Computer Physics Communications

PyFitit: the software for quantitative analysis of XANES spectra using machine-learning algorithms

--Manuscript Draft--

Manuscript Number:	COMPHY-D-19-00267R3
Article Type:	Computer programs in physics paper
Keywords:	X-Ray Absorption Spectroscopy (PACS code: 61.10.Ht); Machine Learning; PCA; XANES Fitting procedure; Spectral decomposition
Corresponding Author:	Andrea Martini, M.D. University of Turin Turin, Turin ITALY
First Author:	Andrea Martini, M.D.
Order of Authors:	Andrea Martini, M.D.
	Sergey A. Guda, Professor
	Alexander A. Guda, Dr.
	Grigory Smolentsev, Dr.
	Alexander Algasov
	Oleg Usoltsev, M.D.
	M. A. Soldatov, Dr.
	Aram Bugaev, Dr.
	Yuri Rusalev
	Alexander V. Soldatov, Professor
Abstract:	X-ray absorption near-edge spectroscopy (XANES) is becoming an extremely popular tool for material science thanks to the development of new synchrotron radiation light sources. It provides information about charge state and local geometry around atoms of interest in operando and extreme conditions. However, in contrast to X-ray diffraction, a quantitative analysis of XANES spectra is rarely performed in the research papers. The reason must be found in the larger amount of time required for calculation of a single spectrum compared to a diffractogram. For such time-consuming calculations, in the space of several structural parameters, we developed an interpolation approach proposed originally by Smolentsev et al. [1]. The current version of this software, named PyFitIt, is a major upgrade version of FitIt and it is based on machine learning algorithms. We have chosen Jupyter Notebook framework to be friendly for users and at the same time being available for remastering. The analytical work is divided in two steps. First, the series of experimental spectra are analysed statistically and decomposed into principal components. Second, pure spectral profiles, recovered by principal components, are fitted by theoretical interpolated spectra. We implemented different schemes of choice of nodes for approximation and learning algorithms including Gradient Boosting of Random Trees, Radial Basis Functions and Neural Networks. The fitting procedure can be performed both for a XANES spectrum or for a difference spectrum, thus minimizing the systematic errors of theoretical simulations. The problem of several local minima is addressed in the framework of direct and indirect approaches.

PyFitit: the software for quantitative analysis of XANES spectra using machine-learning algorithms

A. Martini[1,2]*, S. A. Guda[2,3]*, A. A. Guda[2], G. Smolentsev[4], A. Algasov[3], O. Usoltsev[2], M. A. Soldatov[2], A. Bugaev[2], Yu. Rusalev[2], C. Lamberti[1,2], A. V. Soldatov[2].

[1]Department of Physics, INSTM Reference Center and NIS and CrisDi Interdepartmental Centers, University of Torino, Via P. Giuria 1, I-10125 Torino, Italy
[2]The Smart Materials Research Institute, Southern Federal University, 344090 Sladkova 178/24 Rostov-on-Don, Russia
[3]Institute of mathematics, mechanics and computer science, Southern Federal University, 344090 Milchakova 8a, Rostov-on-Don, Russia
[4]Paul Scherrer Insitute, Villigen 5232, Switzerland

Corresponding authors: andrea.martini@unito.it, gudasergey@gmail.com

Abstract

X-ray absorption near-edge spectroscopy (XANES) is becoming an extremely popular tool for material science thanks to the development of new synchrotron radiation light sources. It provides information about charge state and local geometry around atoms of interest in *operando* and extreme conditions. However, in contrast to X-ray diffraction, a quantitative analysis of XANES spectra is rarely performed in the research papers. The reason must be found in the larger amount of time required for the calculation of a single spectrum compared to a diffractogram. For such time-consuming calculations, in the space of several structural parameters, we developed an interpolation approach proposed originally by Smolentsev et al. [1]. The current version of this software, named PyFitIt, is a major upgrade version of FitIt and it is based on machine learning algorithms. We have chosen Jupyter Notebook framework to be friendly for users and at the same time being available for remastering. The analytical work is divided into two steps. First, the series of experimental spectra are analyzed statistically and decomposed into principal components. Second, pure spectral profiles, recovered by principal components, are fitted by theoretical interpolated spectra. We implemented different schemes of choice of nodes for approximation and learning algorithms including Gradient Boosting of Random Trees, Radial Basis Functions and Neural Networks. The fitting procedure can be performed both for a XANES spectrum or for a difference spectrum, thus minimizing the systematic errors of theoretical simulations. The problem of several local minima is addressed in the framework of direct and indirect approaches.

PROGRAM SUMMARY

Program title: PyFitIt.

Licensing provisions: GNU General Public License 3 (GPL).

Programming language: Python, Jupyther Notebook framework.

Journal Reference of the previous version: J. Synch. Radiat., 13 (2006) 19-29 ; Comput. Mater. Sci., 39 (2007) 569-574.

Does the new version supersede the previous version?: yes

Reasons for the new version: we have rewritten the code in the Jupyter Notebooks to make it available for the use and modification by members of the XAS scientific community. When the number of structural parameters for fitting exceeds 3 the use of polynomial interpolation realized in the previous version was highly non-optimal and inaccurate. Therefore, in the new version, the approximation methods were revised in terms of the use of modern machine learning strategies. Finally, in the new version, an important step of analysis of a series of experimental spectra was added, named Principal Component Analysis and spectral un-mixing procedure.

Summary of revisions: Development of a library of methods able to: i) analyze the experimental set of data (PCA); ii) construct the molecule deformations for a selected set of points in a multidimensional space (grid, random and IHS options are available); iii) run the simulations locally or remotely; iv) training the machine learning algorithms (ridge regression, Radial Basis Functions, Extra Trees, Neural Network, LightGBM … etc) on the set of theoretical spectra and fitting the experimental spectra or their differences using inverse or direct approaches.

Nature of problem: Quantitative structural refinements of the X-ray absorption near-edge structure spectra (XANES). Identification of the pure spectral and concentration profiles associated with an experimental XANES dataset.

Solution method: The fitting procedure of the experimental XANES spectra or of their differences is realized by means of the inverse and direct approaches based on the training set and approximation machine learning algorithms. The spectral resolution method is based on the PCA technique involving the usage of a target transformation matrix.

Additional comments including Restrictions and Unusual features: The current version is compatible with the free FDMNES program package for XANES simulations. However, users can prepare their own matrices of spectra calculated by an arbitrary software and the corresponding structural parameters to perform the fitting procedure in PyFitIt. The complete set of examples is distributed along with the program.

References: PyFitIt web page: http://hpc.nano.sfedu.ru/pyfitit/

Keywords: XANES; Fit; Machine Learning; Direct and Inverse approaches; PCA; Spectral Decomposition.

1. Introduction

The low energy part of the X-ray absorption spectrum, extending from some eV below the rising edge up to several tens of eV above it, is known as X-ray absorption near edge structure (XANES) [2, 3]. It is associated with the excitation process of a core electron to unoccupied states near the ionization threshold. The analysis of a XANES spectrum provides information about the local environment of the absorber, such as the bond lengths (the famous Natoli's semi-empirical rule [4, 5]), the symmetry of the absorbing atoms [6], the charge state and the potential surrounding it [7, 8]. While qualitative information from XANES spectra was obtained already in the early sixties on the basis of the edge position and on the presence of particular fingerprint peaks [9], the quantitative structural determination from a XANES spectrum is much more demanding than in the EXAFS (Extended X-ray Absorption Fine Structure) case [10-12]. Theoretical analysis of XANES spectra requires the

evaluation of the photoelectron wavefunction for a discrete (pre-edge features) and continuous parts of the spectrum. Nowadays several reliable program packages implemented *ab initio* approaches to calculate a XANES spectrum for a given atomic configuration. Feff [13-15], FDMNES [16, 17], Wien2k [18], ADF[19], Orca [20], XSPECTRA [21] are the most popular ones. Quantitative analysis of the atomic structure is based on the comparison between calculated and experimental spectra. The discrepancy between the two curves is evaluated as the Euclidean (L_2) norm (i.e. the integral of the squared difference between the theoretical and experimental XANES spectrum). For EXAFS analysis, the R-factor analysis is a more common technique used to compare theoretical and experimental data [22]. The following strategies can be implemented for structure refinement from X-ray absorption spectra:

(1) The fingerprint approach. Herein, the measured spectrum is compared with a set of experimental or theoretical reference spectra, focusing the attention on the existence of pre-edge peaks, the magnitude of the main maximum (the so-called white line), the edge position and the shapes of the first spectral features beyond it [7]. The best model in terms of the L_2 norm is attributed to the correct one [23-26].

(2) Gradient descent approach. This strategy includes many branches with trial and error steps to optimize the structure. It is currently utilized in the software for EXAFS data analysis, e.g. GNXAS [27] and Viper [28]. The most popular software, Artemis, which is a part of Ifeffit program package [29] uses a Levenberg-Marquardt steepest descent algorithm to find the values of the parameters s which minimize the chi-squared value. MXAN applies the same algorithm for the quantitative analysis of XANES spectra [30, 31] exploiting the full multiple scattering calculations in the *muffin-tin* approximation [32-34]. The optimization process is based on the calculation of the square of the residual function in the parameters space. Herein, the multi-dimensional minima search procedure is performed by the MINUIT function developed in CERN [35]. Typically, hundreds of theoretical spectra are needed to obtain the best fit of the data [36 5].

(3) Indirect approach: prediction of a XANES spectrum for a given set of structural parameters. Approach (2) may lead to inconsistent results when several local minima exist in the region of variation of structural parameters. The repetition of the descendent procedure, starting from different initial conditions, can help to find the global minimum but increases significantly the computational time. Alternatively, the studied space of structural parameters can be sampled with some points where the related XANES spectra are calculated *a priori* before comparison with experimental data. Such an approach is realized in FitIt [1, 37]. It performs the multi-dimensional polynomial interpolation of spectra as a function of determined structural parameters. Once that the polynomial interpolation has been constructed, the minima of a discrepancy between the experimental XANES spectrum and the interpolated one are searched, by varying the structural parameters, under the application of a gradient descendent algorithm. In this way, the required number of *ab initio* calculated XANES spectra is considerably smaller with respect to approach (2).

(4) Direct approach: prediction of a set of structural parameters for a given XANES spectrum. Approach (3) can be inverted to establish the correspondence between points in an experimental XANES spectrum $\mu(E_i)$ and the related structural parameters. This procedure eliminates the trial-error search of structural parameters providing the minimal discrepancy between the theoretical and experimental XANES spectrum. The structural parameters are predicted directly as a function of the $\mu(E_i)$ variables (see section 3.5.2 in [38]). Some authors have already described the application of the direct method to the XANES analysis. Zheng et al. [39] created a large dataset of computed references

of XAS spectra calculated in the *muffin-tin* approximation. Afterward, they applied an Ensembled-Learned Spectra-Identification (ELSI) algorithm to predict the oxidation state and the local environment for a wide set of compounds. The algorithm combines 33 weak "learners" comprising a set of pre-processing steps and a similarity metric, and it can achieve up to 84.2% accuracy. Timoshenko et al. [40] used supervised machine learning (SML) to study the 3D structure of supported platinum nanoparticles. The authors constructed an artificial neural network (NN) in Wolfram Mathematica, using an input layer composed of 129 nodes and two hidden layers having 339-387 nodes in each. A hyperbolic tangent function was used for the activation function while the NN was trained creating a dataset of *ab initio* XANES calculations on different size/shape nanoparticles. Their approach allowed us to reconstruct the average size, shape, and morphology of well-defined platinum nanoparticles from their XANES spectra. The same group successively extended the method to the EXAFS part of the spectrum processed by wavelets [41, 42].

For some experiments, for example, time or space-resolved measurements, a large series of spectra must be analyzed, and each spectrum can be considered as a superposition of a few components. The first method used in this situation is commonly known as a Linear Combination Fit (LCF) [7, 43, 44]. The limitation of LCF consists in the need to know the exact number of reference compounds and their spectra in advance. An alternative approach is the Principal Component Analysis (PCA). Both LCF and PCA were implemented in Athena [29]program package. Fernandez-Garcia et al. [45], for the first time, used a factor analysis procedure on a series of XANES spectra of CuPd/KL zeolites acquired during a temperature-programmed reduction (TPR). Herein they found the correct estimation of pure species in the Cu chemical matrix using PCA technique; then they identified the pure spectra and concentration profiles applying a self-modelling curve resolution algorithm called Iterative Target Transform Factor Analysis (ITTFA) [46, 47]. Afterward a novel technique for the decomposition of spectra, called: Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) [48, 49] appeared and it is rapidly spreading in the field of the analysis of XANES dataset [50-57]. Basically, MCR-ALS is an iterative algorithm employed to decompose an experimental dataset in the product of an optimized pure (i.e. with a chemical/physical meaning) spectral matrix for its related concentration profiles using a set of primary results delivered by such methods as SIMPLISMA [58] or EFA [59] and following an alternating least squares optimization of both spectral and concentration profiles under constraints [60]. Clearly, the iterative refinement of initial guesses often suffers poor convergence and an of certain lack of robustness [61]. However, it has been demonstrated that the usage of some spectral or concentration profiles, as initial guess, having some typical features proper of the dataset, increased the probability to reach the convergence[62-65]. Mathematically speaking,the task of decomposition of spectral series into a set of spectral and concentration values is an indirect problem characterized by multiple solutions, therefore different approaches exist which provides solutions corresponding to different physical, chemical or mathematical constraints.

The FitIt core, provides the indirect approach for structural refinement from XANES [1] and implements a PCA-based mathod for the analysis of spectral series [66]. It was written in Borland C++ Builder. Nowadays it is difficult to maintain the code due to outdated developer's instruments. According to the statistics [67], Python (including Anaconda, keras, scikit-learn) revealed as the most popular programming language for machine learning applications. In particular, Jupyter notebooks are actively used in the scientific community and allow them to perform calculations remotely. These

reasons induced us to develop pyFitIt exploiting the Python library and creating the user interface in the form of Jupyter notebooks based on ipywidgets.

We have realized both the approaches,direct and indirect,for the XANES prediction and the related structural refinement using novel machine learning methods for non-uniform distribution of points and multiple dimensions. These methods,developed for the Big Data (or multidimensional data) tasks, found to be appropriate [38]. Moreover, we have implemented somemethods based on the Singular Value Decomposition (SVD) for the analysis of a large series of experimental spectra. Herein, PyFitIt allows the user to perform PCA and estimate the retrieved set of spectra and concentration profiles by a visual approach based on a set of intuitively clear sliders.

The manuscript is organized as follows. Section 2 describes the methods implemented in the software. Details on the spectral decomposition via the SVD approach are provided in Section 2.1 and 2.2, while the machine learning algorithms are described in Section 2.3. The user interface, communication of local computer with the remote cluster and programming details are provided in Section 3. Sections 4.1 and 4.2 are devoted to some practical applications of PyFitIt. PCA procedure for two components is discussed in Section 4.1 while Section 4.2 describes the fit procedure for the structure refinement of the $Fe(terpy)_2$ molecules upon laser excitation.

2. Methods

Figure 1 summarizes the features implemented in PyFitIt. In sections 2.1-2.3 we describe which methods were used to implement these features.

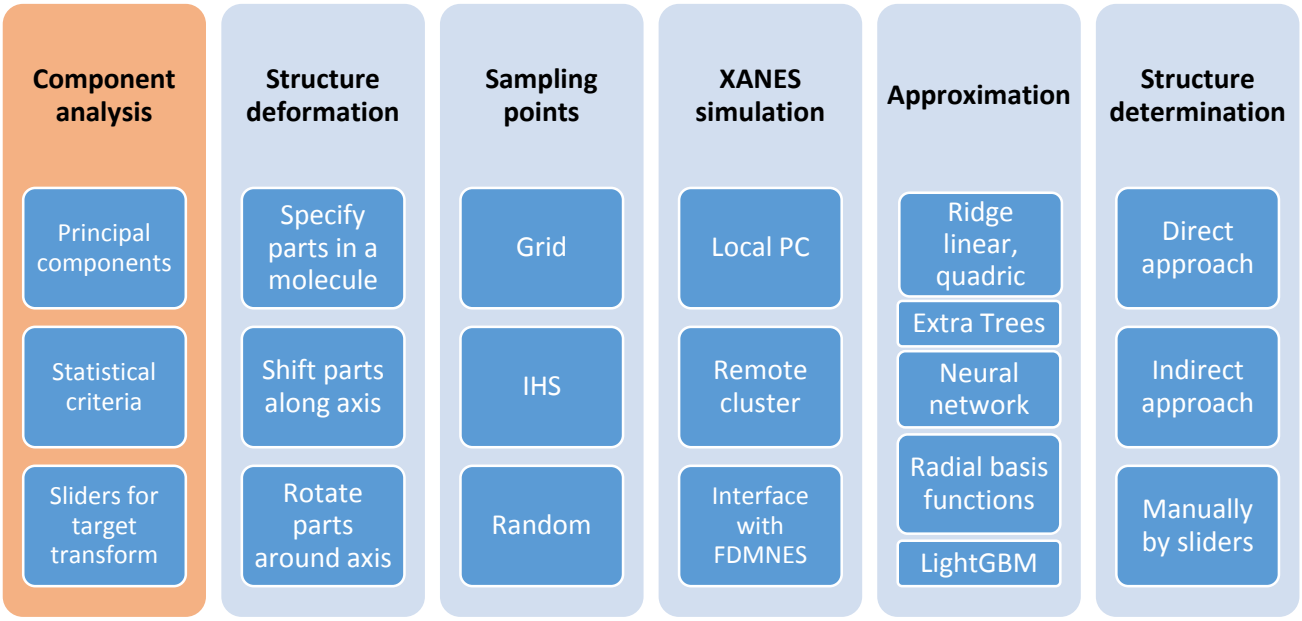

Figure 1. PyFitIt features. A python interfaced is provided to help the user for each step of the analysis.

2.1. Decomposition of spectra using PyFitIt

Usually, for time and spatially-resolved measurements, an experimental set of n spectra could be modelled as the weighted sum of some uncorrelated "pure" (i.e. independent) N (with $N < n$) spectra multiplied for their related concentration profiles. This kind of bilinear decomposition can be realised by means of Singular Value Decomposition (SVD) procedure as follow:

$$\boldsymbol{\mu} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{V} + \mathbf{E} \quad (1)$$

Where $\boldsymbol{\mu}$ is the reconstructed matrix of the experimental spectra composed by m energy points and N spectra, matrices $\mathbf{U} \cdot \boldsymbol{\Sigma}$ $(m \times N)$and $\mathbf{V}$ $(m \times N)$ represent, respectively, the matrix containing the absorption coefficients and the associated concentration values while matrix $\mathbf{E}$ is the residual data matrix that critically depends on the number of components (N) used to approximate $\boldsymbol{\mu}$. A detailed treatment on how this kind of decomposition can be realised, together with a more informative description of $\mathbf{E}$, is provided in Section 1.1 of the Supporting information (S.I.) text. The identification of the correct number of component, characterizing the experimental data matrix, represents a fundamental point in decomposition (1). Different statistical and empirical techniques have been developed to discriminate between components related to the real (observed in nature) signal and components containing only the statistical noise of the experimental spectra. Among them, the most widely used by the XAS community have been developed in PyFitIt and are described in Section 2.2 and in Section 2 of the S.I.

It is worth noting that matrixes, appearing in equation (1) and obtained numerically by SVD, must be considered as a mathematical solution of the problem without any chemical/physical meaning. For this reason, it is necessary to transform them in the real, observed in nature, spectral and concentration profiles. To this purpose, a transformation square matrix $\mathbf{T}$ can be inserted in equation (1) exploiting the identity matrix $\mathbf{I} = \mathbf{T} \cdot \mathbf{T^{-1}}$ as follow:

$$\boldsymbol{\mu} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{T} \cdot \mathbf{T}^{-1} \cdot \mathbf{V} \quad (2)$$

In PyFitIt, the elements T_{ij} of matrix $\mathbf{T}$ can be modified with interactive sliders until reasonable spectra $\mathbf{S} = \mathbf{U} \cdot \boldsymbol{\Sigma} \cdot \mathbf{T}$ and concentration profiles $\mathbf{C} = \mathbf{T^{-1}} \cdot \overline{\mathbf{V}}$ are found. In case that $\mathbf{T}$ is a singular matrix, its inverse can be determined using the Moore-Penrose pseudoinverse technique [68]. However, without any restriction, the number of elements of matrix $\mathbf{T}$ that can be adjusted by user is equal to N^2. In order to reduce this number, some constraints can be imposed. On this basis, the first or the last experimental spectrum (or both) can be set as pure spectra (i.e. boundaries conditions) and, at the same time, the normalization of the extracted spectral profiles can be used as another method to reduce the number of free parameters. The detailed description on how this kind of constraints are mathematically realized is reported in section 1.2 of the S.I. text.

Finally, taking in account each possible combination of the constraints described above, it is possible to justify that the number of adjustable parameters (i.e. sliders in the program interface) that can be used to perform some transformation is given by the following formula:

$$N_{\text{param}} = \begin{cases} N^2 - N : \text{ Normalization imposed;} \\ N^2 - 2N + 1 : \text{ Normalization imposed and first/ last spectrum fixed;} \\ N^2 - 3N + 2 : \text{ Normalization imposed and both first and last spectrum fixed;} \end{cases} \quad (3)$$

In Section 4.1 we report one example where this method has been applied together with the application of the statistical criteria used to identify the correct number of components in a XANES dataset. A second example is reported, for completeness, in section 4.2 of the S.I text.

2.2. Statistical criteria to determine number of principal components.

How described in Section 1 in the S.I., matrix **Σ,** appearing in (1), is a diagonal matrix whose elements are called singular values. It is possible to demonstrate that each singular value s_{ii} is related to the eigenvalues (λ_i) of the covariance matrix of **μ** by the following relation [46]:

$$\lambda_i = {s_{ii}}^2/(m-1) \tag{4}$$

It is worth noting that each λ_i term corresponds to the variance associated to the i[th] component. This means that higher is the variance related to a determinate component then larger is its contribution in the reconstruction of the dataset. Vice versa, components characterized by a low variance will keep account for the noise contributes. The variance values obtained from **Σ** are used in different statistical tests aimed to determine the correct number of principal components (PCs) to consider (i.e. the components related to real signal and not to the noise). Some of the most popular are: the scree plot, the imbedded error function (IE-test), the factor indicator function (IND-function) and the Malinowski F-Test. IE-test and IND-function are described in Section 2 of the S.I. text

In the scree plot, the variance associated with each component is used as a criterion for accepting or rejecting a determined component. The variance can be plotted against the number of components and the position of the elbow on the curve determines the border between components having a real physical/chemical meaning and those related to the data noise. The latter are called secondary components and the associated eigenvalues are called secondary eigenvalues: ${\lambda_i}^0$.

The Malinowski F-Test [46, 69] is a statistical-based method applied to determine the true dimensionality of a dataset. It is based on the observation that the secondary eigenvalues expressed in the reduced form REV_i (see eq. (5)) should be statistically equal.

$$\mathrm{REV}_i = \frac{\lambda_i}{(m-i+1)(n-i+1)} \tag{5}$$

Because the reduced eigenvalues are still proportional to a variance, a Fisher test can be applied. The test starts from the smallest eigenvalue, clearly associated with the noise, and proceeds to the eigenvalues, with higher magnitude, until the first significant one (*i.e.* first signal-related one) is found [70]. The k[th] component is considered significant on the basis of the Fisher test applied on its related standardized F-variable:

$$F(1, n-k) = \frac{\mathrm{REV}_k}{\sum_{k+1}^{n}\lambda_i}\left(\sum_{k-1}^{n}(m-i+1)(n-i+1)\right) \tag{6}$$

where the variable λ_i represents the noise-related (secondary) eigenvalues. If the percentage of significance level (%SL), associated to this variable, is lower than a pre-fixed value (usually it is fixed to 5%) then the k[th] extracted component is accepted as a pure component.

Finally, it is worth remembering that the results coming from the F-Test, IND and IE factors must be considered with caution. These statistical criteria critically depend on the amount of noise in the dataset. In fact, a deviation from the real number of chemical/physical component occurs when the experimental noise (which is not known in advance) is close to the variation in the data, or when some

component species have indistinct spectral features [51] or even when their fractional weight, in the data mixture, is statistically constant [70].

2.3. Implemented machine learning algorithms

A XANES spectrum μ(E, **P**) can be considered as a multi-variable function of *k* -structural parameters **P** = (p_1, p_2, …, p_k) (i.e. distances, angles, coordinates, and different deformations). However when the number of employed parameters becomes large (more than 3) or in case of their broad limits of variation (for example more than 0.4 Å for the first coordination shell distances),the application of polynomial interpolation is problematic. In fact, high order interpolation polynomial exhibit large oscillations known as Runge's phenomenon, while low order interpolation polynomial has a poor approximation quality. Machine learning algorithms work much better in this case and provide a good approximation in the whole region of variation of structural parameters. To predict a spectrum, consisting of many points, Regression Models with multiple target variables should be applied. We have tested ridge linear and ridge polynomial regressions, radial basis functions, and Extra Trees methods. Alternatively, one-target regression models (e.g. gradient boosting of trees) can be also used for each energy point. Similarly to the task of interpolation, machine-learning algorithms require a training set to be calculated first. It consists of theoretical XANES spectra calculated for a given sets of geometrical parameters $\mathbf{P}_1$, $\mathbf{P}_2$, …, $\mathbf{P}_N$. Further approximation depends on the selected method of machine learning. For approximation, in the training set we normalized the variation of the spectral profiles for each each energy point to the interval [-1…1]. Analogously, the range of variation of each structural parameter is scaled to the [-1..1] interval.

Ridge regression approach is described in Section 2 of S.I. The interpolation method based on Radial Basis Functions is a well-proven mesh-free method [71]. In this case, the unknown function $\hat{\mu}(\mathrm{E}, \mathbf{P})$ is represented in terms of a set of basis functions characterised by certain factors and polynomial terms as follow:

$$\hat{\mu}(\mathrm{E}, \mathbf{P}) = \sum_{i=1}^{N} w_i(\mathrm{E}) K(\|\mathbf{P} - \mathbf{P}_i\|) + \mathrm{Polynomial}_E(\mathbf{P}) \quad (7)$$

where $K(\mathrm{r})$ – is the radial basis function, $\mathrm{Polynomial}_E(\mathbf{P})$ – is a polynomial function of *k*-structural parameters (p_1, p_2, …, p_k) with energy dependent coefficients. The unknown factors w_i and the polynomial coefficients are obtained by the least squares method or by ridge regression. Every basis function is a function of distance from the training set point $\mathbf{P}_i$. In our task, some good results were obtained using linear basis functions and a second-order polynomial (we use a polynomial function extracted by the ridge quadric regression method).

ML methods based on decision trees divide the space of geometric parameters into non-intersecting rectangles, in each of them the objective function μ(E, **P**) is approximated by a linear expression $\hat{\mu}_j(\mathrm{E}, \mathbf{P})$(where j is rectangle index) using the least squares method. Each node of the decision tree contains a condition $p_j < t$ for one of the geometrical parameters p_j which divides the training subset into parts (see Figure 2).

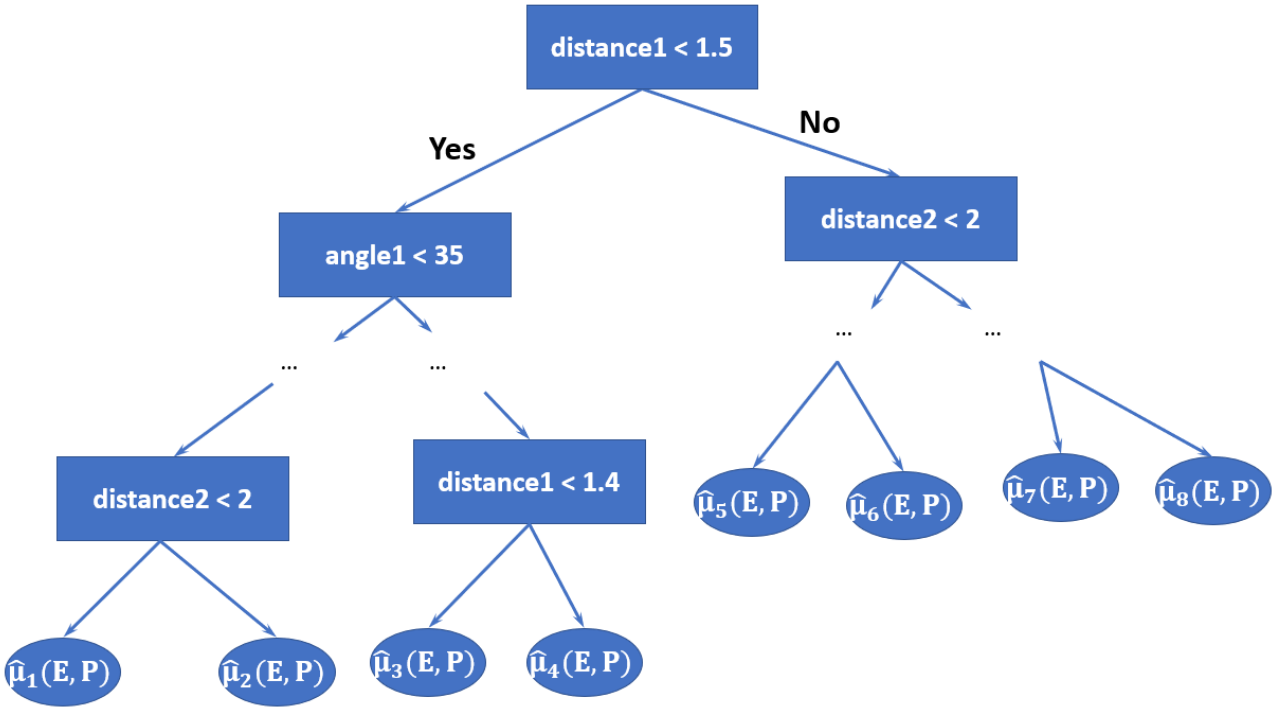

Figure 2. Illustration of a decision tree for dividing a set of theoretical spectra in groups (ellipses, called leaves) corresponding to the combination of restriction on structural parameters (rectangles, called nodes). The branching process continues until the sample size of a node is more than 10 spectra.

Initially, the overall training set $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N$ is randomly divided into training and test subsets which are used to construct and evaluate the quality of the tree correspondingly. Each leaf of the tree (ellipses in Figure 2) contain linear approximations of XANES $\hat{\mu}_1(E,\mathbf{P}) \ldots \hat{\mu}_l(E,\mathbf{P})$, which are constructed based on the training subset for a given leaf (l – is the number of leaves in the tree). Thus, a single tree is a form of specifying a piecewise linear function of geometric parameters $p_1, \ldots, p_k$. The values of separator t for each node are selected from a random set based on the comparison of the results of approximation $\hat{\mu}_i(E,\mathbf{P})$ with exactly known values for the test subset. In the Extra Trees method [72] several random trees are constructed. based on a random subsample of the training set. The results of approximation from all the trees are then averaged. Such methoa d of combining machine learning models is called bagging.

Recently, some more accurate algorithms of model's combination, such as gradient boosting, have been developed. In the gradient boosting of trees, each next tree improves the approximation of previous ones. For this application, the same tree-building algorithm is applied to the training set with weighted samples. This means that the probability that $\mathbf{P}_i$ is included in the training set for the new tree is higher if the error of approximation coming from the previous combination of trees is larger. Unlike the Extra Trees approach, the trees are further summed with different coefficients. Weighted samples correspond to XANES spectra while coefficients correspond to Trees itself. During the construction, of the next tree associated with the gradient boosting, the coefficient of the linear combination is adjusted only for this tree. The next coefficient is chosen on the basis that the addition of a new tree is similar to the step of the gradient descent method. This procedure explains the origin of the name of this method.

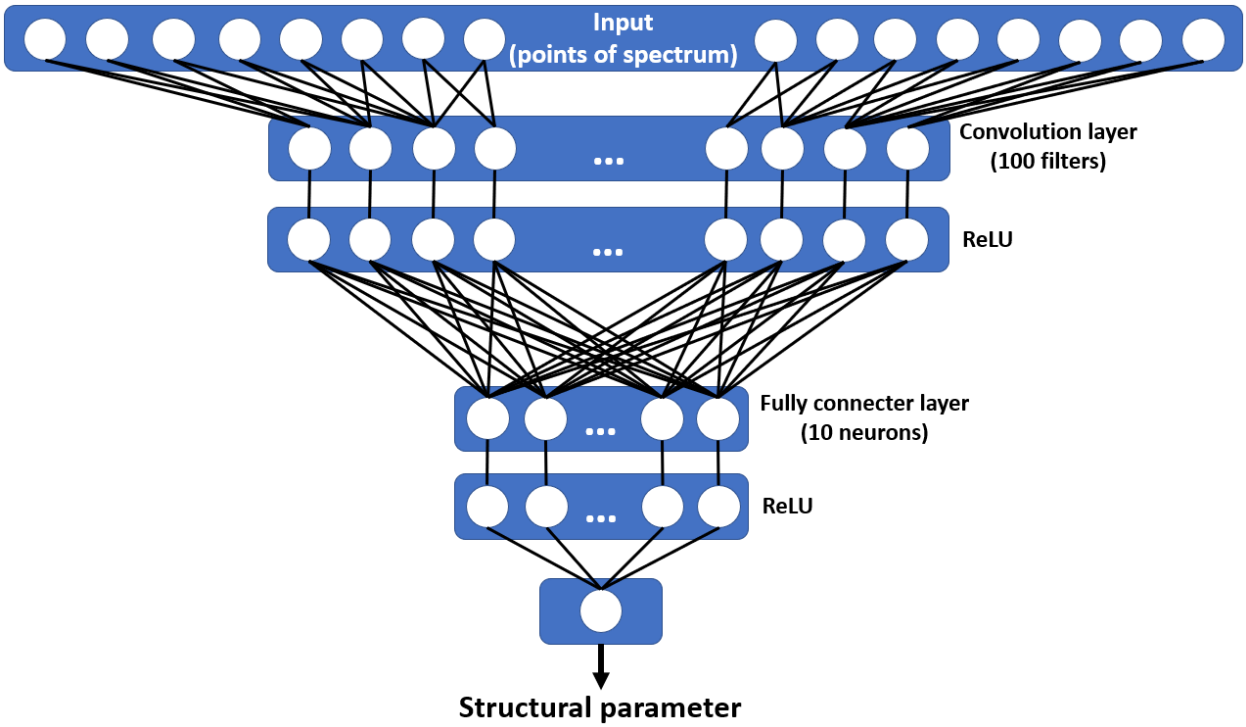

Figure 3. Scheme of the Neural Network implemented in PyFitIt

Finally, we constructed the neural network (NN) for the direct method. The NN was constructed by means of the TensorFlow library associated with the Keras framework. We choose a sequential neural net model with a convolutional first layer (100 filters) then followed by 3 dense layers with sizes of 100, 10, 1 (the activation function is ReLU). Figure 3 demonstrates the sketch diagram of the neural network. To control the quality of the approximation of the function and to estimate the error bar of the predicted values, we apply the cross-validation technique. In order to optimize the weights, we used the SGD optimizer with the following parameters: 50 epochs, batch_size=32.

2.3.1. Selection of sampling points in the space of parameters

Sampling points $\mathbf{P}_1$, $\mathbf{P}_2$, …, $\mathbf{P}_N$ for the training set can be selected in several ways. In PyFitIt different grid building topologies can be chosen: the random distribution of points, grid and improved Latin hypercube sampling (IHS). Figure 4 explains the differences between these approaches.

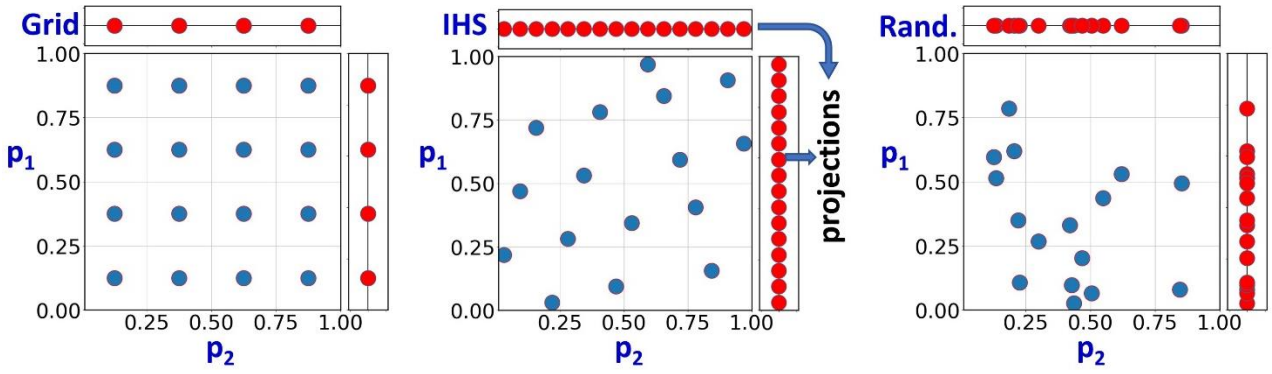

Figure 4. 16 sampling points distributed in the related two-dimensional space of parameters (p_1, p_2) according to the Grid, IHS and Random methods. Red points show the projections of the sampling points on corresponding axes.

It is possible to assert that the IHS approach is superior over grid, since the distance between projections of points along each dimension is smaller. This fact, along with a randomized selection procedure, is important for many dimensional interpolation techniques. Using the same number of points as in grid, we obtain a higher quality of approximation. In the random approach, points are selected randomly in the parameters space. Thus, for a low number of points, problems arising from their grouping ,in a given region of space, are very probable. For a large number of sampling points, the random approach is comparable to IHS. Its additional advantage is associated to itssimplicity. The IHS sampling is constructed specifically for a determined number of parameters and the addition of extra points requires changing all of them (i.e. a larger computational time required to calculate new training set). In a random approach any additional number of points can be added in any region of the space – (e.g. where XANES spectrum variation is larger).

2.3.2. Comparison with an experiment: inverse and direct ML approaches

The methods described in Section 2.3 can be applied in two ways to analyze experimental data: to predict: (a) a XANES spectrum for any set of geometrical parameters; (b) the geometrical parameters **P** for any XANES spectrum. We call the first approach indirect and it considers each point of the XANES spectrum as a function of the structural parameters **P** (i.e. $\mu(E, \mathbf{P})$). The task of the structural fitting is solved when a given set of parameters **P** minimizes the difference between approximated and experimental spectra.

We call the second approach (b) direct since the machine learning model is trained to predict directly the geometry. In the direct approach, the XANES values $\mathbf{M} = (\mu(E_1), \mu(E_2), \ldots, \mu(E_n))$ play the same role of the structural parameters $\mathbf{P} = (p_1, p_2, \ldots, p_k)$ in the indirect approach. It follows that a small number of structural parameters can be considered as functions $\mathbf{P}(\mathbf{M})$ of a large number of spectral points $\mu(E_1), \mu(E_2), \ldots, \mu(E_n)$. The choice of these spectrum points as arguments leads to a multicollinearity problem. Thus, the linear regression cannot be used without regularization methods, while ridge regression is required. Extra Trees and Gradient Boosting can be applied without modifications. In addition, convolutional neural networks can be successfully applied to the direct method.

The direct method has several advantages. First, this approach automatically divides the geometry parameters into two parts: those, which can be accurately determined from the XANES spectrum, and those that cannot be precisely determined. The second advantage is that there is no need to select the comparison metrics between predicted and experimental spectra. The problem of metric arises from the fact that the minimization of objective functions based on different metrics (e.g. the L_2 norm between spectra or their derivatives) leads to different final structures and it is not clear which of them is better.

The problem that characterised both the direct and the indirect approaches is the possible multivaluedness of the retreived parameters , seeFigure 5 for and example. The best way to avoid this problem is to select a proper parameter space.For example, considering the structure showen in Figure 5, it is possible to define a new set of values using (p1+p2)/2 and (p1-p2)/2 and assuming only positive variations . Sometimes this is a complicated task and is not convenient.

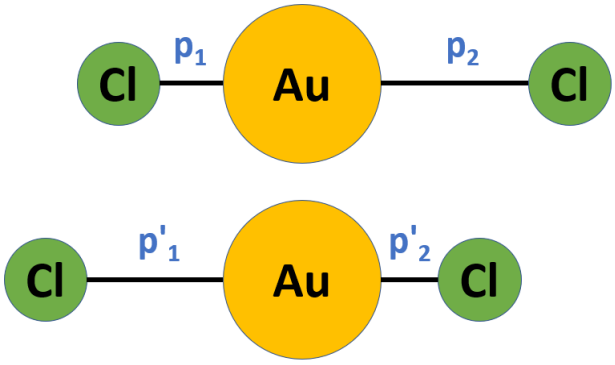

Figure 5. Two structures of $AuCl_2$ complex which have formally different values of structural parameters $\mathbf{P}_1 = (p_1, p_2)$ and $\mathbf{P}_2 = (p'_1, p'_2)$. Though under the condition of $\mathbf{P_1} \neq \mathbf{P_2}$, thecorresponding spectra are identical (i.e. $\mu(E, \mathbf{P_1}) = \mu(E, \mathbf{P_2})$ if $p_1 = p'_2$ and $p_2 = p'_1$). This fact makes impossible the prediction of single value for p_1 and p_2 separately in the direct method.

In such cases, this problem can be overcome in two ways. The cclassification procedure suggests to divide the interval of variation of each parameter in parts, then the ML method learns how to predict the probability that a parameter belongs to each given part of the variation interval. For each parameter, the prediction is performed independently from the predictions made for the other parameters. Another possibility is to predict always a single valued function. In particular, radial and angular distribution functions centered on the absorbing atom can be used instead of several first shell structural parameters.

The direct method is an attractive tool for the structural analysis of X-ray absorption spectra. However, we noted that, currently for XANES spectra, the direct method cannot be used as a black box. The reason must be found in the systematic discrepancies between calculated and experimental spectra even for the best fit model. Due to the simpler theory level required, this problem is negligible in EXAFS region, where theoretical simulations can reproduce experimental data with excellent accuracy. High fit quality, which is common in EXAFS analysis or X-ray diffraction, is currently unavailable for XANES. Therefore, the direct method can lead to unphysical results, when points of experimental data escape from the region of variation in the theoretical training dataset. To reduce such errors, we introduced in PyFitIt also an option able to perform the fitting difference spectra.

3. Installation and workflow

The software is published in The Python Package Index (PyPI) and therefore can be installed using the following command from the python environment (for example, Anaconda):

pip install --upgrade pyfitit

The most update information about the code, for example Notebooks, and related datasets can be downloaded from http://hpc.nano.sfedu.ru/pyfitit/.

3.1. Decomposition of the experimental data into pure spectra

The Jupyter Notebook PyPCA contains a set of functions for the spectral analysis and the decomposition of a XANES series using the approach described in section 2.1. PyPCA requires, as input, a txt file with the series of spectra measured on the same energy grid. The principal components associated with the input data and their related eigenvalues are extracted by means of SVD analysis. Afterward, these values are used to calculate and plot the statistical parameters (which are the Scree Plot, IND Factor, IE, the plot of the percentage of significance level (%SL) used for the F-Test).

PyPCA also plots each abstract component. In this way, it is possible to perform a qualitative/graphical analysis that can be helpful to discard the components related to noise effects. Finally, after the data normalization procedure, the Target Transformation module, see Figure 6, enables the direct access to the elements of the transformation matrix using sliders and toimpose determined constraints used to recover a *pure* set of spectral and concentration profiles.

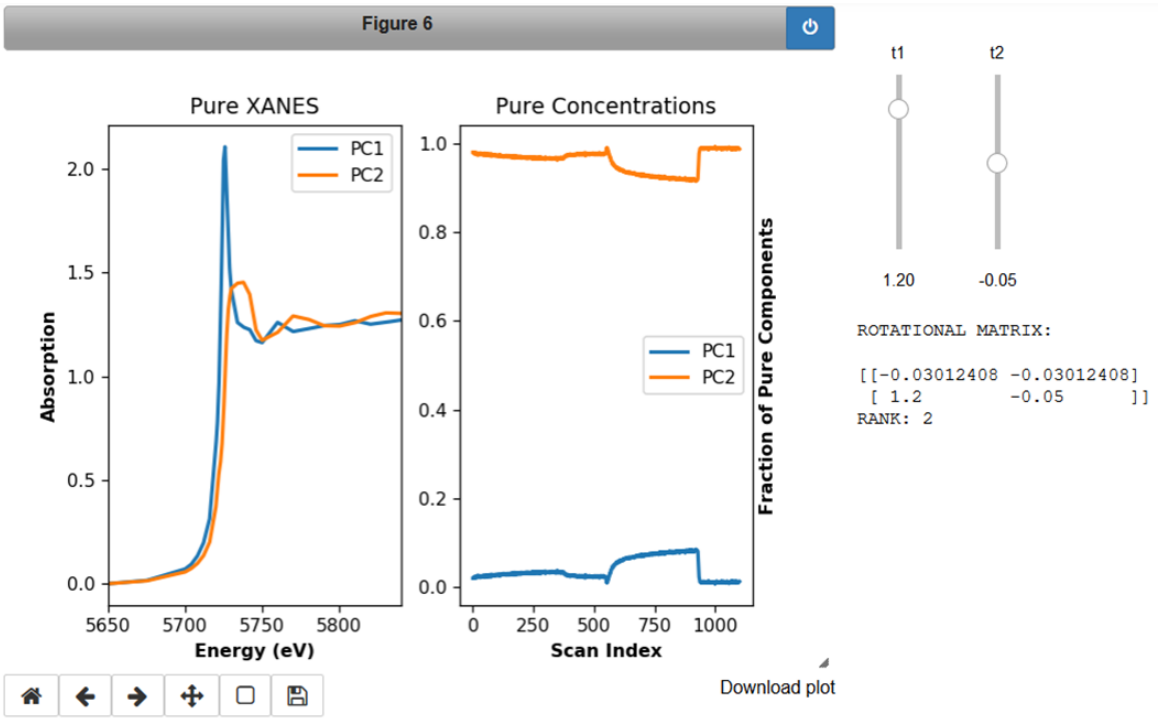

Figure 6: Output of the Target Transformation module. Moving sliders, it is possible to have access to each element of the transformation matrix used to retrieve meaningful spectral and concentration profiles from the experimental input data matrix.

3.2. The fit of the XANES spectra.

When the spectrum of a pure phase is known, it can be compared to the related theoretical simulations. Figure 7 shows a general workflow of PyFitIt for the quantitative analysis of some structural parameters associated with determined XANES spectra. In our examples, each step is confined within a separate Jupyter Notebook, but the user can call functions of the PyFitIt library in any place of his own code. These notebooks help to: (i) specify the structural deformations, (ii) calculate the XANES spectra for each deformation and collect all data from separate calculation folders, (iii) train algorithm on a theoretical dataset, (iv) perform the fit of an experimental spectrum using the inverse approach (automatic search of the minimum value or manual adjustment of parameters using sliders) or direct approach (predict each parameter independently or the radial/angular distribution function).

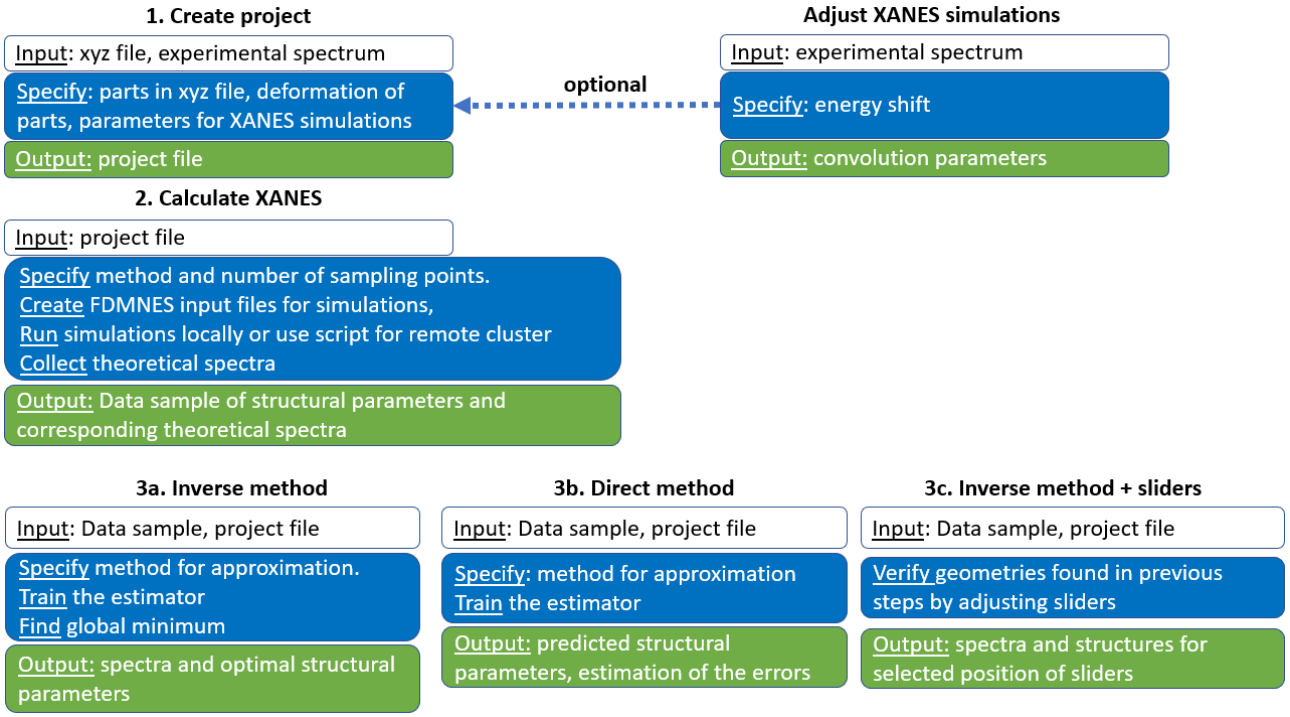

Figure 7: Workflow of PyFitIt

Users can follow this workflow by running Notebooks in series or provide necessary input for any given step and use it as stand-alone. In the Supplementary Section 3 of the S.I. we describe each step in more detail. When the automatic fit is performed (steps 3a and 3b in Figure 7) user can manually change the structural parameters looking at the effect on the related XANES spectrum. Figure 8 shows the panel program for such a fitting procedure.

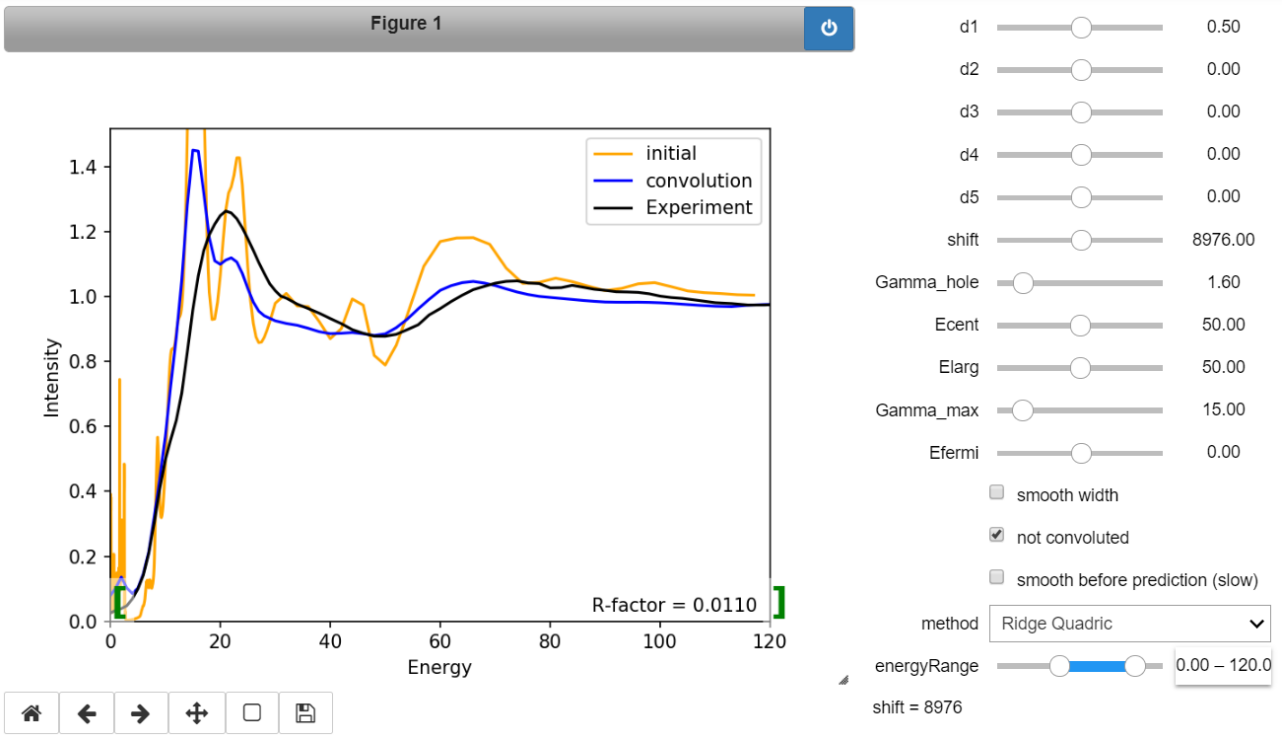

Figure 8. Process of fitting of structural parameters d1-d5 by means of sliders. The shift of the experimental energy scale and the parameters associated to the spectral convolution (taken from FDMNES) can also be adjusted.

In the XANES region, the contribution of the nonstructural parameters in the theoretical calculations can introduce a significant uncertainty on the results coming from the structural refinement. We recommend to select values of non-structural parameters using reference compounds with well-known structure. Additionally, PyFitIt provides a visual interface for varying the structural parameters in a broad range and monitoring the corresponding changes in the spectral profiles in order to identify systematic problems that have a non-structural origin.

4. Case studies

4.1. Spectral decomposition for Ce L_3 XANES in CeO_2/Pt system under redox conditions

Ce L_3 XAS spectra were measured at the SuperXAS beamline of the Swiss Light Source (SLS) at Paul Scherrer Institute, Switzerland using the setup described in [73]. Details are reported in Section 4.1 of the S.I. text.

Figure 9 shows the set of experimental spectra, recorded in the two redox cycles at 26 °C and 90 °C, consisting of 120 s of reduction by CO gas flow and then 60 s of oxidation by O_2. These cycles were repeated for every energy spectral point, selected by the monochromator. In each point, the Ce L_3 fluorescence signal was recorded with 0.3 s time resolution. This results in 1103 Ce L_3 XANES spectra. The differences between the successive spectra can hardly be visible from this figure and thus the quantitative analysis of the principal components is required to clarify the number of pure components in the spectra and their concentrations profiles.

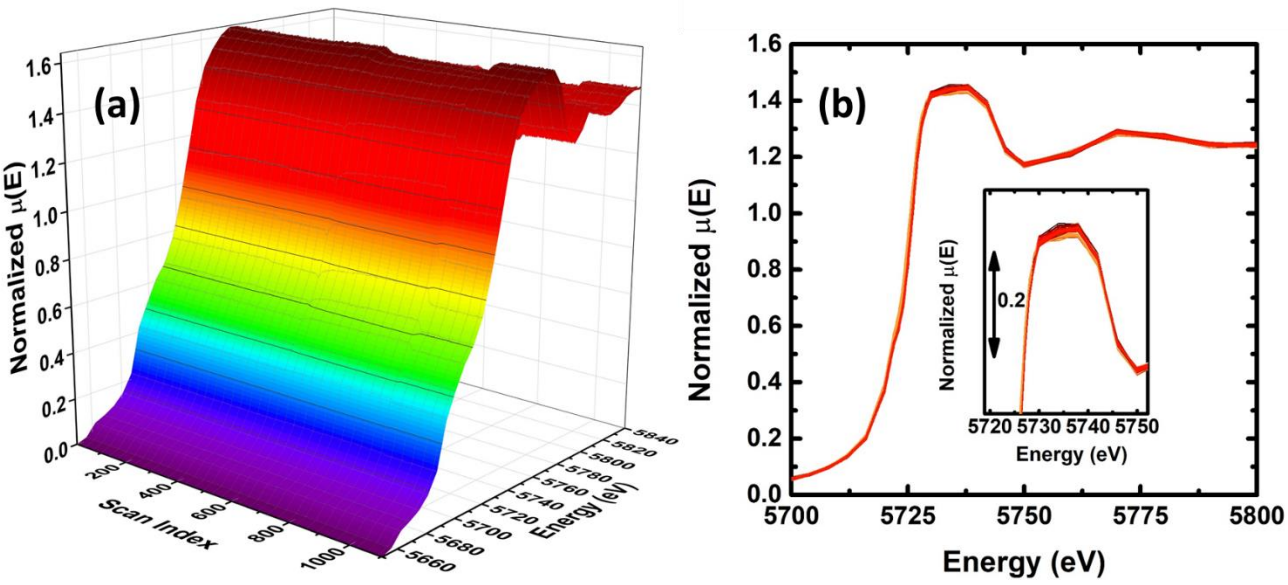

Figure 9: (a) 3D (energy, scan index vs μ(E)) and (b) 2D (energy vs μ(E)) representation of 1103 Ce L_3 XANES spectra. This dataset is used further as input for the analysis of the principal component and the following spectral and concentration decomposition.

In this example, we decompose an experimental dataset, reported in Figures 9 (a) and (b), constituted by 1103 spectra and 38 energy points (from 5650 to 5840 eV). To calculate an accurate value of σ

required for normalization (see equation S.6 in the S.I), the XANES spectra were interpolated with a smaller energy step 0.05 eV.

The statistical estimators plotted *versus* the number of independent components show different results. The scree plot in Figure 10(a) shows an elbow in the proximity of the third component while the magnitude of the fourth eigenvalue stands closely to the line grouping the noise-related components. The same result is provided by the IND factor, see Figure 10(b) which presents a minimum in the proximity of the third component. However, the difference between the second and third components in terms of IND is low. On the contrary, the IE plot and the F-Test, Figure 10(c) and 10(d), (executed with a %SL fixed to 5%) indicate the presence of only two components. Uncertainty remains also by inspecting the graphical representation of the abstract components, reported in Figure 11. Here, while the second component is characterized by a well-defined shape, the third component has less intense and evident signal features. This component seems to be related to the decrease of the white line intensity flattering the spectra that is typical for self-absorption in a fluorescent regime of measurements. These reasons justified our choice to select two PCs for the data decomposition.

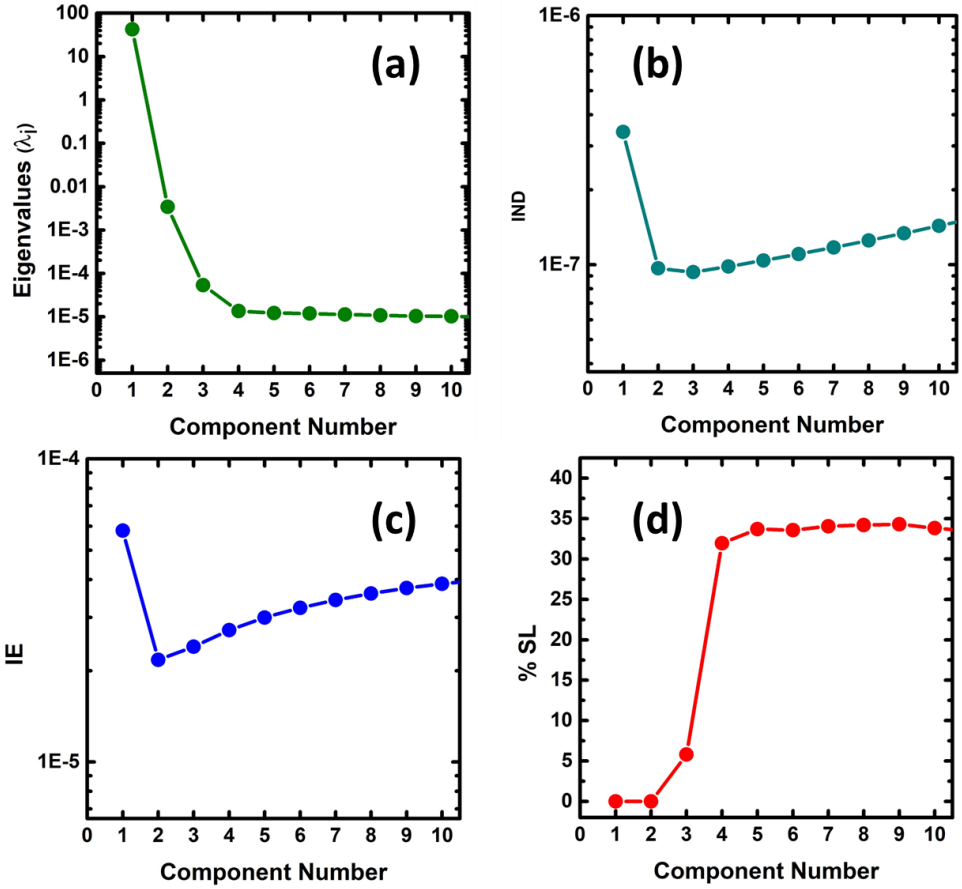

Figure 10: Statistical curves used for determining the number of principal components (i.e. pure spectra in the dataset from figure 9: (a) Scree Plot; (b) IND Factor; (c) IE; (d) plot of the percentage of significance level (%SL) used for the F-Test.

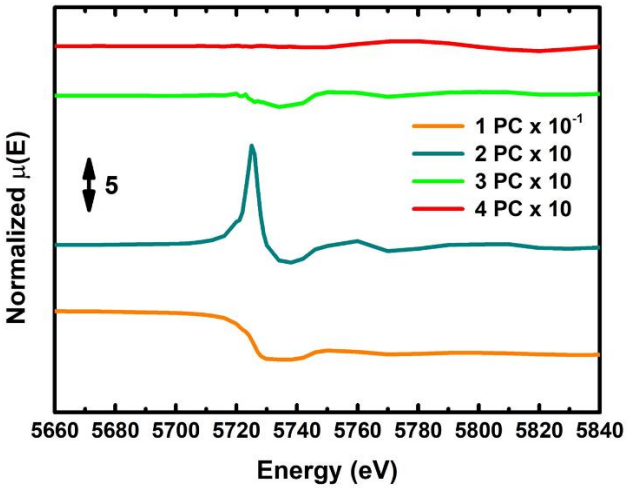

Figure 11: Abstract Spectral Profiles (PCs) extracted from the Ce L_3 XANES spectra shown in Figure 9. The first one and the second components are characterized by a higher magnitude, already shown in the scree plot.

In order to recover the two pure spectral profiles and their related concentrations, characterized by a chemical/physical meaning, we employed the following transformation matrix:

$$\mathbf{T}_{\text{Ceria}} = \begin{pmatrix} -0.03 & -0.03 \\ 1.20 & -0.05 \end{pmatrix}$$

As constraints, we applied the normalization of spectra, the positivity of the spectral profiles and we required that the concentration values are enclosed in the range between 0 and 1. Normalization reduced the number of free parameters from four to two. Finally, each parameter were varied in the range between -2 and 2 with a step of 0.05. One possible solution is reported in Figure 12(a) and (b).

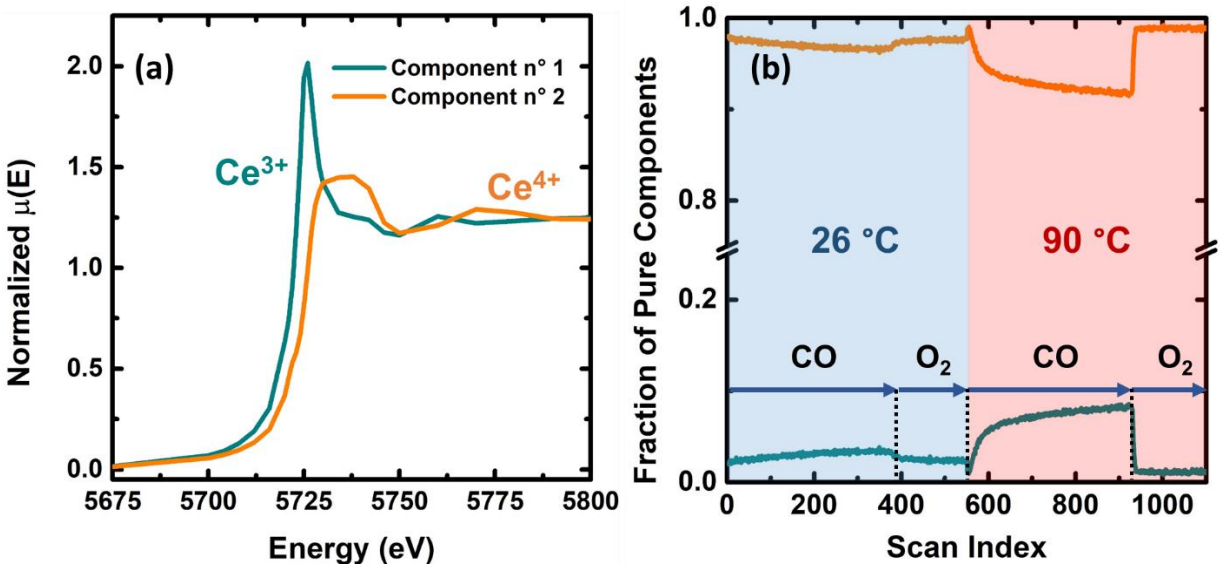

Figure 12: Pure Spectra (a) and Concentration profiles (b) recovered from the Ce L_3 XANES spectra shown in Figure 9. With the blue and orange lines are represented, respectively, the spectra and the related concentration profiles of Ce^{4+} and Ce^{3+} species.

Kinetic curves obtained in Figure 12(b) contain important information about the catalytical system. First, we can directly estimate the concentration of active oxygen atoms that are released from ceria lattice at two temperatures – 26 and 90 °C. Indeed, each released oxygen leaves two electrons in the system which reduce two ceria atoms from Ce^{4+} to Ce^{3+}. The amount of lattice oxygen that can be extracted in the CO atmosphere from CeO_2/Pt increases six times upon temperature increase. Another important feature is that at 26 °C ceria is not fully oxidized in the O_2 atmosphere and some fraction of Ce^{3+} atoms is constantly present in the lattice which was also indicated by another method [74]. Finally, the time-dependant concentrations reported in Figure 12(b) can be further used to calculate kinetic constants of the system which are important to understand the microscopical origin of the catalytic behavior of the ceria-based system [73].

It is worth noting that, in general, an original dataset can be reconstructed with the same quality fit using spectra and concentration profiles, with different shapes, by varying each element of the corresponding transformation matrix $\mathbf{T}_{\mathrm{Ceria}}$. This fact is known as rotational ambiguity [61]. In order to reduce it, different further constraints can be imposed. In the case of the Ce dataset, the transformation matrix is constituted by two free parameters. This means that the areas of feasible solutions (AFS) can be represented graphically (i.e. a 2D plot) and limited by further specific constraints. In this case, we considered only the couples of (T_1, T_2) able to isolate only the "pure" non-negative spectra, with a white line lower than 2.5, and characterized by some concentration values between 0 and 1. The results of this calculation are reported in Figure 13(a,b).

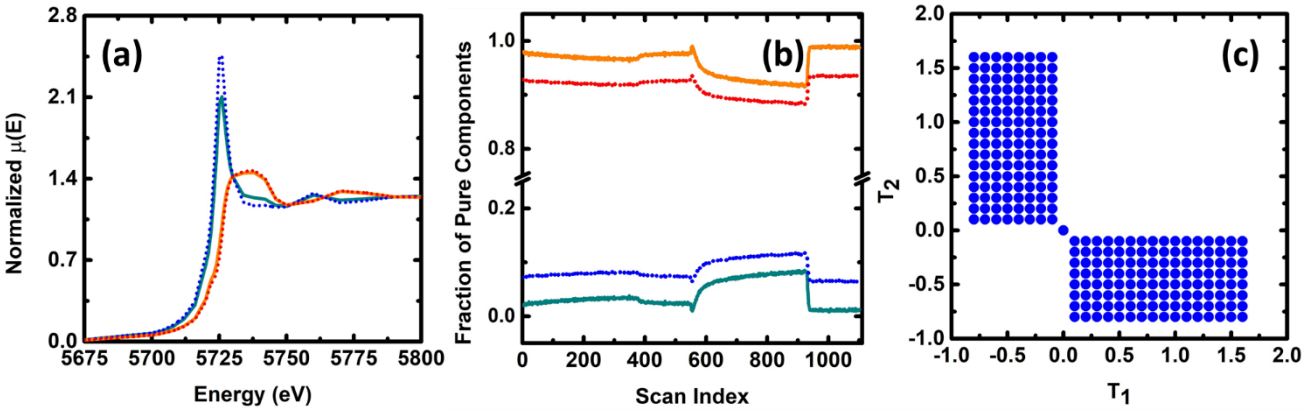

Figure 13: Pure spectral (a) and concentration profiles (b) obtained by setting both the free elements of the matrix $\boldsymbol{T}_{Ceria}$ respectively to : 1.20 and -0.05 (continuous lines) and to 1.60 and -0.15 (dotted lines) .(c) Related area of the feasible solution obtained selecting a range of variation for variables T_1 and T_2 between -2 and 2 with a step of 0.1. The following constraint have been considered: 1) non-negativity of the spectral profiles; 2) maximum intensity of the spectral white line fixed to 2.5; 3) concentration values included in the region between 0 and 1.

From Figure 13(c) it is possible to see the presence of a general "symmetry" in the AFS representation. This behavior can be explained on the basis that there is no inherent order on the columns of the pure spectral matrix **S** and the corresponding rows of **C**. The interchanging the two columns of **S** and rows of **C** is realized by permuting the two columns of the transformation matrix (i.e. swapping the two free elements of $\mathbf{T}_{\mathrm{Ceria}}$). From this example it is, possible to understand that, without a set of appropriate constraints, it is literally impossible to identify a unique solution. For this reason PyPCA, must be seen as a tool to perform an estimation of the pure spectral and concentration profiles which characterized the experimental data matrix. Clearly, a diminution of the ambiguities

associated to the variation of the sliders can be realized using some spectral and concentration references. On this basis, each slider of matrix $\mathbf{T}_{\mathrm{Ceria}}$ can be moved until the best agreement between the reference and the reconstructed spectrum is obtained. On the other hand, the same procedure can be applied to the recovered concentration profiles.

4.2. Structural refinement of the Fe(terpy)$_2$ excited state

Spin-crossover 3d metal complexes are potential candidates for molecular switches, novel data storage devices and optical displays [75]. Iron complexes with octahedral coordination can exist either in a low-spin (LS) or a high-spin (HS) state, depending on the temperature or pressure. Green light pulse can trigger LS to HS transition in solution of $[Fe(terpy)_2]^{2+}$ (terpy: 2,2′:6′,2″-terpyridine) already at room temperature [76]. To understand the fundamental processes upon electron transitions and improve the photoswitching parameters, a detailed characterisation of the excited state is required. In particular, the spin state and its lifetime should be determined at first. Recently, the high spin state of Fe(terpy)$_2$ was characterized by ultrafast time-resolved XANES [77]and the ^{5}E quintet state was identified after irradiation through the laser pulse. Later, Vanko et al. [78] discussed the difficulty in the selection between ^{5}E and $^5B_2$ candidate quintet states. One of the difficulties was that the standard DFT predicted $^5B_2$ to be the more stable while the more sophisticated CASPT2 approach predicted ^{5}E state to be at 150 meV lower in energy. However, geometry optimization is almost impossible at such a high level of theory. The quantitative fitting of the time-resolved XANES data offers a new independent source of structural information complementary to the EXAFS fit and quantum chemistry simulations. We have selected the problem of the structural analysis of the HS state of Fe(terpy)$_2$ as a demonstration of the capabilities of PyFitIt.

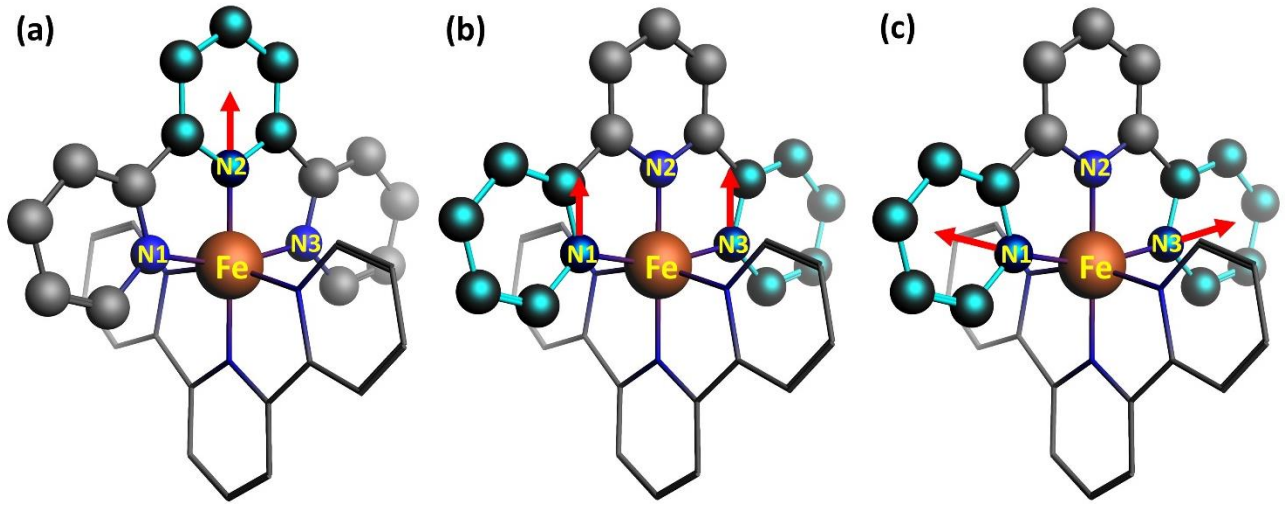

Figure 14. Three structural deformations for one terpy ligand of Fe(terpy)$_2$ molecule. (a) the shift of the axial ring along the Fe-N2 axis. (b) The shift of two equatorial rings along the direction of the Fe-N2 bond. (c) the shift of two equatorial rings along Fe-N1 and Fe-N3 bonds correspondingly. A total number of structural parameters for fitting equals 6 – three for each ligand. Atoms are plotted with spheres only for the upper terpy ligand.

Figure 14 shows the Fe(terpy)$_2$ molecule. To model the distortions of the structure upon spin transition we have split terpy ligand into three rings which are moved independently. Six degrees of freedom were considered in the fit. Three of them for one terpy are shown in Figure 14 and the rest

are identical but for another terpy. The first deformation shown in panel (a) corresponds to the translation of the axial ring along the Fe-N_2 axis. The second deformation is the translation of two equatorial rings along the same Fe-N_2 direction. Finally, the third deformation is the simultaneous symmetric elongation of Fe-N_1 and Fe-N_3 bonds for the equatorial rings. The amplitude of all deformations was set equal to 0.8 Å in the range -0.3 Å … 0.5 Å relative to the crystallographic low spin structure. These values correspond to distances of 1.58 Å … 2.38 Å for the axial Fe-N bond and 1.68 Å … 2.48 Å for the equatorial Fe-N bonds.

The following part of the code is an example of how the deformations are specified in the project. The *xyz* file is stored in the object *m* of class *Molecule*. Method *setParts* groups atoms according to their numbers, specified by the user. Variable *deformation* stores the name for the corresponding slider in the structural fit (see Figure 8). The *Axis* for the translation is specified through the difference of atomic coordinates. Finally, the method *shift* is applied to the parts of the molecule to specify the deformation.

```
m = Molecule('Fe_terpy.xyz')
m.setParts('0', '1-9', '10-19', '20-29', '30-38', '39-48', '49-58')

deformation = d2
   part1 = 2; part2 = 3
   axis = m.atom[1] - m.atom[0]; axis = axis / norm(axis)
   m.part[part1].shift(axis*params[deformation])
   m.part[part2].shift(axis*params[deformation])
```

In the 6-dimensional space of parameters, we specified 729 sampling points according to the IHS scheme. In the case of a grid method for sampling, such a number corresponds to three points along each deformation, i.e. $3^6 = 729$ points. Despite the large values of deformation along each parameter (0.8 Å) the quality of approximation was sufficient for analysis even at such a low number of points. The average uncertainty between the calculated spectra and the approximated ones using radial basis functions is shown in Figure 15. The uncertainties were calculated by means of cross-validation approach when each quarter of points in the training set was excluded from the training and used further for the validation.

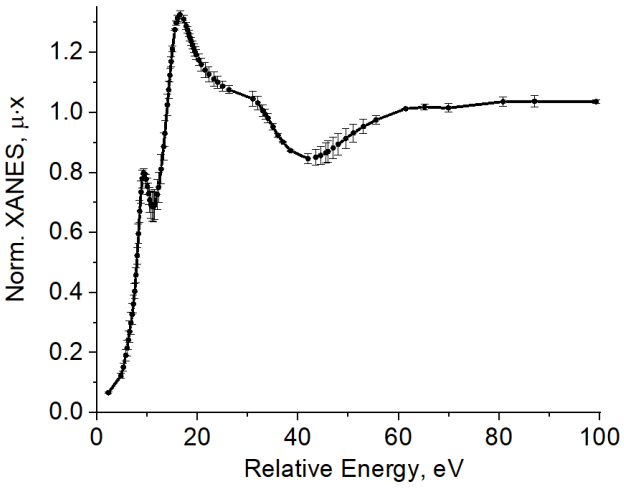

Figure 15: Cross-validation analysis for the training sample composed by Fe K-edge spectra for $Fe(terpy)_2$. Error bars show the mean error of approximation over the 6-dimensional space of parameters between approximated (RBF) and the exactly calculated spectrum.

PyFitit allows performing additional analysis of the approximation quality. For this purpose, we construct a straight line,with dense points, which runs between two points in the multidimensional space of parameters specified by the user. The calculated and approximated XANES in each energy point can be directly compared along this line. Such visual comparison helps to verify properly the training sample size, the points distribution ,and the approximation method. For our task, the best method for approximation was RBF. Methods such as Ridge Regression and Extra Trees require many sampling points in all dimensions around the point of interest, where the spectrum will be predicted. Therefore, they fail to provide an accurate approximation near the edges of the sampling region. Radial basis functions are superior when the region of interest for the prediction approaches the borders of the sampling per-space. However, this is the slowest method among all the others and for the first trial, we recommend the Extra Trees.

Figure 16 shows the resulting fit obtained by the inverse approach. The experimental analyzed data were kindly provided by W. Gawelda from [78]. The red curve is the theoretical spectrum that provides the optimal fit for the experimental one. The theoretical spectrum quantitatively reproduces the energy position and the intensities of maxima A-D in the experimental spectrum. During the fitting procedure, we found several close minima in terms of different combinations of structural parameters. Therefore, we investigate the region near the global minimum in more detail by means of two-dimensional contour maps shown in Figure 17. PyFitIt constructs such contour plots for each possible pair of structural parameters.

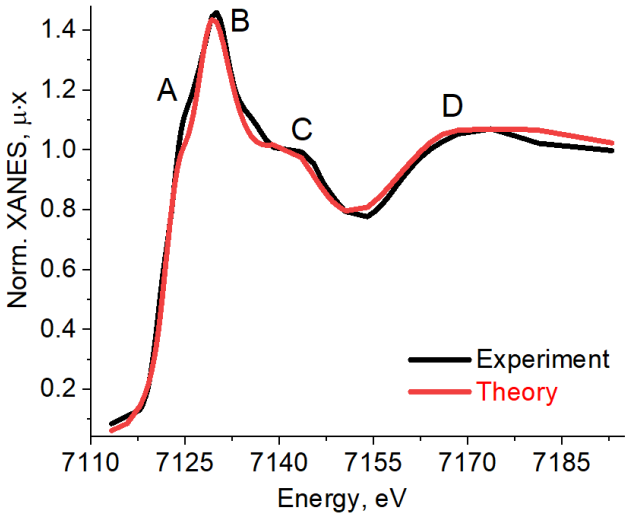

Figure 16: Best fit obtained using the inverse method.

Contour plots for the fit index indicate an obvious correlation between the interatomic distances in the first coordination sphere of iron. As clear from Figure 17(a), a small increment of Fe-N_{axial} distance in one terpy ligand can be compensated by the analogous decrease in the length of Fe-N_{axial} bond in the second terpy ligand. Similar behavior of the fit index is observed for the Fe-N equatorial distances in the two ligands as shown in Figure 17(c), where the minimum valley is located along the line $y = -x + b$. An interesting observation can be made from Figure 17(b). We plot here the contour

plot for the variation of the fit index in the plane of Fe-N_{axial} and Fe-$N_{equatorial}$ distances averaged over the two ligands. Since the equatorial distortion involves two nitrogen atoms, the valley of a minimum of the fit index is located along the line $y = -0.5 \cdot x + b$ (i.e. the shift of one axial nitrogen by 0.1 Å is compensated by a shift of the two equatorial nitrogen atoms of 0.05 Å).

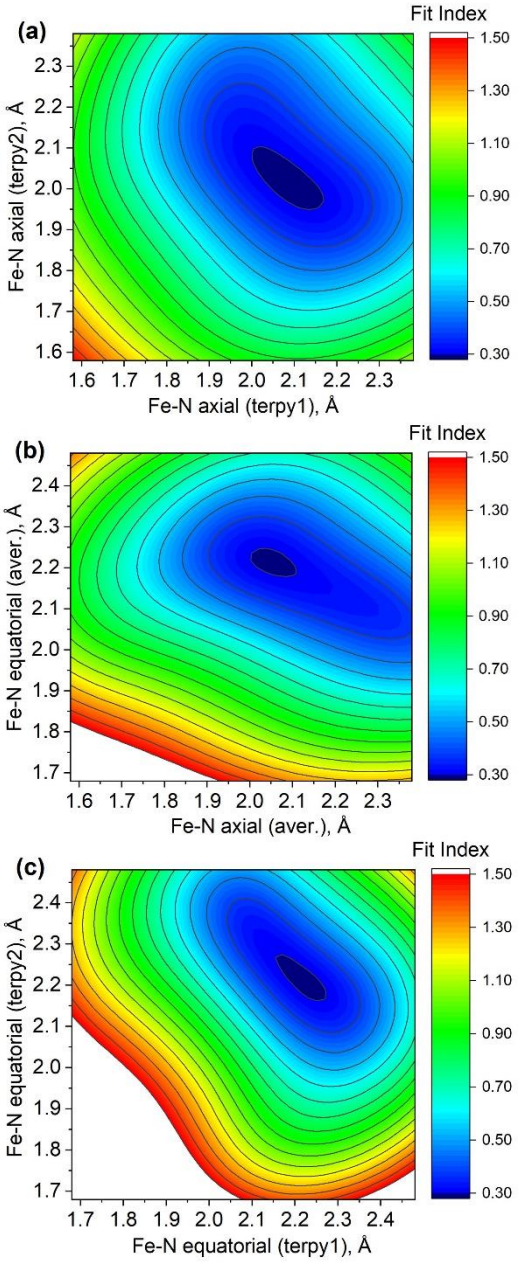

Figure 17. Contour plots for the fit index as a function of two selected structural parameters. The remaining parameters were fixed in the best minimum position (a).

Due to the correlations indicated in Figure 17, we report the averaged Fe-N distances over the two terpy ligands. In Table 1 the values obtained through the fit of the XANES spectrum are compared to the structural parameters obtained in DFT simulations. The latter predicts the difference between Fe-N axial and equatorial distances equal to 0.1 Å in ^5E state and only 0.03 in $^5B_2$ state. Results of our fit agree better with the ^5E model of the excited state, which also emerges from the CASPT2 simulations of Vanko et.al. [78]. However, from a statistical point of view, the ^5E state is difficult to

distinguish from $^5B_2$ state exploiting the XANES fit. In fact, the valley of minima in the multidimensional space of parameters contains many structures with similar fit indices but representing both the ^5E and $^5B_2$ excited spin states (see Figure 5 in S.I. for the comparison between the spectra for these two spin states). The global minimum depends on the choice of the parameters used for the convolution, the energy region for calculating the L_2 norm and the energy shift between the theoretical and experimental spectrum. The effect of these parameters can be partially reduced by using difference XANES, which is also supported by PyFitIt.
A detailed comparison between the fitted structural parameters for difference and normal XANES spectra, the influence of convolutional parameters of theoretical spectra and their energy shift will be the topics for a separate further study using the tools developed and provied by PyFitIt.

Table 1. Structural parameters obtained using the inverse approach. Due to correlations between distances in two terpy ligands (Figure 17 a-c) we show the averaged values for the axial and equatorial parameters. The uncertainty in determining the structural parameters from XANES fit was evaluated by defining a region around the best fit point where changes of XANES spectra, due to variations of structural parameters, are less than error in approximation of XANES as estimated in the cross-validation procedure.

	Fe-N Axial (Å)	**Fe-N Equatorial (Å)**
XANES fit	2.05± 0.02	2.23± 0.02
Exafs fit [78]	2.08 ± 0.02	2.20 ± 0.01
DFT ^5E [78]	2.10	2.20
DFT ^5E [77]	2.12	2.22
DFT $^5B_2$ [78]	2.16	2.19
DFT $^5B_2$ [77]	2.18	2.21

The application of the direct method for $Fe(terpy)_2$ is not straightforward. The reason is sited in the ambiguity associated with structural parameter prediction due to the symmetry of the molecule. The problem is described above in Section 2.3.2. The direct method predicts a set of structural parameters as a function of the input XANES spectrum. The procedure crashes when two different sets of structural parameters correspond to similar spectra. This is the case, for example, of the $Fe\text{-}N_{axial}$ bond lengths in the two symmetric terpy ligands (see Figure 5 for a detailed explanation). Therefore, when all six parameters are predicted independently, the large Root Mean Square (RMS) error makes the direct method comparable to constant prediction. To overcome this problem, we reduced the training set to three deformations shown in Figure 14 but applied simultaneously to all ligands and conserving the initial symmetry of the molecule. In this way, we obtained a $Fe\text{-}N_{axial}$ bond length equal to 2.06 Å and a $Fe\text{-}N_{equatorial}$ distance equal to 2.27 Å. These results, together with their RMS error calculated in the direct method, are presented in Figure 18a. The predicted distances reproduce the expected elongation of the Fe-N bonds in the excited state as well as the asymmetry split for the axial and equatorial distances.

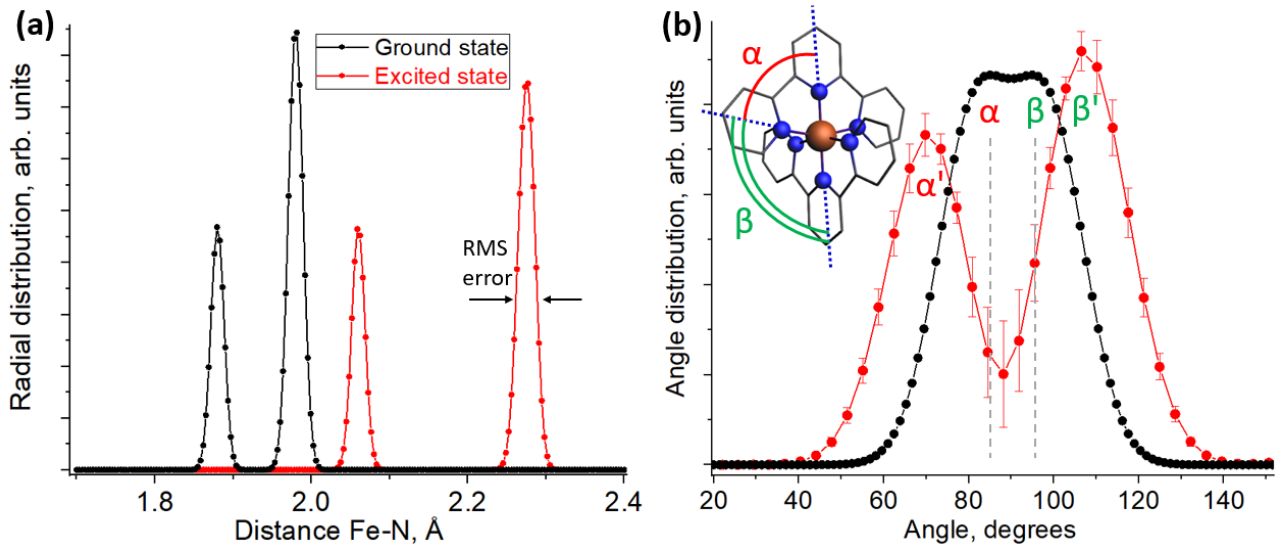

Figure 18: Comparison between radial distribution function for the ground state of Fe(terpy)$_2$ and predicted in direct method (Extra Trees) for the high spin state.

To demonstrate the different possibilities provided by the software, another algorithm was trained to predict the N_i-Fe-N_j angles. For this task, we adopted a modified approach and we predicted the whole function instead of the single values. For every molecule in the training set the angular distribution function (ADF) was calculated and the correspondence ADF → XANES was set up. Thereafter, the ADF can be predicted for the given experimental data. Figure 18(b) shows the initial ADF for the ground state along with the ADF for the high spin state of Fe(terpy)$_2$. Upon elongation of Fe-N axial and equatorial bonds, the FeN_6 polyhedron is distorting from the octahedral configuration. We observed the decrease of the N-Fe-N′ angles when N and N′ belong to the same terpy ligand and the opposite trend when N and N′ belong to different ligands. The broadening of ADF is a parameter in the training algorithm. For smaller FWHM of Gaussians, constituting ADF (each nitrogen atom was modeled with one Gaussian function placed on the corresponding Fe-N distance thus producing smearing of angle) the quantitative interpretation of results becomes clearer, however the error of prediction increases.

Conclusions

We present a new version of the FitIt code, named PyFitIt, for quantitative analysis of XANES spectra. The software is realized in Jupyter Notebooks and utilizes the capabilities of Python language – the most popular tool for machine learning applications. We developed a library of methods able to: 1) perform the evaluation of the experimental set of data (principal component analysis), 2) construc the molecular deformations for a set of selected points constituting a multidimensional space, 3) run the simulations locally or remotely, train the machine learning algorithm on the set of theoretical spectra, 4) fit the experimental spectra or their differences using the inverse or direct approac. The implemented methods, Extra Trees, LightGBM, Ridge Regression, Neural Networks are superior to the polynomial approach used in an older version of FitIt especially for multidimensional problems when 3 and more structural parameters should be fitted within broad ranges of their variation. To reduce the number of required sampling points and to cover the space of parameters uniformly, we implemented the IHS method of sampling of the parameters space. In this work, we show the applicability of the methods to some practical cases. In the first problem, we

showed the PCA analysis and the spectral decomposition procedure performed over a set of experimental data containing two and three independent components (this last example is reported in section 4.2 of the S.I.). While the second problem focused on the structural deformations associated with the spin-crossover complex $Fe(terpy)_2$ upon laser excitation.

Acknowledgement

AVS acknowledges the Russian Foundation for Basic Research (project № 18-02-40029) for the development of machine learning methods and software. AAG acknowledges the Russian Science Foundation (grant No. 17-72-10245) in terms of data analysis in Section 4.1. A. Bugaev acknowledges the support from the President's Grant of Russian Federation for young scientists МК-2554.2019.2. (№ 075-15-2019-1096) for the Pd K-edge data analysis (Section 4.2 in SI).

Bibliography

[1] G. Smolentsev, A.V. Soldatov, FitIt: New software to extract structural information on the basis of XANES fitting, Comput. Mater. Sci., 39 (2007) 569-574.
[2] A. Bianconi, Surface X-ray absorption spectroscopy: Surface EXAFS and surface XANES, Appl. Surf. Sci., 6 (1980) 392-418.
[3] Y. Joly, S. Grenier, Theory of X-Ray Absorption Near Edge Structure, in: J.A. van Bokhoven, C. Lamberti (Eds.) X-Ray Absorption and X-Ray Emission Spectroscopy: Theory and Applications, John Wiley & Sons, Chichester (UK), 2016, pp. 73-97.
[4] A. Bianconi, M. Dell'Ariccia, A. Gargano, C.R. Natoli, Bond Length Determination Using XANES, in: A. Bianconi, L. Incoccia, S. Stipcich (Eds.) EXAFS and Near Edge Structure. Springer Series Chem. Phys., vol 27. , Springer, Berlin, 1983, pp. 57-61.
[5] C.R. Natoli, Distance Dependence of Continuum and Bound State of Excitonic Resonances in X-Ray Absorption Near Edge Structure (XANES), in: K.O. Hodgson, B. Hedman, J.E. Penner-Hahn (Eds.) EXAFS and Near Edge Structure III. Springer Proc. Phys., Vol 2. , Springer, Berlin, 1984, pp. 38-42.
[6] T.E. Westre, P. Kennepohl, J.G. DeWitt, B. Hedman, K.O. Hodgson, E.I. Solomon, A Multiplet Analysis of Fe K-Edge 1s → 3d Pre-Edge Features of Iron Complexes, J. Am. Chem. Soc., 119 (1997) 6297-6314.
[7] L. Mino, G. Agostini, E. Borfecchia, D. Gianolio, A. Piovano, E. Gallo, C. Lamberti, Low-dimensional systems investigated by x-ray absorption spectroscopy: a selection of 2D, 1D and 0D cases, J. Phys. D-Appl. Phys., 46 (2013) 72.
[8] M.A. Soldatov, A. Martini, A.L. Bugaev, I. Pankin, P.V. Medvedev, A.A. Guda, A.M. Aboraia, Y.S. Podkovyrina, A.P. Budnyk, A.A. Soldatov, C. Lamberti, The insights from X-ray absorption spectroscopy into the local atomic structure and chemical bonding of Metal–organic frameworks, Polyhedron, 155 (2018) 232-253.
[9] R.A. Van Nordsthand, The Use of X-Ray K-Absorption Edges in the Study of Catalytically Active Solids, Adv. Catal., 12 (1960) 149-187.
[10] Y. Iwasawa, X-ray absorption fine structure for catalysts and surfaces, Iwasawa, Y. ed., World Scientific, 1996.
[11] D.C. Koningsberger, R. Prins, X-ray absorption: principles, applications, techniques of EXAFS, SEXAFS, and XANES, John Wiley and Sons,New York, NY; None, 1988.
[12] J.A. Van Bokhoven, C. Lamberti, X-ray absorption and X-ray emission spectroscopy: theory and applications, John Wiley & Sons, 2016.
[13] A.L. Ankudinov, B. Ravel, J.J. Rehr, S.D. Conradson, Real-space multiple-scattering calculation and interpretation of x-ray-absorption near-edge structure, Physical Review B, 58 (1998) 7565-7576.

[14] J.J. Rehr, J.J. Kas, M.P. Prange, A.P. Sorini, Y. Takimoto, F. Vila, Ab initio theory and calculations of X-ray spectra, C. R. Phys., 10 (2009) 548-559.
[15] J.J. Rehr, J.J. Kas, F.D. Vila, M.P. Prange, K. Jorissen, Parameter-free calculations of X-ray spectra with FEFF9, Phys. Chem. Chem. Phys., 12 (2010) 5503-5513.
[16] Y. Joly, X-ray absorption near-edge structure calculations beyond the muffin-tin approximation, Physical Review B, 63 (2001) 10.
[17] S.A. Guda, A.A. Guda, M.A. Soldatov, K.A. Lomachenko, A.L. Bugaev, C. Lamberti, W. Gawelda, C. Bressler, G. Smolentsev, A.V. Soldatov, Y. Joly, Optimized Finite Difference Method for the Full-Potential XANES Simulations: Application to Molecular Adsorption Geometries in MOFs and Metal-Ligand Intersystem Crossing Transients, J. Chem. Theory Comput., 11 (2015) 4512-4521.
[18] P. Blaha, K. Schwarz, G.K.H. Madsen, D. Kvasnicka, J. Luitz, R. Laskowski, F. Tran, L.D. Marks, WIEN2K, An Augmented Plane Wave Local Orbitals Program for Calculating Crystal Properties, Karlheinz Schwarz, Techn. Universität Wien, Austria, Wien, Austria, 2018.
[19] G. te Velde, F.M. Bickelhaupt, E.J. Baerends, C.F. Guerra, S.J.A. Van Gisbergen, J.G. Snijders, T. Ziegler, Chemistry with ADF, Journal of Computational Chemistry, 22 (2001) 931-967.
[20] F. Neese, Software update: the ORCA program system, version 4.0, Wiley Interdisciplinary Reviews-Computational Molecular Science, 8 (2018).
[21] C. Gougoussis, M. Calandra, A.P. Seitsonen, F. Mauri, First-principles calculations of x-ray absorption in a scheme based on ultrasoft pseudopotentials: From alpha-quartz to high-T-c compounds, Physical Review B, 80 (2009).
[22] B. Ravel, Quantitative EXAFS Analysis, in: C. Lamberti, J.A. van Bokhoven (Eds.) X- Ray Absorption and X- Ray Emission Spectroscopy: Theory and Applications, John Wiley & Sons, Ltd, 2016.
[23] E. Borfecchia, K. Lomachenko, F. Giordanino, H. Falsig, P. Beato, A. Soldatov, S. Bordiga, C. Lamberti, Revisiting the nature of Cu sites in the activated Cu-SSZ-13 catalyst for SCR reaction, Chem. Sci., 6 (2015) 548-563.
[24] K.A. Lomachenko, E. Borfecchia, C. Negri, G. Berlier, C. Lamberti, P. Beato, H. Falsig, S. Bordiga, The Cu-CHA deNOx Catalyst in Action: Temperature-Dependent NH3-Assisted Selective Catalytic Reduction Monitored by Operando XAS and XES, J. Am. Chem. Soc., 138 (2016) 12025-12028.
[25] L. Braglia, E. Borfecchia, A. Martini, A.L. Bugaev, A.V. Soldatov, S. Oien-Odegaard, B.T. Lonstad-Bleken, U. Olsbye, K.P. Lillerud, K.A. Lomachenko, G. Agostini, M. Manzoli, C. Lamberti, The duality of UiO-67-Pt MOFs: connecting treatment conditions and encapsulated Pt species by operando XAS, Phys. Chem. Chem. Phys., 19 (2017) 27489-27507.
[26] 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 (2017) 14961–14975.
[27] K. Hatada, F. Iesari, L. Properzi, M. Minicucci, A. Di Cicco, Iop, New Graphical User Interface for EXAFS analysis with the GNXAS suite of programs, in: 16th International Conference on X-Ray Absorption Fine Structure, 2016.
[28] K.V. Klementev, Deconvolution problems in x-ray absorption fine structure spectroscopy, J. Phys. D-Appl. Phys., 34 (2001) 2241-2247.
[29] B. Ravel, M. Newville, ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT, J. Synchrot. Radiat., 12 (2005) 537-541.
[30] M. Benfatto, A. Congiu-Castellano, A. Daniele, S.D. Longa, MXAN: a new software procedure to perform geometrical fitting of experimental XANES spectra, J. Synchrot. Radiat., 8 (2001) 267-269.

[31] M. Benfatto, S. Della Longa, C.R. Natoli, The MXAN procedure: a new method for analysing the XANES spectra of metalloproteins to obtain structural quantitative information, J. Synchrot. Radiat., 10 (2003) 51-57.
[32] C.R. Natoli, M. Benfatto, S. Doniach, Use of general potentials in multiple-scattering theory, Physical Review A, 34 (1986) 4682-4694.
[33] A. Bianconi, J. Garcia, M. Benfatto, A. Marcelli, C.R. Natoli, M.F. Ruiz-Lopez, Multielectron excitations in the K-edge x-ray-absorption near-edge spectra of V, Cr, and Mn $3d^{10}$ compounds with tetrahedral coordination, Physical Review B, 43 (1991) 6885-6892.
[34] T.A. Tyson, K.O. Hodgson, C.R. Natoli, M. Benfatto, General multiple-scattering scheme for the computation and interpretation of x-ray-absorption fine structure in atomic clusters with applications to SF_6, $GeCl_4$ and Br_2 molecules, Physical Review B, 46 (1992) 5997-6019.
[35] F. James, M. Roos, MINUIT: a system for function minimization and analysis of the parameter errors and corrections, Comput. Phys. Commun., 10 (1975) 343-367.
[36] K. Hayakawa, K. Hatada, S. Della Longa, P. D'Angelo, M. Benfatto, Progresses in the MXAN fitting procedure, in: B. Hedman, P. Painetta (Eds.) X-Ray Absorption Fine Structure-Xafs13, Amer Inst Physics, Melville, 2007, pp. 111-113.
[37] G. Smolentsev, A. Soldatov, Quantitative local structure refinement from XANES: multi-dimensional interpolation approach, J. Synchrot. Radiat., 13 (2006) 19-29.
[38] 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, Catalysis Today, (2018).
[39] C. Zheng, K. Mathew, C. Chen, Y.M. Chen, H.M. Tang, A. Dozier, J.J. Kas, F.D. Vila, J.J. Rehr, L.F.J. Piper, K.A. Persson, S.P. Ong, Automated generation and ensemble-learned matching of X-ray absorption spectra, npj Comput. Mater., 4 (2018) 9.
[40] J. Timoshenko, D.Y. Lu, Y.W. Lin, A.I. Frenkel, Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles, J. Phys. Chem. Lett., 8 (2017) 5091-5098.
[41] J. Timoshenko, A. Anspoks, A. Cintins, A. Kuzmin, J. Purans, A.I. Frenkel, Neural Network Approach for Characterizing Structural Transformations by X-Ray Absorption Fine Structure Spectroscopy, Physical Review Letters, 120 (2018) 225502.
[42] J. Timoshenko, C.J. Wrasman, M. Luneau, T. Shirman, M. Cargnello, S.R. Bare, J. Aizenberg, C.M. Friend, A.I. Frenkel, Probing Atomic Distributions in Mono- and Bimetallic Nanoparticles by Supervised Machine Learning, Nano Lett., (2019) doi: 10.1021/acs.nanolett.1028b04461.
[43] A.I. Frenkel, O. Kleifeld, S.R. Wasserman, I. Sagi, Phase speciation by extended x-ray absorption fine structure spectroscopy, J. Chem. Phys., 116 (2002) 9449-9456.
[44] A. Piovano, G. Agostini, A.I. Frenkel, T. Bertier, C. Prestipino, M. Ceretti, W. Paulus, C. Lamberti, Time Resolved in Situ XAFS Study of the Electrochemical Oxygen Intercalation in SrFeO2.5 Brownmillerite Structure: Comparison with the Homologous SrCoO2.5 System, J. Phys. Chem. C, 115 (2011) 1311-1322.
[45] M. Fernandezgarcia, C.M. Alvarez, G.L. Haller, XANES-TPR study of Cu-Pd-bimetallic catalysts - Application of factor analysis, J. Phys. Chem., 99 (1995) 12565-12569.
[46] E.R. Malinowski, Factor analysis in chemistry, Wiley, 2002.
[47] A.L. Pomerantsev, Chemometrics in excel, John Wiley & Sons, 2014.
[48] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: new features and applications, Chemometr. Intell. Lab., 140 (2015) 1-12.
[49] 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., 76 (2005) 101-110.

[50] 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 V2O5 Lithium Battery, Anal. Chem., 82 (2010) 3629-3635.
[51] 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 (2018) 1396-1407.
[52] 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, The Journal of Physical Chemistry C, 121 (2017) 18544-18556.
[53] 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 (2014) 20-27.
[54] 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 (2017) 6836-6851.
[55] 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 (2017) 886-895.
[56] 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 (2013) 2855-2867.
[57] 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 (2014) 114-122.
[58] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem., 63 (1991) 1425-1432.
[59] M. Maeder, Evolving factor analysis for the resolution of overlapping chromatographic peaks, Anal. Chem., 59 (1987) 527-530.
[60] C. Ruckebusch, Resolving Spectral Mixtures: With Applications from Ultrafast Time-Resolved Spectroscopy to Super-Resolution Imaging, Elsevier, 2016.
[61] M. Vosough, C. Mason, R. Tauler, M. Jalali-Heravi, M. Maeder, On rotational ambiguity in model-free analyses of multivariate data, J. Chemometr., 20 (2006) 302-310.
[62] A. de Juan, S.C. Rutan, R. Tauler, Two-Way Data Analysis: Multivariate Curve Resolution Iterative Resolution Methods, Elsevier Science Bv, Amsterdam, 2009.
[63] A. de Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures - Revisiting latest trends in multivariate resolution, Anal. Chim. Acta, 500 (2003) 195-210.
[64] A. de Juan, R. Tauler, Multivariate curve resolution (MCR) from 2000: Progress in concepts and applications, Critical Reviews in Analytical Chemistry, 36 (2006) 163-176.
[65] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: a review of advanced and tailored applications and challenges, Anal. Chim. Acta, 765 (2013) 28-36.
[66] 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 (2009).
[67] G. Piatetsky, Python eats away at R: Top Software for Analytics, Data Science, Machine Learning in 2018: Trends and Analysis, in, 2018.
[68] A. Ben-Israel, T. N. E. Greville, Generalized Inverses. Theory and Applications, 2nd ed., Springer, New York, NY, 2003.
[69] R.G. Brereton, Chemometrics: data analysis for the laboratory and chemical plant, John Wiley & Sons, 2003.

[70] A. Manceau, M. Marcus, T. Lenoir, Estimating the number of pure chemical components in a mixture by X-ray absorption spectroscopy, J. Synchrot. Radiat., 21 (2014) 1140-1147.
[71] G.E. Fasshauer, Meshfree Approximation Methods with Matlab, World Scientific, 2007.
[72] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Machine Learning, 63 (2006) 3-42.
[73] 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/CeO2 catalyst, J. Synchrot. Radiat., 25 (2018) 989-997.
[74] R. Kopelent, J.A. van Bokhoven, J. Szlachetko, J. Edebeli, C. Paun, M. Nachtegaal, O.V. Safonova, Catalytically Active and Spectator Ce^{3+} in Ceria-Supported Metal Catalysts, Angewandte Chemie-International Edition, 54 (2015) 8728-8731.
[75] J.A. Real, A.B. Gaspar, M.C. Munoz, Thermal, pressure and light switchable spin-crossover materials, Dalton Transactions, (2005) 2062-2079.
[76] F. Renz, H. Oshio, V. Ksenofontov, M. Waldeck, H. Spiering, P. Gutlich, Strong field iron(II) complex converted by light into a long-lived high-spin state, Angewandte Chemie-International Edition, 39 (2000) 3699-3700.
[77] S.E. Canton, X.Y. Zhang, L.M.L. Daku, A.L. Smeigh, J.X. Zhang, Y.Z. Liu, C.J. Wallentin, K. Attenkofer, G. Jennings, C.A. Kurtz, D. Gosztola, K. Warnmark, A. Hauser, V. Sundstrom, Probing the Anisotropic Distortion of Photoexcited Spin Crossover Complexes with Picosecond X-ray Absorption Spectroscopy, J. Phys. Chem. C, 118 (2014) 4536-4545.
[78] G. Vanko, A. Bordage, M. Papai, K. Haldrup, P. Glatzel, A.M. March, G. Doumy, A. Britz, A. Galler, T. Assefa, D. Cabaret, A. Juhin, T.B. van Driel, K.S. Kjaer, A. Dohn, K.B. Moller, H.T. Lemke, E. Gallo, M. Rovezzi, Z. Nemeth, E. Rozsalyi, T. Rozgonyi, J. Uhlig, V. Sundstrom, M.M. Nielsen, L. Young, S.H. Southworth, C. Bressler, W. Gawelda, Detailed Characterization of a Nanosecond-Lived Excited State: X-ray and Theoretical Investigation of the Quintet State in Photoexcited Fe(terpy)(2) (2+), J. Phys. Chem. C, 119 (2015) 5888-5902.