Chapter 6
Estimating a Set of Pure XANES Spectra from Multicomponent Chemical Mixtures Using a Transformation Matrix-Based Approach

Andrea Martini, Alexander A. Guda, Sergey A. Guda, Anastasiia Dulina, Francesco Tavani, Paola D'Angelo, Elisa Borfecchia, and Alexander V. Soldatov

Abstract In this work, we propose a new method for the analysis of time-resolved X-ray absorption near edge structure (XANES) spectra. It allows to decompose an experimental dataset as the product of two matrices: a *pure* spectral matrix, composed by XANES spectra associable to well-defined chemical species/sites, and their related concentration profiles. This method combines the principal component analysis and the application of a transformation matrix whose elements are directly accessible by the user. We demonstrate the potential of this approach applying it to a series of XANES spectra acquired during the direct conversion of methane to methanol (DMTM) over a Cu-exchanged zeolite characterized by the ferrierite topology. Possibilities and limitations of this methodology are discussed together with a critical comparison with the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) algorithm that, in the field of X-ray absorption spectroscopy (XAS), is imposing itself as a widely used method for spectral decomposition.

A. Martini (✉) · E. Borfecchia
Department of Chemistry, INSTM Reference Center and NIS and CrisDi Interdepartmental Centers, University of Torino, Via P. Giuria 7, 10125 Turin, Italy
e-mail: andrea.martini@unito.it

A. Martini · A. A. Guda · S. A. Guda · A. V. Soldatov
The Smart Materials Research Institute, Southern Federal University, Sladkova 178/24, 344090 Rostov-on-Don, Russia

S. A. Guda
Institute of Mathematics, Mechanics and Computer Science, Southern Federal University, Milchakova 8a, 344090 Rostov-on-Don, Russia

A. Dulina · F. Tavani · P. D'Angelo
Dipartimento di Chimica, Università di Roma "La Sapienza", P.le A. Moro 5, 00185 Rome, Italy

6.1 Introduction

One of the main dreams for researchers working in the field of chemistry and materials science consists in having a mathematical tool which allows to obtain an atomic-scale movie of a chemical reaction at realistic working conditions [1]. This request can be thought as the simultaneous determination of the spectra and concentrations of all the species involved in the analysed chemical reaction (i.e. reactants, intermediates and products), monitored by one or more characterization methods as a function of time. In this way, a reliable correlation between structure, kinetic and functionality can be properly identified. Focusing on the chemical speciation, X-ray absorption near edge structure (XANES) spectroscopy demonstrated to be an extremely useful technique, principally thanks to its local sensitivity and element selectivity, together with the possibility to simultaneously access both to the electronic and structural information of the material under study [2]. This fact led to the development of different strategies to decompose a dataset of XANES spectra acquired during a chemical/physical process, into a set of spectral and concentration profiles. However, most of them are based on the usage of particular constraints (i.e. the presence of a unique chemical specie at the beginning or at the end of the process) or references that, in some cases, are difficult or even impossible to measure, making their application unrealizable [3–5]. The work by Tauler et al. made a substantial contribution towards the solution of the spectral un-mixing problem. The authors proposed an automated data processing technique referred to as Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) which has been largely used during the last two decades in different fields of research, ranging from chromatography to image analysis [6, 7]. MCR-ALS is basically an iterative algorithm which allows the separation of the experimental data set into pure, chemically/physically meaningful, spectra and their associated concentrations without the use of any reference. In the last years, an increasing numbers of research groups have begun to use it in the analysis of large XAS datasets relevant to different scientific fields, such as battery research [8], quantum-dots formation [9], solid-state chemistry [10] and heterogeneous catalysis [11–14]. However, the possibility retrieve, from this method, a proper set of pure spectra and concentration profiles having a spectroscopic meaning seems to depend on the amount of the variance of the XANES dataset and on the initialization of the MCR-ALS routine [15]. There are, in fact, some XANES dataset, such as the ones reported by Guda and Bugaev [16, 17], showing only the variation of small spectral features causing, in this way, the failure of the MCR-ALS analysis. This fact lead to the development of a new approach (part of the PyFitIt software [18]) based on the joined application of Principal Component Analysis (PCA) and of a user-defined transformation matrix. In general, no particular standards are required to drive the output of this method towards a meaningful solution. Nonetheless, some background knowledge of the system under study (e.g. from complementary characterization techniques or computational analysis) appears to be greatly helpful for a robust interpretation of the results. In Sect. 6.3.1.2, this new method is applied to a dataset constituted of a series of Cu K-edge XANES spectra, collected on Cu-exchanged

ferririte zeolite (Cu-FER) during the direct conversion of CH_4 to CH_3OH. Finally, the obtained results are critically discussed and compared in Sect. 6.3.1.2.2 with the ones retrieved using the Multivariate Curve Resolution—Alternating Least Squares (MCR-ALS) method.

6.2 Method

6.2.1 The Transformation Matrix Approach

Let us consider an experimental XANES dataset μ_{ij} composed by M energy points and L spectra (i.e. $\dim(\mu_{ij}) = M \times L$), acquired during an experiment, where one or more physical or chemical variables are varying (e.g. time, temperature, pressure, pH …). Each spectrum μ_i of the dataset μ_{ij} can be expressed as a linear combination of N *pure* spectral components s_j (with $N < L$) as follow:

$$\mu_i = \sum_{j=1}^{N} c_{ij} s_j + \varepsilon_i \tag{6.1}$$

Equation (6.1) is the so-called Lambert and Beer equation [19]. Under this representation, μ_i and s_j are one-dimensional vectors with length equal to M, while the scalar term c_{ij} is the fraction of the jth component acquired during the ith scan (with $i = 0, 1, …, L$). Finally, the vector ε_i represents the experimental noise values associated to the ith vector in the dataset. It is worth noting that each of the N components must refer to a determined chemical species present in the analysed data mixture and must show some well-defined spectroscopic features able to visually characterize it (e.g. edge position, intensity/shape of the white line peak; number, energy position, and intensity and pre-edge and rising-edge peak …).

Considering Eq. (6.1), one would recover, starting from each experimental spectrum μ_i, the related pure spectra s_j and the associated concentration values c_{ij}. This request can be seen as an inverse problem. Herein, we present a mathematical method based on the usage of a transformation matrix able to find a solution of (6.1) realizing this kind of bilinear separation, entering in this way, in the family of the Multivariate Curve Resolution (MCR) methods [19, 20].

The first step of this approach foresees the application of the singular value decomposition (SVD) on the experimental dataset μ_{ij} as follow:

$$\mu_{ij} = u_{ik} \sigma_{kl} v_{lj} \tag{6.2}$$

where u_{ik} is the absorption coefficient for the component k, σ_{kl} is the element of a diagonal matrix, called singular values matrix, having the diagonal elements sorted in decreased order while the product $w_{kj} = \sigma_{kl} v_{lj}$ can be considered as the concentration

value associated to the kth specie. Different statistical and empirical criteria can be employed, on the basis of the analysis of σ_{kl}, to define how many components correspond to the real pure species with different absorption coefficient (i.e. N) and which of them are instead associated to the experimental noise (L–N). Among all of them, due to its effective interpretability, we employed in this work the analysis of the scree plot, as reported afterwards in Fig. 6.4a.

It is worth noting that the decomposition of μ_{ij} into the product of multiple spectral and concentration matrices is not unique. Equation (6.2) can be rewritten as:

$$\mu_{ij} = u_{ip} T_{pk} T_{kh}^{-1} w_{hj} \tag{6.3}$$

where T_{pk} is a square invertible matrix, called *transformation matrix*, having the property: $T_{pk} T_{kh}^{-1} = \delta_{ph}$. The inversion of T_{pk} can be used to realise decomposition (6.1) as: $s_{ik} = u_{ip} T_{pk}$ and $c_{kj} = T_{kh}^{-1} w_{hj}$. This step is fundamental. In fact, following the Eckhart-Young theorem, it is possible to state that the spectral and concentration profiles obtained directly form the SVD decomposition are able to guarantee the best approximation of μ_{ij} [21]. However, these values represent only a mathematical solution of (6.1) without any inherent chemical/physical meaning (see Fig. 6.4c). The transformation matrix allows, in this way, to convert the set of mathematical spectral and concentration profiles into a set of solutions of (6.1) having a physical/chemical interpretation. In the PCA section of the PyFitIt software [18], the elements of the transformation matrix are accessible by user and can be varied using sliders. Clearly, a proper set of constraints must be defined in order to reduce the number of elements of T_{pk} to be used (which goes as N^2) and their range of variation. Dealing with XANES spectra, it is possible to include the non-negativity of the spectral and concentration profiles and the mass balance condition, as stated by Conti et al. in their pioneering work regarding the application of the MCR-ALS approach (see Sect. 6.3.1.2.2) to the analysis of a set of XAS data [8]. While the first two constraints can be implemented looking for a set of parameters T_{pk} able to provide absorption coefficients and concentration values that are non-negative, the mass balance condition is less straightforward to realise. Indeed, it requires the normalization of the experimental spectral profiles. For our analysis, we used the following formula:

$$\ell_i = \sqrt{(1/(E_{max} - E_{min})) \int_{E_{min}}^{E_{max}} dE \mu_i(E)^2} \tag{6.4}$$

where ℓ_i is the normalization factor associated to the ith spectrum while E_{min} and E_{max} are respectively the minimum and maximum energy values of the XANES region. The requirement of the dataset normalization ensures the equality between the element of the first abstract concentration component of (6.3) (i.e. w_{h1}) and the normalization coefficient related to the first abstract spectrum: $w_{h1} = \ell_u$, where
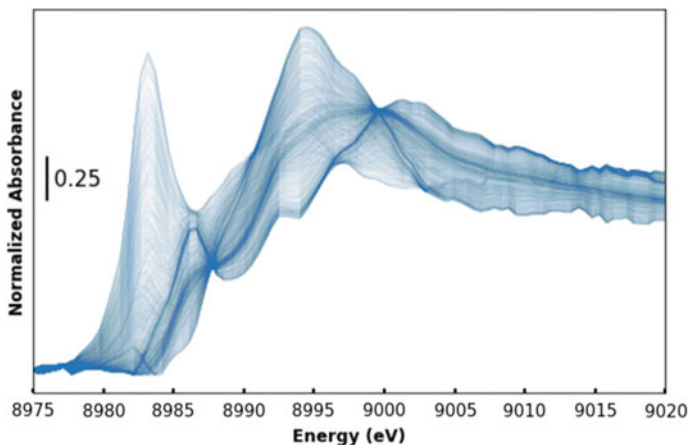
Fig. 6.1 Set of theoretical spectra, solution of Eq. (6.1), obtained from the XANES dataset described in Sect. 6.3.1.2. In order to identify them, the following constraints have been adopted: non-negativity of the spectral and concentration profiles together with the mass-balance condition

$\ell_u = \sqrt{(1/(E_{max} - E_{min})) \int_{E_{min}}^{E_{max}} dE u_{1p}(E)^2}$. This result can be used to guarantee the condition $\sum_{j=1}^{N} c_{ij} = 1$. In fact, it is possible to show that the normalization of the components reduces the number of matrix transformation elements from N^2 to $N^2 - N$ and determines the following simplification:

$$\sum_{j=1}^{N} c_{kj} = \sum_{j=1}^{N} T_{kh}^{-1} w_{hj} = w_{h1}/\ell_u = 1 \tag{6.5}$$

Similarly to the case of the Linear Combination Analysis (LCA) the uniform normalization of the experimental XANES spectra plays a fundamental role in the identification of spectroscopically interpretable results, in this case a set of pure spectral and concentration profiles. If the dataset is not properly normalised the condition reported in Eq. (6.5) cannot be satisfied leading to a set of concentration values whose sum for each scan can slightly differ from 1. At the same time it is possible to retrieve a series of pure spectra, characterized by a range of XANES points sited usually above the edge, which can deviate from the global profile of the XANES dataset, as described by Calvin in [22].

The presence of these constraints obviously limits the range of variation of the elements of T_{pk} and only the construction of a proper set of strongly selective constraints can lead to the isolation of a series of XANES components extremely close to the real physical/chemical solution. However, as showed, a unique solution of (6.1) cannot be identified. An ensemble of feasible XANES spectra is represented

in Fig. 6.1. Herein, this dataset has been generated considering the XANES data described in Sect. 6.3 and imposing the constraints described before.

The entire data analysis reported in this work has been realized using PYTHON 3.7. All the scripts can be provided by the corresponding author under request.

6.3 Case of Study

6.3.1 Spectral Decomposition for Cu K-Edge XANES of Cu-FER During the DMTM Conversion

6.3.1.1 Experimental Setup and Description of the Protocol Followed

XAS data were collected during the DMTM conversion at beamline BM31 [23] of the European Synchrotron Radiation Facility (ESRF, Grenoble, France). For the measurements, we used 3 mg of a Cu-FER sample with Cu/Al = 0.20 and Si/Al = 11. Details about the synthesis of this Cu-exchanged zeolite can be found in Ref. [24]. The sample was inserted in a 1 mm diameter quartz capillary with the powdered sample placed between glass wool plugs. The capillary was then fixed on a metal bracket and used as a fixed bed reactor. Finally, the gas inlet was connected to a dedicated gas flow setup. The process consisted of three steps: O_2 activation at 500 °C (120 min, 100% O_2), CH_4 loading at 200 °C (180 min, 100% CH_4) and H_2O assisted CH_3OH extraction at 200 °C (ca. 60 min). The temperature of the sample was controlled using a heat gun and the heating/ cooling ramps were performed with a 5 °C/min rate. The flow at each step was set to 2 ml/min using dedicated mass flow controllers (MFCs).

Cu K-edge XAS spectra were collected in transmission mode, using a water-cooled flat-Si (111) double crystal monochromator. The incident and transmitted X-ray intensities were detected using 30 cm long ionization chambers filled with He/Ar mixture. Scans in the range of 8800–9300 eV were continuously collected, binned with a constant energy step of 0.5 eV with the acquisition time being ca. 5 min/scan.

6.3.1.2 Data Analysis

In order to obtain more insights into the conversion mechanism of CH_4 to CH_3OH mediated by Cu-FER, we focused our analysis on the set of data acquired after the O_2 activation (see Fig. 6.2), starting from the He flushing till the extraction of CH_3OH by means of steam. The collected dataset shown in Fig. 6.3 is composed by 30 XANES spectra properly normalized to the unity edge jump using the Athena software from the Demeter package [25].
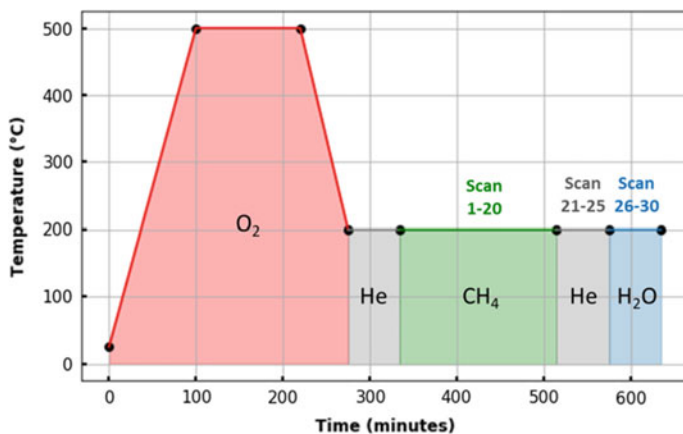
Fig. 6.2 Graphical representation of the protocol followed: 120 min O_2 activation at 500 °C (red), 180 min CH_4 loading at 200 °C (green), steam-assisted CH_3OH extraction at 200 °C for ca. 60 min (blue). The sample and the lines were flushed with He (grey segments) after O_2 activation and CH_4 loading for ca. 60 min
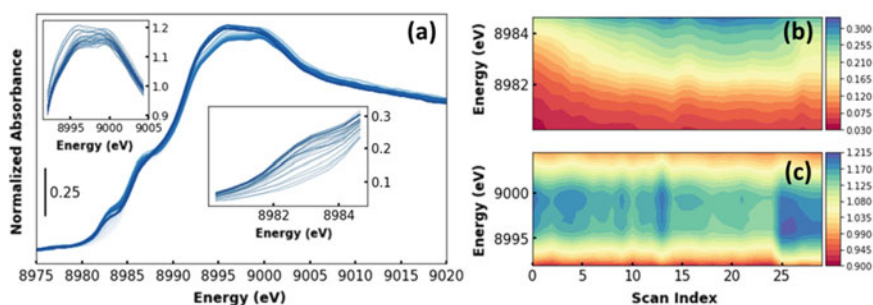

Fig. 6.3 **a** Plot of the analysed time-resolved XANES dataset: the insets contain the magnification of the spectral regions showing the highest variations during the followed experimental protocol: white line variations (upper left inset), rising-edge peak variation (central inset). **b** Contour maps associates to the insets reported in Figure (**a**)

As it is possible to see from Fig. 6.3a, during the entire MTM process, only small variations in the XANES spectra occur. In particular, these variations involve the intensities of the XANES white line and the rising-edge transitions (see the insets of Fig. 6.3a). Analysing these spectral modifications together with the variation of the scan index (that can be imagined as a temporal variable, being the adopted sampling time in our experiment 5 min/scan) some interesting trends appear. By sending CH_4, scans 1–20, the energy edge is shifted progressively towards lower values, the XANES white line magnitude becomes lower, while the intensity of the $1s \rightarrow 4p$ dipolar transition at ca. 8983 eV (characteristic of the Cu(I) ions) increases, as showed in Fig. 6.3b. This phenomenon can be interpreted as the reduction of a
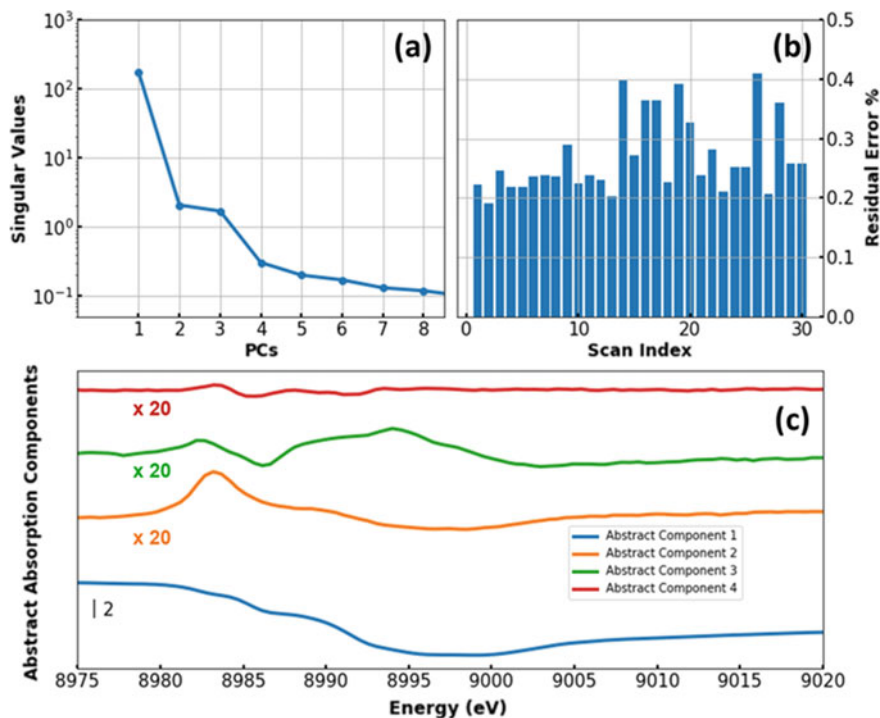
Fig. 6.4 PCA output. **a** Scree plot (logarithmic plot of the singular values extracted by the SVD vs. the number of PCs). **b** %R-factor (residual error) associated to the reconstruction of each spectrum of the experimental dataset shown in Fig. 6.3a using three components. **c** First four abstract components retrieved by PCA. All the abstract components, except for the first one, have been multiplied for a factor 20 in order to enhance their main spectral features

certain quantity of framework-coordinated Cu(II) sites, previously formed during the activation process in the presence of O_2, to Cu(I) sites, always coordinated to the zeolite lattice oxygens [2, 24, 26]. During the extraction of CH_3OH with water, scans 26–30, the edge energy is re-shifted towards higher energy, the intensity of the Cu(I) $1s \rightarrow 4p$ transition is abated and the XANES white line feature grows up again (see Fig. 6.3c). These evidences underline the presence of a higher abundance of Cu(II) sites in the chemical mixture, plausibly encompassing both Cu(II) aquo-complexes and framework-coordinated Cu(II) ions.

In order to identify the proper number of chemical species present in the analysed mixture, we applied the Principal Component Analysis (PCA) on the dataset showed in Fig. 6.3a. The results of this approach are reported in Fig. 6.4.

The analysis of the singular values, extracted by the SVD of the experimental dataset is reported in Fig. 6.4a. It is worth to note that each singular value is tight to the data variance explained by the related PC by the following relation: $s_i = \sigma_{ii}^2/(M-1)$, where the subscript i denotes the ith component [21]. It follows that those components, that are associated to the noise, contribute in the same way to the

dataset reconstruction and, for this reason, they are characterised by similar singular values. In the graph, an elbow is evident in proximity of the third component while from the fourth one onwards all the singular values lay approximately on a flat line. This trend suggests the presence of three PCs able to characterise the entire dataset. The fourth PC presents only some rather weak features if compared to the first three PCs, as evidenced in Fig. 6.4c and, for these reasons, it should be associated to some noise contribution or to the presence of a highly diluted specie. It is interesting to observe that the dataset reconstruction process with three PCs, shows an increase of the %R-factor values in proximity of two groups of scans: 14, 16, 17, 20 and 26, 28. The R-factor, for each scan, is defined as follows:

$$\%R_{Factor} = 100 \times \frac{\sum_{i=1}^{M} \left| \mu_{ij}^{PC} - \mu_{ij} \right|}{\sum_{i=1}^{M} \left| \mu_{ij} \right|} \tag{6.6}$$

where μ_{ij}^{PC} is the dataset reconstructed with three PCs. For the first group of scans, it is interesting to underline the correlation between the higher error values with the increasing of the spectral white line and the shift of the edge energy, as showed in Fig. 6.3b, c. On the other hand, the error associated with the second group of scans seems to be related to the appearance of CH_3OH during the steam-assisted extraction step. This analysis suggests that some transient chemical species are present for the mentioned scans, influencing the experimental spectra. Probably, these small variations in the dataset could be represented by the fourth and fifth component. However, based on the scree plot analysis results and on the error on the reconstruction using three PCs (lower than 0.45%) we decided to retrieve only three PCs.

Application of the Transformation Matrix Approach and Interpretation of the Results

We applied the transformation matrix approach on the experimental dataset showed in Fig. 6.3a. Each spectrum was initially normalised using Eq. (6.4). Then, employing the *target Transformation* function of PyFitIt [18] for three PCs, we defined a 3 × 3 transformation matrix. Thanks to the normalization constraint, we reduced the number of sliders to adjust from nine to six. The analysis of the raw data shows that the background profile due to the atomic Cu K-edge absorption process is similar for all the recorded spectra. As already pointed out by Giorgetti et al. in [27], this behaviour indicates that there are no secondary processes such as the loss or dissolution of a part of the sample during the entire reaction process or the movement of the powder inside the capillary. This fact justified the application, in this case, of the mass balance condition closure described in Sect. 6.2.1. Finally, the elements of the transformation matrix were moved according to the non-negativity of the spectra and concentration profiles.
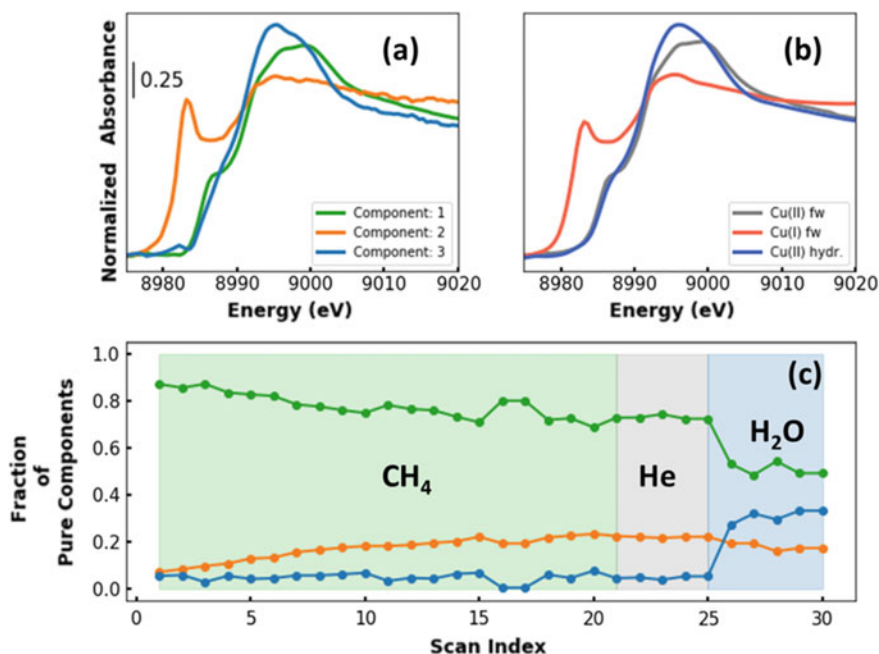
Fig. 6.5 **a**, **c** Spectral and concentration profiles retrieved using the transformation matrix approach. **b** Cu-references used to test visually the goodness of the spectral decomposition

A retrieved solution of Eq. (6.1) having a well-defined chemical/physical meaning is given by matrix $T_{pk} = \begin{pmatrix} 1/\ell & 1/\ell & 1/\ell \\ 3.40 & -1.05 & -0.70 \\ 0.45 & 1.50 & -0.30 \end{pmatrix}$, with $1/\ell = -0.18$ and it is showed in Fig. 6.5a, c.

It is possible to see that the identified spectral profiles are extremely similar to a set of references showed in Fig. 6.5b. These include a pseudo-octahedral Cu(II) aquo-complex (*Cu(II) hydr.*) as well as two framework-coordinated Cu(II) and Cu(I) species referred to as *Cu(II)* and *Cu(I) fw*, respectively. The *Cu(II) hydr.* was obtained measuring a Cu(II) acetate aqueous solution at RT. The *Cu(I) fw* reference was collected at RT after heating the sample up to 400 °C in vacuum. Finally the XANES acquired in He at 200 °C, just before the CH$_4$ loading step, was used as a *Cu(II) fw* reference.

The extracted profiles seem to be affected by a small amount of noise. This fact can be explained remembering that if the correct number of components is chosen, the PCA acts as a filter removing the highest amount of noise characterizing the dataset. However, as described by Malinowski [28], there is always a fraction of residual noise depending on the quality of the measurement mixed in the pure spectral and concentration profiles which cannot be removed deleting the unnecessary components.

The analysis of the concentration profiles associated to the pure spectra extracted showed in Fig. 6.5c and can lead to the following interpretation.

Scan 1 corresponds to the first state when the CH_4 is sent over the investigated Cu-FER sample at 200 °C. As it is possible to see from the concentration profiles (Fig. 6.5c), the amount of the second and third component is almost zero and it is possible to conclude that this scan is dominated by framework-coordinated Cu(II) sites (component n° 1, green spectrum in Fig. 6.5a). A precise assessment on the nature of this Cu(II) site is not straightforward. Depending on the zeolite topology, a number of Cu(II)-oxo species potentially active towards DMTM have been proposed to form during the high-temperature activation in O_2 and their structures are still debated in the literature [2, 24, 26, 29]. Among them, we can mention mono(μ-oxo) dicopper(II) cores, dicopper(II) peroxides and monocopper(II) superoxides. XANES simulations carried out on selected monomeric and dimeric Cu_xO_y moieties demonstrated that there is no sharp spectroscopic contrast in terms of spectral features among them [30, 31]. If follows that the first component profile is associated to a pure spectrum but it can be attributed to different Cu(II) species that, during the entire reaction, can coexist, making their identification impossible to be achieved using this technique.

During the sample interaction with CH_4, we observe the partial reduction of Cu(II) to Cu(I) (component n° 2, orange spectrum in Fig. 6.5a), see scans 1–25 in Fig. 6.5c. Focusing on the Cu(I) species, it is interesting to note that the maximum development of the related concentration profile occurs relatively early, around scan n° 7. Subsequently, concentration values tend to stabilize, indicating saturation of some Cu(II) reactive species. The Cu(I) spectrum, retrieved by the transformation matrix approach, can be associated to a two-fold coordinated Cu(I) specie. In fact, assuming the mono(μ-oxo) dicopper(II) as the active site for the CH_4 oxidation, the Cu(I) site supports the opening of the Cu-(μ-O)-Cu bridge in the mono(μ-oxo) dicopper cores upon (μ-O) methylation giving rise to the Z[Cu(I)(OCH$_3$)Cu(II)]Z intermediate (where Z denotes coordination to two zeolite framework oxygen atoms in the proximity of a charge-balancing framework Al site) [26]. Starting from this last structure, a proposed scenario involves the di-copper core dissociation into proximal Cu(I)/Cu(II) units, e.g. a bare ZCu(I) ion, having a spectral signature equal to component 2 of Fig. 6.5a and a methoxide Z[Cu(II)(OCH$_3$)] complex represented by a spectrum expected to be indistinguishable by *classic* XAS spectroscopy from the one associated to component 1. Novel insights about the identification of these intermediates could be obtained using High Energy Resolution Fluorescence Detected (HERFD) XANES, proven to be extremely helpful for the detection of the small variations of the XAS features that can characterize these species [15, 32].

Considering the scans associated with the CH_3OH extraction (26–30), it is interesting to see from Fig. 6.5c the presence of two processes triggered by water: the diminution of components n° 1 and n° 2, associated to framework-coordinated Cu(II) and Cu(I) species, and the appearance of a third component (blue spectrum and concentration profile) associated to a Cu(II) hydrated state. The framework-coordinated Cu(II) fraction diminution can be explained by the hydrolysis mechanism involving the methoxide group of the Z[Cu(II)(OCH$_3$)] complex while the small
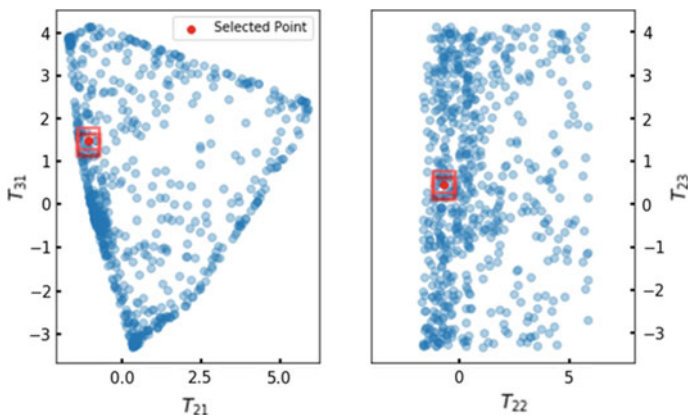
Fig. 6.6 Graphical representation of two of the fifteen AFS, related to dataset showed in Fig. 6.3, for the couples of variables: (T_{21}, T_{31}) and (T_{22}, T_{23}). These distributions have been obtained using a Monte Carlo approach, initializing and minimizing Eq. (6.7) for 1000 times. The initialization has been realized generating random numbers between -10 and 10, while for the minimization process, the Nelder-Mead algorithm has been employed [34]. With the red points are represented the sets of parameters able to provide the solution of Fig. 6.5, while the red cubes indicates, pictorially, the projections of a six-dimensional hypercube with a side of 0.3 over the 2D plane defined by these couples of parameters

abatement of the Cu(I) concentration values can be associated with H_2O-mediated re-oxidation pathways.

As previously discussed in Sect. 6.2.1, the solution obtained by the matrix transformation method depends on the values of the elements of T_{pk} and it is not unique. In order to quantify the maximum and minimum values of the spectral and concentration profiles for the solutions of (6.1) having a chemical/physical meaning, we proceeded with the following protocol:

First, we defined an objective function P as [33]:

$$P(T_{21}, T_{22}, T_{23}, T_{31}, T_{32}, T_{33}) = \sum_{i=1}^{L}\sum_{j=1}^{N} H_s(s_{ij})s_{ij}^2 + \sum_{k=1}^{M}\sum_{j=1}^{N} H_c(c_{kj})c_{kj}^2 \quad (6.7)$$

Due to the normalization constraint, P does not depend on the first row of T_{pk}, fixed to $1/\ell$. In (6.7) H_s is a Heaviside function that returns 0 if the spectral values s_{ij} are higher or equal to zero and 1 for their negative values, while H_c is a second function, associated with the concentrations profiles, that returns 0 for concentrations within 0 and 1 while it is equal to 1 if this last condition is not satisfied. Initializing randomly function P and minimizing it for a considerable number of iterations (i.e. 1000 or more) it is possible to obtain a graphical representation of all the combination of the elements of matrix T_{pk} satisfying the required constraints, called Area of Feasible Solutions (AFS), see Fig. 6.6. The ensemble of spectra associated to every minimum point of (6.7) is showed in Fig. 6.1.

The geometric shapes of the obtained AFS can be explained taking into account the portions of a $\mathbb{R}^6$ space enclosed in a subspace limited by the conditions $s_{ij} \geq 0$ and $0 \leq c_{ij} \leq 1$ [33]. Despite the large range of variation of the elements of the transformation matrix, only a small number of combinations of these parameters are acceptable. The retrieved spectra must satisfy the imposed constraints as showed by Figs. 6.1 and 6.6, but, at the same time, they must be characterized by determined spectral features physically and chemically interpretable. This fact reduces drastically the number of spectra of Fig. 6.1 and consequently the related AFS showed in Fig. 6.6. Unfortunately, at the moment, there is no technique available able to automatedly assess if a XANES spectrum, generated by a determined combination of parameters T_{pk}, has a physical/chemical meaning. The transformation matrix approach is not able to realize the so-called *blind source separation* of the experimental signal and only the user's intuition and the knowledge of the system under study can lead to a meaningful solution. It is opinion of the authors that the creation of a large dataset of reference XANES (experimental and simulated) spectra together with a solid Machine Learning algorithm for spectral comparison could improve the quality of the results. However, it is possible to select a region surrounding a feasible point and try to identify the maximum and minimum band boundaries of the feasible solutions having a physical/chemical meaning. To do this, we exploited the idea of Tauler [35] and we defined the following scalar function:

$$f_n\left(T_{ij}\right) = \frac{s_{in}\left(T_{ij}\right)c_{nj}\left(T_{ij}\right)}{\mu_{ij}} \tag{6.8}$$

where the operator $\|\cdot\|$ indicates the Frobenius norm. This function gives the ratio between the contribution of a particular nth specie with respect to the total contribution coming from all the components μ_{ij}. The optimization of this objective function, either maximized or minimized under the constraints, will give respectively the maximum and the minimum boundary for each chemical specie present in the dataset. In our case, we considered a subspace of AFS consisting of a six-dimensional hypercube having a side equal to 0.3 (six times the step variation used as a standard values in PyFitIt [18]) surrounding the point which provides the spectra and concentrations of Fig. 6.5. Afterwards, we minimised and maximised Eq. (6.8) changing progressively the components. This step was realised under constraints (described before) using the Sequential Least Squares Programming method [36].

The obtained results are showed in Fig. 6.7.

Analysing this picture, it is interesting to see that the lines constituting the spectral variation bounds are extremely close to each other. Some small differences appear in the rising-edge region (especially for the 1s $\rightarrow$ 4p peak of the Cu(I) component) and for the white line peak. Vice-versa, larger variations are observable for the related concentration profiles. The explanation must be found in the selection of the subspace of the T_{pk} parameters used for the minimization procedure [37]. The chosen hypercube has been defined in order to incorporate only the spectral profiles characterized by interpretable spectroscopic features. This '*user-based*' constraint limited
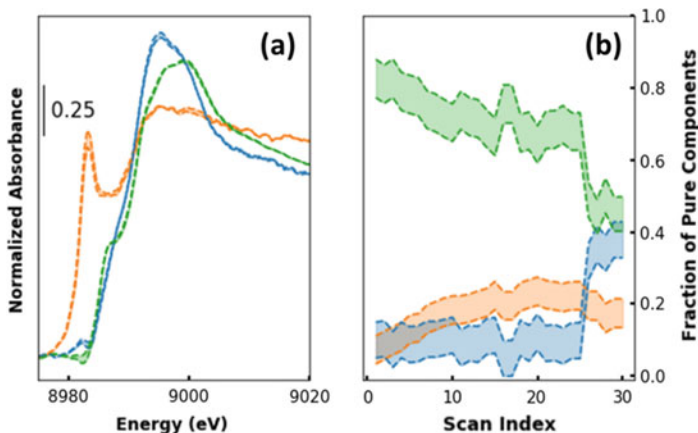
Fig. 6.7 Spectral (**a**) and concentration (**b**) band boundaries calculated for the profiles of Fig. 6.5 minimizing and maximizing for six times Eq. (6.8)

the shape of the pure spectral profiles that can be isolated but not their concentrations that, in the selected range of variation of the T_{ij} can undergo significant variations. Possible strategies to reduce the concentration band boundaries amplitude could rely on the introduction of additional concentration constraints or by fixing a reference spectrum as a *pure* component in the analysed system.

Application of the MCR-Alternate Regression (MCR-AR) Method on the Analysed Dataset

For the sake of comparison, we performed the decomposition of the experimental dataset of Fig. 6.3 according to Eq. (6.1) using a different MCR method based on an alternate regression algorithm [38]. This technique is becoming extremely popular in the field of the XAS analysis, especially for time or space-resolved measurements when a large series of spectra must be analysed or when a high number of components (i.e. >3) characterize the experimental dataset. The MCR algorithm requires an initial set of spectral s_{ih}^0 or concentrations profiles c_{hj}^0. If, as an example, the algorithm is initialized using s_{ih}^0, then the concentration profiles related to step $k = 1$ will be given by the following minimization:

$$c_{hj}^1 = \underset{c_{hj}^0}{\operatorname{argmin}}[\mathcal{F}_C(s_{ih}^0 c_{hj}^0)] \qquad (6.9)$$

where $\mathcal{F}_C$ is an objective function. Once the concentration profiles have been defined, a new set of spectral values can be retrieved minimizing a second objective function $\mathcal{F}_S$:
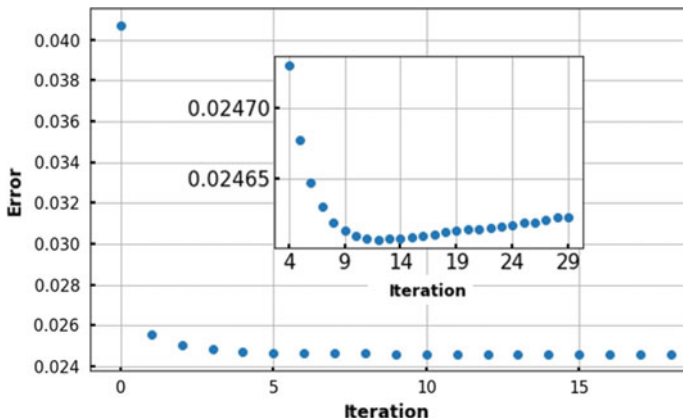
Fig. 6.8 Plot of the error in the dataset reconstruction, calculated using Eq. (6.11), versus the number of iterations employed. The MCR-ALS routine has been initialized using spectra showed in Fig. 6.5b. The insets report the magnification of minimum region located at iteration n° 12

$$s_{ih}^1 = \underset{s_{ih}^0}{\mathrm{argmin}}[\mathcal{F}_S(s_{ih}^0 c_{hj}^1)] \tag{6.10}$$

Both the minimization processes (6.9) and (6.10) must be performed under constraints. Among all the different regressors available in Python, we found particularly suitable for the XANES decomposition the OLS (ordinary least squares) regressor, which minimizes the L_2-norm (residual sum of squares) among the original dataset μ_{ij} and the reconstructed-one. In the literature, the MCR method based on the multiple OLS regression is usually named as MCR-ALS (where ALS stands for alternating least squares) [6]. Herein, the *classical* XANES constraints can be imposed (i.e. spectral and concentration non-negativity and mass balance condition) allowing one to drive the set of minimizations towards a feasible solution. The scheme of multiple regression described above can be easily extended to *k*-iterations. For each step, as a function of the retrieved s_{ih}^k and c_{hj}^k, an expression describing the goodness of the reconstruction can be calculated. In our analysis, we adopted $\mathcal{E}_k$ described by the following equation [39]:

$$\mathcal{E}_k = 100 \times \langle \sqrt{\frac{\langle \left(\mu_{ij} - s_{ih}^k c_{hj}^k\right)^2 \rangle_i}{\langle \mu_{ij}^2 \rangle_i}} \rangle \tag{6.11}$$

where the operator $\langle \cdot \rangle_i$ denotes the mean over the columns' matrix while $\langle \cdot \rangle$ represents the mean calculated on a one-dimensional vector. Usually, if the difference between the errors associated to two consecutive iterations is lower than 0.1% the routine is stopped. In the case of the Cu-FER dataset in Fig. 6.3, the error trend related to the MCR-ALS method versus the iteration number is reported in Fig. 6.8. It is
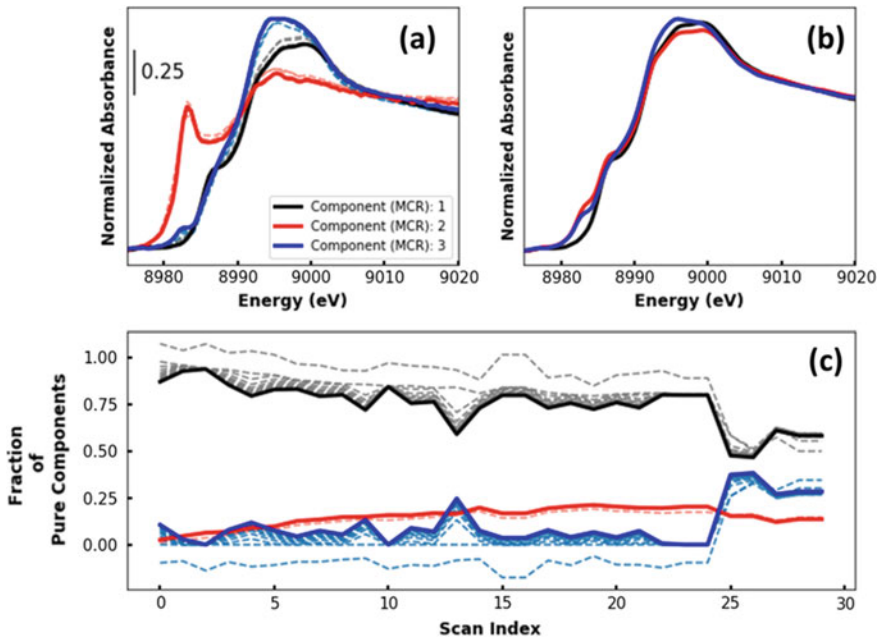
Fig. 6.9 **a**, **c** Spectral and concentration profiles obtained by MCR-ALS algorithm initialized with the spectral references showed in Fig. 6.5b. The dashed lines represent the intermediate spectra and concentration before reaching the minimum values of the error in the dataset reconstruction process. **b** Set of spectral profiles obtained by the same method initialized using SIMPLISMA algorithm

interesting to see that after three iterations the difference $\Delta\mathcal{E}_{23} = (\mathcal{E}_2 - \mathcal{E}_3) < 0.1\%$; after the third iteration only small variations occur, indicating that this set of spectra is already a good candidate to represents properly the dataset. However, for the sake of completeness, we assumed as the final state of the refinement process the one associated to the minimum value of the error function $\mathcal{E}_k$, that corresponds to the 12th iteration.

The power of this method stands principally in its *blindness* regarding the system under study. However, the entire routine is extremely sensitive to the kind of initialization used. Different statistical techniques such as EFA and SIMPLISMA can be applied to generate or isolate a proper set of spectra or concentration profiles suitable for the subsequent minimization routine [40, 41]. Nevertheless, these methods strongly depend on the amount of variation of spectra in the dataset [15]. If these variations are low, as for the dataset under study, MCR-ALS algorithm often fails, proposing a minimum characterized by spectra and concentrations, which minimize the error associated to the reconstruction but are still a mixture of pure components; see Fig. 6.9b.

The solutions to this problem are multiple but involve further measurements or a deeper knowledge of the system under study. Different datasets supposed to be characterized by the same components can be merged together in order to increase

the variance associated to the data, helping, in this way to identify a proper initial set of spectral and concentration profiles. An example where this strategy provided good results can be found in [42], where multiple XANES datasets collected on Cu-zeolites (chabazite) samples with different Si/Al and Cu/Al ratios, during the same activation process (from 25 to 400 °C) were joined in one larger dataset. Another strategy could be fixing some components to determined references (supposed to be present in the data mixture) or the initialization of the ALS routing using always selected references or some spectral profiles supposed to be connected with almost *pure* species. This last method, employing the reference spectra in Fig. 6.5b, was the one that we used to retrieve the set of spectral and concentration profiles, reported in Fig. 6.9a, c. Herein, the isolated components have a well-defined chemical-physical meaning and differ from the spectra used for the initialization only for small variations in the pre-edge and on the white-line. Finally, it is also interesting to note that the identified MCR-ALS concentration profiles lye in the band boundaries region showed in Fig. 6.7b, confirming the comparability of this method with the transformation matrix approach.

6.4 Conclusions

In this work, we firstly demonstrated that the transformation matrix approach is an efficient technique for the analysis of a generic experimental XANES dataset, even when characterized by small spectral variations, as it is the case for the Cu K-edge XANES dataset described in Sect. 6.3, collected during DMTM over Cu-FER. After-wards, we compared the results obtained through the application of this method with the ones derived by the MCR-ALS approach. We showed that both techniques are able to isolate similar pure XANES spectra. However, we stressed the fact that the set of spectral and concentration profiles provided by the MCR-ALS approach seem to depend strongly on the degree of the variation characterizing the experimental dataset and on the methods adopted for the initialization of the routine. On the other hand, despite the inability to identify a unique solution, the application of constraints can drastically reduce the number of solutions provided by the transformation matrix approach, leading to a set of chemically/physically interpretable spectra and concentration profiles. At the same time, the multiple minimization and maximization of Eq. (6.8) provides a valid method to define the variation bounds associated to the pairs of spectral and concentration profiles identified by this new technique.

References

1. G. Smolentsev, G. Guilera, M. Tromp, S. Pascarelli, A.V. Soldatov, Local structure of reaction intermediates probed by time-resolved X-ray absorption near edge structure spectroscopy. J. Chem. Phys. **130**, 9 (2009)
2. D.K. Pappas, E. Borfecchia, M. Dyballa, I.A. Pankin, K.A. Lomachenko, A. Martini, M. Signorile, S. Teketel, B. Arstad, G. Berlier, C. Lamberti, S. Bordiga, U. Olsbye, K.P. Lillerud, S. Svelle, P. Beato, Methane to methanol: structure activity relationships for Cu-CHA. J. Am. Chem. Soc. **139**, 14961–14975 (2017)
3. T. Ressler, WinXAS: a program for X-ray absorption spectroscopy data analysis under MS-Windows. J. Synchrotron Radiat. **5**, 118–122 (1998)
4. Q. Wang, J.C. Hanson, A.I. Frenkel, Solving the structure of reaction intermediates by time-resolved synchrotron X-ray absorption spectroscopy. J. Chem. Phys. **129**, 7 (2008)
5. S.M. Webb, SIXpack: a graphical user interface for XAS analysis using IFEFFIT. Phys. Scr. **T115**:1011–1014 (2005)
6. J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. Chemometr. Intell. Lab. Syst. **76**, 101–110 (2005)
7. R. Tauler, Multivariate curve resolution applied to second order data. Chemometr. Intell. Lab. Syst. **30**, 133–146 (1995)
8. P. Conti, S. Zamponi, M. Giorgetti, M. Berrettoni, W.H. Smyrl, Multivariate curve resolution analysis for interpretation of dynamic Cu K-edge X-ray absorption spectroscopy spectra for a Cu doped V_2O_5 lithium battery. Anal. Chem. **82**, 3629–3635 (2010)
9. B.L. Caetano, V. Briois, S.H. Pulcinelli, F. Meneau, C.V. Santilli, Revisiting the ZnO Q-dot formation toward an integrated growth model: from coupled time resolved UV–Vis/SAXS/XAS data to multivariate analysis. J. Phys. Chem. C **121**, 886–895 (2017)
10. H.W.P. Carvalho, S.H. Pulcinelli, C.V. Santilli, F. Leroux, F. Meneau, V. Briois, XAS/WAXS time-resolved phase speciation of chlorine LDH thermal transformation: emerging roles of isovalent metal substitution. Chem. Mat. **25**, 2855–2867 (2013)
11. W.H. Cassinelli, L. Martins, A.R. Passos, S.H. Pulcinelli, C.V. Santilli, A. Rochet, V. Briois, Multivariate curve resolution analysis applied to time-resolved synchrotron X-ray Absorption Spectroscopy monitoring of the activation of copper alumina catalyst. Catal. Today **229**, 114–122 (2014)
12. J.P. Hong, E. Marceau, A.Y. Khodakov, L. Gaberova, A. Griboval-Constant, J.S. Girardon, C. La Fontaine, V. Briois, Speciation of ruthenium as a reduction promoter of silica-supported Co catalysts: a time-resolved in situ XAS investigation. ACS Catal. **5**, 1273–1282 (2015)
13. A. Rochet, B. Baubet, V. Moizan, E. Devers, A. Hugon, C. Pichon, E. Payen, V. Briois, Intermediate species revealed during sulfidation of bimetallic hydrotreating catalyst: a multivariate analysis of combined time-resolved spectroscopies. J. Phys. Chem. C **121**, 18544–18556 (2017)
14. A. Voronov, A. Urakawa, W.V. Beek, N.E. Tsakoumis, H. Emerich, M. Rønning, Multivariate curve resolution applied to in situ X-ray absorption spectroscopy data: an efficient tool for data processing and analysis. Anal. Chim. Acta **840**:20–27 (2014)
15. A. Martini, E. Alladio, E. Borfecchia, Determining Cu-speciation in the Cu-CHA zeolite catalyst: the potential of multivariate curve resolution analysis of in situ XAS data. Top. Catal. **61**, 1396–1407 (2018)
16. A.L. Bugaev, O.A. Usoltsev, A.A. Guda, K.A. Lomachenko, I.A. Pankin, Y.V. Rusalev, H. Emerich, E. Groppo, R. Pellegrini, A.V. Soldatov, J.A. van Bokhoven, C. Lamberti, Palladium carbide and hydride formation in the bulk and at the surface of palladium nanoparticles. J. Phys. Chem. C **122**, 12029–12037 (2018)
17. A.A. Guda, A.L. Bugaev, R. Kopelent, L. Braglia, A.V. Soldatov, M. Nachtegaal, O.V. Safonova, G. Smolentsev, Fluorescence-detected XAS with sub-second time resolution reveals new details about the redox activity of Pt/CeO_2 catalyst. J. Synchrot. Radiat. **25**, 989–997 (2018)

18. A. Martini, S.A. Guda, A.A. Guda, G. Smolentsev, A. Algasov, O. Usoltsev, M.A. Soldatov, A. Bugaev, Y. Rusalev, C. Lamberti, A.V. Soldatov, PyFitit: the software for quantitative analysis of XANES spectra using machine-learning algorithms. Comput. Phys. Commun. 107064 (2019)
19. C. Ruckebusch, *Resolving Spectral Mixtures: With Applications from Ultrafast Time-Resolved Spectroscopy to Super-Resolution Imaging* (Elsevier, Amsterdam, 2016)
20. J. Timoshenko, A.I. Frenkel, "Inverting" X-ray absorption spectra of catalysts by machine learning in search for activity descriptors. ACS Catal. **9**, 10192–10211 (2019)
21. I. Markovsky, Structured low-rank approximation and its applications. Automatica **44**, 891–909 (2008)
22. S. Calvin, *XAFS for Everyone* (CRC Press, Boca Raton, 2013)
23. P.M. Abdala, O.V. Safonova, G. Wiker, W. van Beek, H. Emerich, J.A. van Bokhoven, J. Sa, J. Szlachetko, M. Nachtegaal, Scientific opportunities for heterogeneous catalysis research at the SuperXAS and SNBL beam lines. Chimia **66**, 699–705 (2012)
24. D.K. Pappas, E. Borfecchia, K.A. Lomachenko, A. Lazzarini, E.S. Gutterod, M. Dyballa, A. Martini, G. Berlier, S. Bordiga, C. Lamberti, B. Arstad, U. Olsbye, P. Beato, S. Svelle, Cu-exchanged ferrierite zeolite for the direct CH_4 to CH_3OH conversion: insights on Cu speciation from X-ray absorption spectroscopy. Top. Catal. **62**, 712–723 (2019)
25. B. Ravel, M. Newville, ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT. J. Synchrotron Radiat. **12**, 537–541 (2005)
26. K.A. Lomachenko, A. Martini, D.K. Pappas, C. Negri, M. Dyballa, G. Berlier, S. Bordiga, C. Lamberti, U. Olsbye, S. Svelle, P. Beato, E. Borfecchia, The impact of reaction conditions and material composition on the stepwise methane to methanol conversion over Cu-MOR: an operando XAS study. Catal. Today **336**, 99–108 (2019)
27. M. Giorgetti, S. Mukerjee, S. Passerini, J. McBreen, W.H. Smyrl, Evidence for reversible formation of metallic Cu in Cu0.1V2O5 xerogel cathodes during intercalation cycling of Li⁺ ions as detected by X-ray absorption spectroscopy. J. Electrochem. Soc. **148**, A768–A774 (2001)
28. E.R. Malinowski, *Factor Analysis in Chemistry* (Wiley, Hoboken, 2002)
29. E.M.C. Alayon, M. Nachtegaal, A. Bodi, J.A. van Bokhoven, Reaction conditions of methane-to-methanol conversion affect the structure of active copper sites. ACS Catal. **4**, 16–22 (2014)
30. A. Martini, I.A. Pankin, A. Marsicano, K.A. Lomachenko, E. Borfecchia, Wavelet analysis of a Cu-oxo zeolite EXAFS simulated spectrum. Radiat. Phys. Chem. 108333 (2019)
31. I.A. Pankin, A. Martini, K.A. Lomachenko, A.V. Soldatov, S. Bordiga, E. Borfecchia, Identifying Cu-oxo species in Cu-zeolites by XAS: a theoretical survey by DFT-assisted XANES simulation and EXAFS wavelet transform. Catal. Today (2019)
32. D.K. Pappas, A. Martini, M. Dyballa, K. Kvande, S. Teketel, K.A. Lomachenko, R. Baran, P. Glatzel, B. Arstad, G. Berlier, C. Lamberti, S. Bordiga, U. Olsbye, S. Svelle, P. Beato, E. Borfecchia, The nuclearity of the active site for methane to methanol conversion in Cu-mordenite: a quantitative assessment. J. Am. Chem. Soc. **140**, 15270–15278 (2018)
33. K. Sasaki, S. Kawata, S. Minami, Constrained nonlinear method for estimating component spectra from multicomponent mixtures. Appl. Opt. **22**, 3599–3603 (1983)
34. J.A. Nelder, R. Mead, A simplex method for function minimization. Comput. J. **7**, 308–313 (1965)
35. R. Tauler, Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. J. Chemometr. **15**, 627–646 (2001)
36. D. Kraft, *A Software Package for Sequential Quadratic Programming* (DFVLR, Köln, 1988)
37. A.C. Olivieri, R. Tauler, The effect of data matrix augmentation and constraints in extended multivariate curve resolution-alternating least squares. J. Chemometr. **31**, 10 (2017)
38. C.H. Camp, pyMCR: a python library for multivariate curve resolution analysis with alternating regression (MCR-AR). J. Res. Natl. Inst. Stand. Technol. **124**, 10 (2019)
39. A.A. Guda, S.A. Guda, K.A. Lomachenko, M.A. Soldatov, I.A. Pankin, A.V. Soldatov, L. Braglia, A.L. Bugaev, A. Martini, M. Signorile, E. Groppo, A. Piovano, E. Borfecchia, C. Lamberti, Quantitative structural determination of active sites from in situ and operando XANES spectra: from standard ab initio simulations to chemometric and machine learning approaches. Catal. Today **336**, 3–21 (2019)

40. M. Maeder, Evolving factor-analysis for the resolution of overlapping chromatographic peaks. Anal. Chem. **59**, 527–530 (1987)
41. W. Windig, J. Guilment, Interactive self-modeling mixture analysis. Anal. Chem. **63**, 1425–1432 (1991)
42. A. Martini, E. Borfecchia, K.A. Lomachenko, I.A. Pankin, C. Negri, G. Berlier, P. Beato, H. Falsig, S. Bordiga, C. Lamberti, Composition-driven Cu-speciation and reducibility in Cu-CHA zeolite catalysts: a multivariate XAS/FTIR approach to complexity. Chem. Sci. **8**, 6836–6851 (2017)