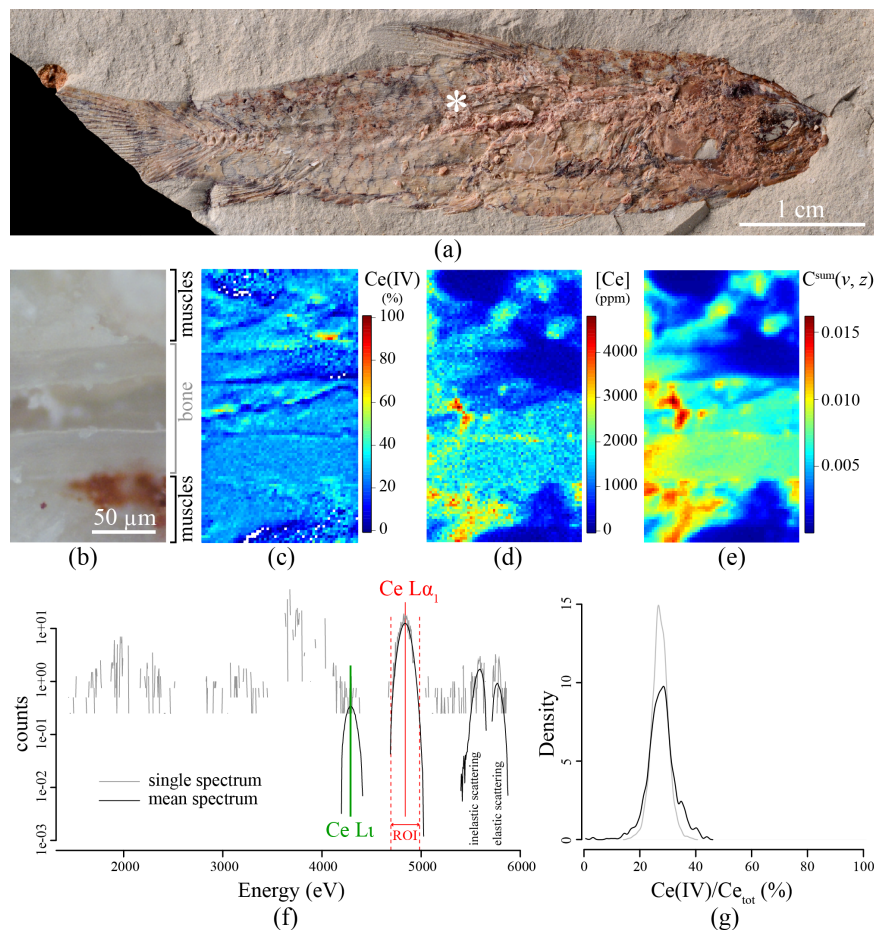


Fig. 1. (a) Graphical illustration of the acquisition process. (b) Schematic of the geometry and of the geometrical notations used.

520

521



522

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: $2 \times 2 \mu\text{m}^2$, 6466 pixels). (d) Total cerium content [Ce(tot)] determined from post-edge spectral decomposition (5900 eV). (e) $C^{\text{sum}}(v, z)$ map. Note the similarity and improved signal-to-noise with respect to the cerium map in (d). (f) μ XRF 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.

523

Table 1. *Notations used in the article.*

K	Number of species present (indexed by $k \in \llbracket 1, K \rrbracket$)
N	Number of energies at which measurement is performed (indexed by $n \in \llbracket 1, N \rrbracket$)
\mathbf{r}, dV	Position and volume, respectively, of an elementary volume (<i>voxel</i>)
I_0	Incident beam intensity
I_f^{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 <i>complete mass attenuation coefficient</i> 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 $l_c(E_f, v, z) = \max_E(1/\mu_{cm}(E, E_f, v, z))$
d	Sample thickness
μ_k	Mass attenuation coefficient of the species k of the probed element
$\hat{\mu}_k$	Normalized reference spectra for the species k
C_k	Concentration of species k
$E, E^{\text{edge}}, \delta E$	Incident energy, reference energy and deviation from reference: $E = E^{\text{edge}} + \delta E$
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
$\gamma_{k, E_f}, \gamma_{E_f}$	Yield of species k , and excited element (irrelevant of the species considered) to produce fluorescence at energy E_f
$\mathcal{A}_m, \mathcal{B}_m$	Coefficients in the affine model of μ_m
$\mathcal{A}^{\text{edge}}, \mathcal{B}^{\text{edge}}$	Coefficients in the affine model of <i>out of edge</i> mass attenuation of the excited element, that is contribution to the mass attenuation of non-excited electrons
$a^{\text{edge}}, b^{\text{edge}}$	Coefficients in the affine model for μ^{edge}
A, B	Affine model of the XAS baseline on <i>relative scale</i> , when using normalized reference spectra
$C_0, C_{\delta E}$	Linear regression coefficients, respectively, intercept and δE coefficient
C_k	Linear regression coefficients, relative to each of the K tested species
C^{sum}	Sum of the C_k ; $C^{\text{sum}} = \sum_{k \in \llbracket 1, K \rrbracket} C_k$
D_m	A proportionality factor depending on the experimental setup, matrix composition and E_f used to measure the fluorescence signal.

524