Chapter 5.18

Statistical analysis in XANES spectroscopy

Anna Kubacka* and Marcos Fernández-García*

Instituto de Catálisis y Petroleoquímica, CSIC, Calle de Marie Curie 2, 28049 Madrid, Spain. *Correspondence e-mail: ak@icp.csic.es, mfg@icp.csic.es

Statistical procedures to analyse sets of X-ray absorption near-edge structure (XANES) spectra are briefly described. Correlation and factor-analysis procedures are among the most widely used. Representative examples of their application to XANES analysis in a broad range of scientific disciplines are discussed.

1. Brief introduction

As detailed in general references and in previous chapters of this book, X-ray absorption near-edge structure (XANES) is a rather powerful element-specific X-ray absorption technique that renders useful local electronic and geometric information for each element in a material (van Bokhoven & Lamberti, 2016). XANES spectroscopy can render information in almost all experimental conditions whatever the temperature, the pressure, the surrounding medium (gas, liquid or solid phase) of the element or any other environmental variable. A critical issue for this chapter is that a XANES spectrum can routinely be acquired at synchrotrons with an adequate signal-to-noise ratio in less than a few seconds (micro-XANES would take an order of magnitude more time), allowing the collection of large set(s) of spectra which would naturally be submitted to statistical analysis to extract information. This is almost mandatory in the case of multiple local geometries of the element subjected to study. Generally speaking, the study of multiple local geometries of a chemical element using a set of XANES spectra involving either different local environments in a single material, different materials (samples), different (spatial) positions within a material and/or different experimental conditions (or in other words, temporal variation in response to an external perturbation) appear to be ideal problems for the application of statistical tools. In addition, sets of XANES data considering different absorption edges of an element and/or different experimental techniques (including XANES as one of them) may become fruitful subjects for specific statistical procedures.

In general, the application of statistical tools is not limited by any property of the material subjected to analysis and can be applied to any problem as long as the information contained in the set of spectra is sufficient to render useful physical or chemical information. To this end, statistical tools can be divided into two main groups. One group is led by correlation analysis, a tool of general application in spectroscopy. The second group considers the use of so-called factor or principal component analysis and relies on the fact that the absorbance of a XANES spectrum can be expressed as a linear combination of the different local geometries of the element under consideration plus noise.

Related chapters

2. Correlation analysis

In spectroscopy, two-dimensional (2D) correlation analysis is the most broadly used technique based on the mathematical concept of correlation. The use of two dimensions to describe spectral features simplifies the detection of events. The (j, k) element of a 2D (synchronous or asynchronous) spectrum for a set of n XANES spectra (X, where each spectrum is defined as the normalized absorbance as a function of energy ε) can be expressed as

$$C(\varepsilon_j, \varepsilon_k) = \frac{A}{(n-1)\sigma(\varepsilon_j)\sigma(\varepsilon_k)} Y^{\mathrm{T}}(\varepsilon_j) \cdot Y(\varepsilon_k), \qquad (1)$$

where Y^{T} and Y are the row (transposed) and column vector of the set of data Y obtained from X by subtracting a reference spectrum, often the average spectrum of the data. In equation (1) $\sigma(\varepsilon_{j,k})$ are the corresponding standard deviations and A is a constant that takes values of 1 for synchronous analysis, 0 for asynchronous analysis if $j = k$, and $1/[\pi(k - j)]$ for asynchronous analysis if $j \neq k$. In the definition of the constant A, j and k are row and column numbers, respectively, that take values from 1 to n that are obtained during the so-called perturbation that occurs through the series and leads to changes in the XANES signal (Noda & Ozaki, 2004). This formulation can easily be generalized to correlate XANES data sets corresponding to different absorption edges (the same or different elements of a material) or to correlate a XANES set of spectra with any other data set coming from other experimental technique if obtained synchronously.

Thus, the synchronous 2D spectrum from a set of XANES spectra X displays the autocorrelation function of spectral intensities. It is symmetric relative to the diagonal, which contains positive cross-peaks. Positive and negative off-diagonal cross-peaks indicate that the changes at the two energies ($\varepsilon_1, \varepsilon_2$) occur in the same and in opposite directions, respectively. The asynchronous 2D spectrum allows analysis of the sequence of events occurring in the set of XANES spectra. If the change at one energy precedes the change at another (higher) energy, the cross-peak would have the same sign as the corresponding synchronous peak. If the change at one energy follows the change at the other (higher) energy, the asynchronous cross-peak has the opposite sign to the synchronous peak.

Application of 2D correlation analysis to XANES differs to some extent from more conventional applications, which are mostly connected with vibrational spectroscopies. In the latter case, the main issue to solve is to distinguish between peak overlapping and shifting, while the main point for XANES is to correlate potential changes taking place in different regions (pre-edge, edge and continuum resonances) of the spectrum. A representative example of application involves analysis of the temperature behaviour of metal (La, Gd) endohedral C_{82} fullerenes. The authors analysed the thermal evolution of the $C2_v$ and C_s isomers of these materials from low temperature (35 K) and found that the former showed subtle changes involving metal-to-ligand hybridization before structural changes occurred. On the other hand, such changes dominate the XANES thermal behaviour in the case of the C_s materials (Marcelli et al., 2009). The most important limitations in 2D correlation analysis involve cases in which the changes in the XANES spectra along the series under study display a non-monotonic behaviour. This led to the development of several techniques. A simple technique is so-called progressive 2D correlation analysis, which involves repetition of the analysis including an increasing number of spectra each time. Such a technique has been applied in catalysis to elucidate the behaviour of Co-MCM-41 materials used for the production of carbon nanotubes during reduction and CO treatments up to 700°C. Combination with factorial analysis allows one to conclude that a Co^+ intermediate species is responsible for the initial stages of nanotube production. This species seems to be more reactive than those centres based on Co^{2+} (Haider et al., 2005; Ciuparu et al., 2005). A more elaborate technique is the so-called perturbation 2D correlation moving window. This technique involves the measurement of correlation spectra but selecting a series of $2m + 1$ ($m \ll n$) spectra around each spectrum, taking the window as central. The average perturbation $[\langle P(p_k) \rangle]$ is obtained and used in the scalar product in equation (1), which would now read $\langle Y^{\mathrm{T}}(\varepsilon_j) \rangle \cdot \langle P(p_k) \rangle$. This leads to synchronous and asynchronous $C(\varepsilon_j, p_k)$ 2D correlation spectra proportional to the spectral gradient ('perturbation' first derivative) and the negative rate of the spectral gradient change (perturbation second derivative), respectively. Such a technique has been used to correlate subtle changes in Ni (K-edge) pre-edge and white-line changes that take place in an Ni-MCM-41 material subjected to hydrogen reduction (do Nascimento et al., 2007).

3. Factor analysis

Factor analysis (FA) attempts to analyse multicomponent systems in two steps. The first step, which is common to all FA analytical techniques, follows the developments presented by Malinowski and others (Malinowski, 2002). This step is usually called principal component analysis and was applied to XANES for first time in 1995 (Fernández-García et al., 1995). The target of this step is to decompose the XANES X matrix constituted of c spectra into a row matrix R of 'basic' spectra (eigenvectors) and a column matrix C of weights (eigenvalues), so that

$$X_{r\cdot c} = R_{r\cdot n} \cdot C_{n\cdot c}, \qquad (2)$$

where the subscripts indicate the dimensions of the matrices, with r being the number of data points in a spectrum and n the number of factors or principal components (PCs). Equation (2) is solved by diagonalization of the covariance matrix of X and the factors are the number of 'free variables' that explain the variability of data set X. Several procedures define the n PCs required to reconstruct X within experimental error: (i) the decrease in eigenvalues, which ranks PCs according to their importance in reproducing the variance of X, (ii) a semiempirical indicator function IND, (iii) a specific level of significance for Malinowski's F-test of the variance associated with the kth eigenvalue and the summed variance associated

with 'noise' components (from $k + 1$ to c), and (iv) the 'normalized sum squared difference' estimator that measures the degree to which the set of n abstract eigenvectors represents a set of 'denoised' spectra (filtered signal expressed as an energy-related variable defined in such a way that spectral features below a typical width are eliminated) with respect to the real spectra (X) (Fernández-García, 2002; Manceau *et al.*, 2014). Studies usually consider all or at least some of these tests to define n and are applied to the normalized XANES spectra or the first derivative (Caballero *et al.*, 2005). More complex cases related to data sets with important noise or having a minor contribution from a factor can profit from the additional use of so-called evolving factor analysis (EFA). This procedure, applied to sets with inherent order in the data (for example materials where local environments of an element appear or disappear as the spectrum number increases/decreases), analyzes the evolution of the eigenvalues in successive runs upon increasing the number of spectra considered in the analysis procedure, and starting from the beginning and the end of the set. The crossing between the forward and backward (logarithm of the) eigenvalues as a function of the number of spectra included in the analysed X matrix defines the number of PCs as well as their regions of existence (Márquez-Álvarez *et al.*, 1997; Conti *et al.*, 2010).

Once the number of PCs is fixed, the XANES set of data is analysed by different procedures to obtain physical/chemical insights. There are two important points general to all of the methods. The first point is related to the fact that the XANES shape (*i.e.* spectroscopy) shows a strong sensitivity to size in the nanometre range, particularly below 15 nm (Fernández-García, 2002). This means that any analytical procedure that uses external references should be considered with caution, as such references are frequently obtained using bulk-type materials. Also, this is connected to the fact that nanomaterials present relatively wide particle-size distributions. The XANES spectrum is always an average spectrum and 'true' size information cannot be obtained (with the exception of exceptionally narrow or marked multimodal particle-size distributions), independently of the use of FA or any other analytical procedure. The second point relates to the fact that a final objective of FA could be to obtain factors with physical or chemical meaning, the so-called 'pure chemical species', which is not always possible, as detailed below.

Target testing is probably the simplest procedure and is conceptually close to linear fitting. This consists of testing for the presence of reference spectra in the R matrix using a test measuring how well the reference reproduces the data matrix. It establishes a range of values for acceptance, an uncertain situation or for rejection of the hypothesis (Malinowski, 2002). Such a procedure has been used to check for the presence of different copper oxidation states in copper-containing bimetallic PdCu materials subjected to reduction treatments (Fernández-García *et al.*, 1995), to analyse the local environment of sulfur in humic acids (Beauchemin *et al.*, 2002), to check the organic/inorganic nature of lead-containing materials in soils as well as their bioaccessibility (Smith *et al.*, 2011) and for the analysis of the interaction of uranium with iron-containing oxides (generated by the corrosion of iron containers), which is of interest for the safe disposal of nuclear fuels (Pidchenko *et al.*, 2017). Using target testing and a linear combination of references, the latter work showed that both the time of uranium–iron contact and the specific iron materials affect the oxidation state of uranium.

Another type of procedure uses the C matrix in equation (2) to measure the projection of each original spectrum over the orthogonal basis set of n eigenvectors (principal components). Such a method is utilized in spatially resolved studies using micro-beams at synchrotrons as well as microscopes. Analysis of the projection(s) module over the principal components can be carried out by classification techniques. k-means clustering is the most broadly used, allowing the spectrum associated with each spatial point of the XANES data set to be ascribed to a specific group, with the number of groups fixed arbitrarily as an initial guess. The average spectrum of each group is usually confronted with external references, but the method does not ensure that 'pure' chemical species will be obtained, as linear combinations of them are possible. Such a procedure has been utilized, for example, to analyse the oxidation state of arsenic in historic paintings or furniture (Keune *et al.*, 2015) and to carry out iron-containing chemical phase speciation in $LiFePO_4$ macro-crystals (Boesenberg *et al.*, 2013).

A more general procedure for obtaining the XANES spectra and concentration profiles of the 'pure' chemical species contributing to a multicomponent XANES data set is so-called iterative transformation factor analysis (ITFA). This exploits the fact that a physically meaningful C matrix should be non-negative to establish a iterative procedure consisting of rotating the R matrix until the differences in reproducing the X matrix are within experimental error (Malinowski, 2002). The mathematical procedure used to carry out ITFA is not unique. For successful application (*i.e.* to render 'pure' chemical species), a two-step procedure is commonly carried out. The first step is an initial rotation of the orthogonal basis to maximize the projection of the original spectra onto a new basis. Orthogonal (quartimax, equamax and particularly varimax) and oblique (promax and particularly quartimin and direct oblimin) rotations are the most popular initial rotation steps (Costello & Osborne, 2005). From this new basis set, use of the non-negativity of the C matrix allows the corresponding (physically meaningful) R matrix to be obtained. Whatever the mathematical details, examples of the fruitful use of ITFA can be found in many research works and fields, but a (short) selection could be headed by a review article considering the speciation of metal(loids) in environmental samples (Gräfe *et al.*, 2014). Cation-containing particles emitted from gasoline automobiles were also subjected to quantitative chemical speciation using ITFA–XANES (Ressler *et al.*, 2000). Catalysis is another field of frequent use of ITFA. The number of copper species exchanged in a ZSM-5 zeolite and their behaviour under CO and H_2 showed the stabilization of a copper(I) intermediate for the first gas (Neylon *et al.*, 2002). The evolution of palladium-based three-way catalysts was also studied in the nanometre range (2–5 nm) under CO–NO–O_2

atmospheres, showing a strong size dependence of the oxidation state of palladium and the capability to eliminate CO/NO (Iglesias-Juez *et al.*, 2011). EFA (in more or less elaborate versions) is also coupled to ITFA in many cases. For example, in electrochemistry the cell-charging step (lithium release) of Cu–V xerogels used in lithium batteries has been studied, indicating that copper goes from copper(II) to copper(0) through a copper(II) intermediate in which contact with vanadium has been demonstrated by this procedure and by EXAFS (Conti *et al.*, 2010). The application of EFA–ITFA analysis to the $LiVO_4F$–VPO_4F system in lithium batteries also showed the presence of three phases in charge/discharge cycles: the two mentioned above and a Li_xVO_4F phase with variable lithium content x (Piao *et al.*, 2014). Finally, another contribution using ITFA studied the behaviour of NiO electrodes in lithium batteries. Starting from reduced nickel, XANES was able to show that an intermediate is produced during the oxidation step (lithium incorporation) in the pathway ending in NiO. The intermediate appears as a metallic-type phase with oxygen at typical distances for chemisorption (Boesenberg *et al.*, 2014).

Funding information

References

Beauchemin, S., Hesterberg, D. & Beauchemin, M. (2002). *Soil Sci. Soc. Am. J.* **66**, 83–91.

Boesenberg, U., Marcus, M. A., Shukla, A. K., Yi, T., McDermott, E., Teh, P. F., Srinivasan, M., Moewes, A. & Cabana, J. (2014). *Sci. Rep.* **4**, 7133–7142.

Boesenberg, U., Meirer, F., Liu, Y., Shukla, A. K., Dell'Anna, R., Tyliszczak, T., Chen, G., Andrews, J. C., Richardson, T. J., Kostecki, R. & Cabana, J. (2013). *Chem. Mater.* **25**, 1664–1672.

Caballero, A., Morales, J. J., Cordón, A. M., Holgado, J. P., Espinos, J. P. & Gonzalezelipe, A. (2005). *J. Catal.* **235**, 295–301.

Ciuparu, D., Haider, P., Fernández-García, M. M., Chen, Y., Lim, S., Haller, G. L. & Pfefferle, L. (2005). *J. Phys. Chem. B*, **109**, 16332–16339.

Conti, P., Zamponi, S., Giorgetti, M., Berrettoni, M. & Smyrl, W. H. (2010). *Anal. Chem.* **82**, 3629–3635.

Costello, A. B. & Osborne, J. (2005). *Pract. Assess. Res. Eval.* **10**, 7.

do Nascimento, M. A., Paskocimas, C. A., Silva, A. J. N. & Ambrosio, R. C. (2007). *J. Phys. Chem. C*, **111**, 6813–6820.

Fernández-García, M. (2002). *Catal. Rev.* **44**, 59–121.

Fernández-García, M., Marquez Alvarez, C. & Haller, G. L. (1995). *J. Phys. Chem.* **99**, 12565–12569.

Gräfe, M., Donner, E., Collins, R. N. & Lombi, E. (2014). *Anal. Chim. Acta*, **822**, 1–22.

Haider, P., Chen, Y., Lim, S., Haller, G. L., Pfefferle, L. & Ciuparu, D. (2005). *J. Am. Chem. Soc.* **127**, 1906–1912.

Iglesias-Juez, A., Kubacka, A., Fernández-García, M., Di Michiel, M. & Newton, M. (2011). *J. Am. Chem. Soc.* **133**, 4484–4489.

Keune, K., Mass, J., Meirer, F., Pottasch, C., van Loon, A., Hull, A., Church, J., Pouyet, E., Cotte, M. & Mehta, A. (2015). *J. Anal. At. Spectrom.* **30**, 813–827.

Malinowski, E. R. (2002). *Factor Analysis in Chemistry*, 3rd ed. New York: John Wiley & Sons.

Manceau, A., Marcus, M. & Lenoir, T. (2014). *J. Synchrotron Rad.* **21**, 1140–1147.

Marcelli, A., Xu, W., Liu, L., Wang, C., Chu, W. & Wu, Z. (2009). *J. Nanophoton.* **3**, 031975.

Márquez-Álvarez, C., Rodríguez-Ramos, I., Guerrero-Ruiz, A., Haller, G. L. & Fernández-García, M. (1997). *J. Am. Chem. Soc.* **119**, 2905–2914.

Neylon, M. K., Marshall, C. L. & Kropf, A. J. (2002). *J. Am. Chem. Soc.* **124**, 5457–5465.

Noda, I. & Ozaki, Y. (2004). *Two-dimensional Correlation Spectroscopy – Applications in Vibrational and Optical Spectroscopy.* Chichester: John Wiley & Sons.

Piao, Y., Qin, Y., Ren, Y., Heald, S. M., Sun, C., Zhou, D., Polzin, B. J., Trask, S. E., Amine, K., Wei, Y., Chen, G., Bloom, I. & Chen, Z. (2014). *Phys. Chem. Chem. Phys.* **16**, 3254–3260.

Pidchenko, I., Kvashnina, K. O., Yokosawa, T., Finck, N., Bahl, S., Schild, D., Polly, R., Bohnert, E., Rossberg, A., Göttlicher, J., Dardenne, K., Rothe, J., Schäfer, T., Geckeis, H. & Vitova, T. (2017). *Environ. Sci. Technol.* **51**, 2217–2225.

Ressler, T., Wong, J., Roos, J. & Smith, I. (2000). *Environ. Sci. Technol.* **34**, 950–958.

Smith, E., Kempson, I. M., Juhasz, A. L., Weber, J., Rofe, A., Gancarz, D., Naidu, R., McLaren, R. G. & Gräfe, M. (2011). *Environ. Sci. Technol.* **45**, 6145–6152.

Van Bokhoven, J. A. & Lamberti, C. (2016). Editors. *X-ray Absorption and X-ray Emission Spectroscopy.* Chichester: John Wiley & Sons.