A spectrum deconvolution method based on grey relational analysis

Rui Mu [a], Yujie Zheng [a], Andreas Lambertz [b], Regan G. Wilks [c,d], Marcus Bär [c,d,e,f], Yufeng Zhang [g,*]

[a] School of Aerospace Engineering, Xiamen University, Xiamen 361102, China
[b] Institut für Energie- und Klimaforschung (IEK5-Photovoltaik), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
[c] Department Interface Design, Helmholtz-Zentrum Berlin für Materialien und Energie GmbH (HZB), Albert-Einstein-Str. 15, 12489 Berlin, Germany
[d] Energy Materials In-Situ Laboratory Berlin (EMIL), HZB, Albert-Einstein-Str. 15, 12489 Berlin, Germany
[e] Department of Chemistry and Pharmacy, Friedrich-Alexander-Universität Erlangen-Nürnberg, Egerlandstr. 3, 91058 Erlangen, Germany
[f] Helmholtz Institute Erlangen-Nürnberg für Renewable Energy (HI ERN), Albert-Einstein-Str. 15, 12489 Berlin, Germany
[g] College of Physical Science and Technology, Xiamen University, Xiamen 361005, China

ARTICLE INFO

ABSTRACT

The extensive usage of X-ray spectroscopies in studying complex material systems is not only intended to reveal underlying mechanisms that govern physical phenomena, but also used in applied studies focused on an insight-driven performance improvement of a wide range of devices. However, the traditional analysis methods for X-ray spectroscopic data are rather time-consuming and sensitive to errors in data pre-processing (e.g., normalization or background subtraction). In this study, a method based on grey relational analysis, a multi-variable statistical method, is proposed to analyze and extract information from X-ray spectroscopic data. As a showcase, the valence bands of microcrystalline silicon suboxides probed by hard X-ray photoelectron spectroscopy (HAXPES) were investigated. The results obtained by the proposed method agree well with conventionally derived composition information (e.g., curve fit of Si 2p core level of the silicon suboxides). Furthermore, the uncertainty of chemical compositions derived by the proposed method is smaller than that of traditional analysis methods (e. g., the least square fit), when artificial linear functions are introduced to simulate the errors in data pre-processing. This suggests that the proposed method is capable of providing more reliable and accurate results, especially for data containing significant noise contributions or that is subject to inconsistent data pre-processing. Since the proposed method is less experience-driven and error-prone, it offers a novel approach for automate data analysis, which is of great interest for various applications, such as studying combinatorial material "libraries".

Introduction

Modern optoelectronic devices, such as light emitting diodes and photovoltaic cells, consist of multiple functional materials (e.g., absorber, back contact, front contact, and buffer layer) with different optoelectronic properties [1]. The performance of the devices depends not only on the characteristics of the functional materials, but also is crucially determined by the nature and quality of the interfaces between the materials [2], because a variety of unique phenomena (e.g., multi-ferroicity, spin-Hall effect, superconductivity) arise from symmetry breaking at interfaces due to strong coupling of charge, spin, and orbit [3]. Also, interfaces critically influence charge transfer and separation in e.g., photovoltaics. With an ever-increasing demand for better device

performance, the tendency toward using complex material systems (e.g., ternary cathodes for lithium-ion batteries [4], multinary perovskite compound – based solar cells[5]) in devices posts a greater challenge – and opportunity – for device optimization. Hence, it is critical to fully understand the chemical and electronic structures of those complex materials and interfaces, in order to optimize the devices and shed light on the underlying physical mechanisms.

As an example, hydrogenated microcrystalline silicon oxide ($\mu c\text{-}SiO_x$:H) has attracted great attention as a functional material in photovoltaic devices, due to its unique properties (e.g., wide band gap, adaptable refractive index, and high conductivity) [6]. By utilizing a plasma to provide energy for the chemical reaction to take place during deposition, the processing temperature can be greatly reduced in plasma enhanced

---

* Corresponding author.
  E-mail address: yufengzhang@xmu.edu.cn (Y. Zhang).

chemical vapor deposition (PECVD), which makes PECVD suitable to prepare thin films on heat-sensitive materials or when thermal cycles are of great concerns [7,8]. Therefore, PECVD has been widely used in preparing silicon oxides and nitrides [9,10]. However, it was reported that the optoelectrical properties of the μc-SiO$_x$:H prepared by PECVD is greatly influenced by its chemical compositions [6].

To establish correlation between the chemical compositions (and electronic structures) and the physical properties (e.g., optoelectrical properties) for materials, one of the important characterization tools is X-ray spectroscopy (e.g. X-ray photoelectron spectroscopy, X-ray absorption spectroscopy, X-ray emission spectroscopy) [11,12]. However, X-ray spectroscopic data incorporate a wide range of information, which originates from not only the materials or interfaces of interest, but also other relevant or irrelevant substances (e.g., surface adsorbates, buried layers, etc.). The traditional spectrum analysis methods, such as least square fit (LSF) or curve fit, focus on deriving a best linear combination of reference data or specific functions (e.g., Gaussian or Voigt profiles) for spectrum deconvolution, respectively. These methods are usually time-consuming and possess many limitations, such as subjectivity in setting the number of functions and corresponding parameters. These limitations render them inefficient at handling the spectroscopic data of large sample sets, which are increasingly common, due in part to the increasing complexity of materials (e.g., binary, ternary, multinary compounds) but also to increased automation in measurement processes and the study of combinatorial material "libraries."

Furthermore, analysis uncertainties can originate from noise and artifacts in the data induced by fluctuation of photon flux, detector non-linearity, different degrees of sample contamination and sample misalignment. Generally, data pre-processing techniques, such as baseline correction and noise suppression, are often employed before any further analysis. These processes are commonly experience-driven and thus prone to personal criteria or inconsistence in data evaluation; even though certain protocols have been developed to remediate the problem [13]. As an example, in X-ray photoelectron spectroscopy, there are at least four different approaches to account for the background: linear, polynomial, Shirley, or Tougaard – based backgrounds. However, on the question of which one should be applied and how to do so appropriately in the practical application is not universally agreed upon [14,15]. Recently, it has been reported that the oxidation states of materials can be identified by studying X-ray absorption near-edge absorption spectroscopy data with supervised machine learning techniques, which reduce interference of human judgment [16]. However, insufficient training data might limit the extent to which the learning algorithm can capture the complete data structure, and it may miss weak features. Hence, it is an important task to develop and evaluate methods to extract information from the X-ray spectroscopic data, even under the influence of noise and/or deviation in the data pre-processing as well as limited knowledge of the composition of such materials or interfaces (due to diffusion, defects, etc.).

The grey relational analysis (GRA) is a multi-variable statistical method capable of revealing correlations between multiple objects with limited information [17,18], which is successfully applied in a variety of fields, such as optimization of process parameters for electrical discharge machining or laser beam drilling [19–21], failure mode analysis [22], state-of-health estimation [23], and pattern recognition [24]. In general, GRA identifies the relationship between comparison and reference sequences by studying their geometry proximity with a GRA algorithm (e.g., ranking of the grey relational grade) [17], which can be easily implanted in a program for streamlining data analysis. Hence, a quantitative method based on GRA would be useful to extract information from X-ray spectroscopic data, which could also provide automated decision making in data acquisition (e.g., setting measurement time or designing measurement parameters for a combinatorial materials library).

To sum up, X-ray spectroscopies are widely employed to investigate the electronic and chemical structures of materials (e.g., μc-SiO$_x$:H),

which are important for understanding the underlying physics and device optimization. However, the traditional methods (e.g., LSF) for analyzing X-ray spectroscopic data are usually time-consuming and sensitive to noise or inconsistence in data pre-processing. Meanwhile, by comparing sequences based on their geometric proximity and similarity, GRA has been successfully applied in various fields to support multi-criteria decision making. Therefore, the objective of this study is to develop a novel analysis method based on GRA to extract critical information (e.g., chemical compositions) from X-ray spectroscopic data. A set of hard X-ray photoelectron spectroscopy (HAXPES) data in particular of the valence band region of μc-SiO$_x$:H is used as a showcase example to evaluate the effectiveness of the proposed grey relational method (GRM). It is found that the results derived by GRM are more accurate than those derived by LSF, especially when the spectra are noisy or include a complex background.

Material preparation

p-type hydrogenated microcrystalline silicon suboxides (μc-SiO$_x$:H) layers with different oxygen contents ([O] = O/(Si + O)) were deposited by RF (13.56 MHz) PECVD technique at 185 °C substrate temperature and a power density of 300 mWcm^{-2}, using a mixture of silane (SiH$_4$), carbon dioxide (CO$_2$), hydrogen (H$_2$), and trimethylboron (TMB) diluted in helium gas. The thickness of the investigated layers was in the range between 300 and 600 nm, which significantly extends beyond a typical range of the structurally inhomogeneous development of the film during initial growth of doped silicon oxide [6]. [O] of the studied samples was 0, 0.17, 0.25, 0.33, and 0.49, respectively – as derived by Rutherford backscattering spectroscopy (RBS, see Ref. [6] for details).

Characterization methods and material properties

The μc-SiO$_x$:H samples were characterized by HAXPES with 2003 eV photons at the HiKE endstation on the KMC-1 beamline of the BESSY-II electron storage ring [6] focusing on the Si 2p and valence band region. A pass energy of 100 eV was used for the measurements, resulting in a total experimental resolution of approx. 0.25 eV. The binding energy scale was calibrated using a clean Au foil, setting the binding energy of the Au 4f$_{7/2}$ line to 84.00 eV or the center of the Fermi edge to 0 eV. The curve fit analysis of the Si 2p HAXPES spectra suggests that the Si^0 oxidation state fraction is 0.81, 0.68, 0.56, 0.47, and 0.23, respectively [6], with increasing CO$_2$ concentration during sample preparation. The detailed fit of the Si 2p additionally reveals the fractions of Si^{1+}, Si^{2+}, Si^{3+}, and Si^{4+} (see Table 1 and Ref. [6] for details), with mainly the Si^{4+} contribution increasing at the expense of the Si^0 fraction. Assuming that the formation of corresponding SiO$_x$ suboxides determine the Si oxidation states, one can compute the oxygen content as [O]$^{\text{HAXPES}}$ = O/(Si + O)) = $(0 \times \text{Si}^0 + 0.5 \times \text{Si}^{1+} + 1 \times \text{Si}^{2+} + 1.5 \times \text{Si}^{3+} + 2 \times \text{Si}^{4+})/[(\text{Si}^0 + \text{Si}^{1+} + \text{Si}^{2+} + \text{Si}^{3+} + \text{Si}^{4+}) + (0 \times \text{Si}^0 + 0.5 \times \text{Si}^{1+} + 1 \times \text{Si}^{2+} + 1.5 \times \text{Si}^{3+} + 2 \times \text{Si}^{4+})]$. The computed values are shown in Table 1. The values somewhat deviate from the RBS derived [O] ([O]$^{\text{RBS}}$), which can be

Table 1

Oxygen content ([O] = O/(Si + O)) of the studied μc-SiO$_x$:H samples derived by RBS [6], and the fractions of the different Si oxidation states calculated based on the curve fit analysis of the Si 2p HAXPES data (see Ref. [6] for details). The HAXPES-derived oxygen content [O]$^{\text{HAXPES}}$ was computed based on the Si oxidation states assuming the formation of respective SiO$_x$ (sub)oxides.

RBS [O]$^{\text{RBS}}$	HAXPES					
	Si^0/ Si$^{\text{total}}$	Si^{1+}/ Si$^{\text{total}}$	Si^{2+}/ Si$^{\text{total}}$	Si^{3+}/ Si$^{\text{total}}$	Si^{4+}/ Si$^{\text{total}}$	[O]$^{\text{HAXPES}}$
0	0.81	0.03	0	0.01	0.14	**0.24**
0.17	0.68	0.06	0	0.08	0.18	**0.34**
0.25	0.56	0.06	0	0.07	0.30	**0.43**
0.33	0.47	0.07	0	0.09	0.35	**0.47**
0.49	0.23	0.07	0.10	0.10	0.49	**0.56**

explained by the different information depth of RBS and HAXPES. While RBS is bulk sensitive, HAXPES probes the near-surface region of the sample, which is significantly affected by surface contamination and oxidation, rationalizing $[O]^{HAXPES} > [O]^{RBS}$.

Theoretical model

Generally, the ranking of grey relational grades (GRG) plays a key role in revealing the correlation between comparison and reference sequences in the GRA. However, when applied to spectroscopic data, this ranking cannot directly provide information such as chemical composition [17], as is elaborated upon in the Supplementary document S1. Hence, a new algorithm was developed to calculate GRG for analyzing spectroscopic data of samples that consists of multiple chemical species. If Z_0 is defined as comparison sequence (i.e., the spectrum of the studied sample), and Z_1, Z_2, …, Z_n are corresponding reference sequences (i.e., the spectra of standard samples), which consist of N data points, the GRG between the comparison and the ith reference sequence is

$$\gamma_{0i} = \frac{\frac{1}{N}\sum_{j=1}^{N}\frac{1}{[Z_0(j)-Z_i(j)]^2+1}}{\frac{1}{\left[\overline{Z_0}-\overline{Z_i}\right]^2+1}} \quad (1)$$

where $\overline{Z}_0$, $\overline{Z}_i$ is the average of comparison and reference sequences. If the differences between two sequences are small ($\ll 1$), then a Taylor expansion is used to simplify Eq. (1) to

$$\gamma_{0i} = 1 + \overline{(Z_0 - Z_i)}^2 - \overline{(Z_0 - Z_i)^2} \quad (2)$$

For a comparison sequence, which is a linear combination of the reference sequences (e.g., a sample consisting of n species), the difference between corresponding GRGs yield a linear relationship with their chemical composition, which is elucidated in the Supplementary document S2.

$$\begin{aligned}
\gamma_{01} - \gamma_{02} &= a_{11}k_1 + a_{12}k_2 + \cdots + a_{1(n-1)}k_{n-1} + a_{1n} \\
\gamma_{01} - \gamma_{03} &= a_{21}k_1 + a_{22}k_2 + \cdots + a_{2(n-1)}k_{n-1} + a_{2n} \\
&\vdots \\
\gamma_{01} - \gamma_{0(n-1)} &= a_{(n-1)1}k_1 + a_{(n-1)2}k_2 + \cdots + a_{(n-1)(n-1)}k_{n-1} + a_{(n-1)n}
\end{aligned} \quad (3)$$

where k_1, k_2, … k_n are the chemical composition of respective species, and sum of them equals to one. This suggests that it is possible to derive the chemical composition of samples with unknown chemical composition by solving the linear equations, deriving the coefficients (e.g., a_{11}, a_{12}).

Results and discussion

In order to validate the proposed method, a set of simulation data sequences was assembled by a linear combination of reference spectra

$$Z_0 = k_1 Z_1 + k_2 Z_2 \quad k_1 + k_2 = 1 \quad (4)$$

where Z_1 and Z_2 are reference data. In this case, they are the valence band HAXPES data of $\mu c\text{-}SiO_x\text{:}H$ with $[O]^{RBS} = 0$ and 0.49, respectively. It is worth noting that, unlike core levels, the valence band is normally hard to investigate using traditional analysis methods (e.g., curve fit), because the features generally are broad and convoluted due to involvement of more delocalized valence electrons. The k_1 and k_2 indicate the chemical compositions, changing from 0 to 1 in 0.1 steps. In this scenario, the differences between GRGs satisfy the following equations

$$\gamma_{01} - \gamma_{02} = a_{11}k_1 + a_{12} \quad (5)$$

Hence, with known chemical compositions (i.e., values of k_1 and k_2 from the simulation data), the corresponding coefficients (i.e., a_{11}, a_{12}) are determined, as shown in the Supplementary document S3. A linear fit of the differences of GRGs and known chemical compositions was

performed to further improve accuracy of the coefficients.

Then, a set of test data sequences is generated using different chemical compositions

$$Z_0' = k_1' Z_1 + k_2' Z_2 \quad k_1' + k_2' = 1 \quad (6)$$

The comparison between chemical compositions calculated by GRM and the set values (i.e., the values used to generate the test data) suggests that the deviation (δ, absolute difference between the set and calculated values) induced by the proposed model is less than 6.3×10^{-5}, as shown in Supplementary document S4. This indicates that the GRM possesses sufficient accuracy under ideal situation (e.g., without interfering impact from noise or data pre-processing procedures).

However, the spectra of standard samples generally possess a better signal-to-noise ratio than 'real-world' experiments. Therefore, Poisson noise (P), generated by the built-in function in OriginLab, was introduced into the test data (resulting in the data set T1, T2, T3, T4, and T5, see Fig. 1) to simulate the impact of noise on data analysis.

$$ZP = Z_0' + P \quad (7)$$

Poisson noise is chosen because this type of noise is commonly present in spectra recorded by detectors based on charge-coupled devices (CCD), which are widely used in X-ray spectroscopy measurement setups. As discussed in the following paragraphs, the type of noise does not introduce much difference on the analysis. The strength of noise in this study is defined as the ratio between the amplitude of noise and intensity of spectral signal (i.e., difference between maximum and minimum of the spectroscopic data). In the following discussion, the strength of added Poisson noise is about 10% of the intensity of a spectrum.

The test data as well as the reference spectra and a typical Poisson noise are shown in Fig. 1. The chemical compositions derived from the test data calculated by GRM and via LSF are listed in Table 2. The
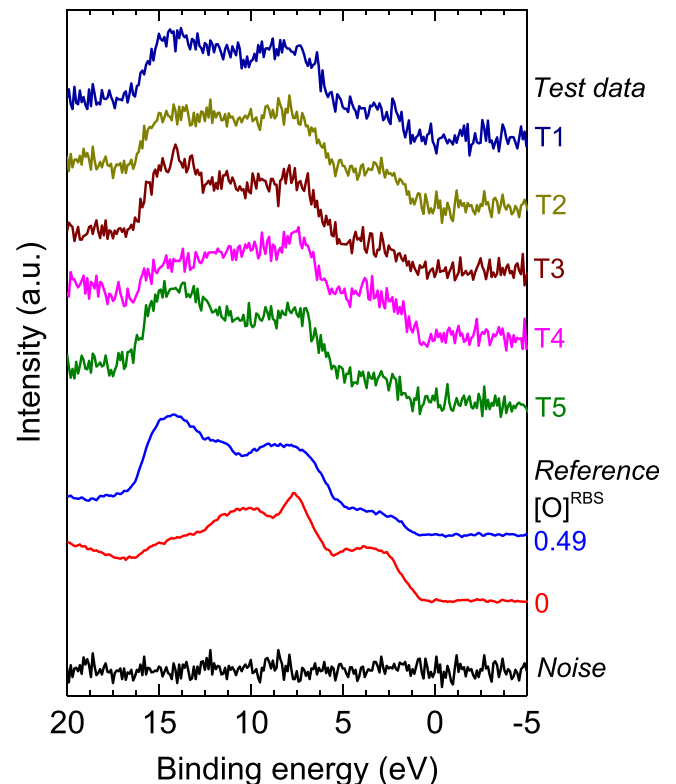
Fig. 1. HAXPES valence band spectra of reference $\mu c\text{-}SiO_x\text{:}H$ samples with $[O]^{RBS} = 0$ and 0.49, as well as the test data T1, T2, T3, T4, and T5. A Poisson noise is shown as a black line.

Table 2

Set chemical composition parameters (k'_1, i.e. fraction of SiO_x with $[O]^{RBS} = 0$, and k'_2, i.e. fraction of SiO_x with $[O]^{RBS} = 0.49$) for the generated test data T1–T5 and calculated composition derived by GRM compared to that obtained by LSF.

Test data			GRM		LSF	
	Set k'_1	Set k'_2	Calc. k'_1	δ	Calc. k'_1	δ
T1	0.03	0.97	0.05	2.4%	0.05	2.3%
T2	0.76	0.24	0.76	0.4%	0.77	0.5%
T3	0.14	0.86	0.15	1.5%	0.15	0.8%
T4	0.44	0.56	0.45	1.0%	0.44	0.4%
T5	0.24	0.76	0.26	1.8%	0.25	0.8%

deviation (δ) between the true and calculated value using GRM increases (up to 2.4%) after introducing the noise; a similar effect is observed for the LSF approach (up to 2.3%). Hence, it is clear that the proposed method provides results as reliable as the traditional method, even when the data incorporates significant noise. Note that test data T4 is used below to study additional impacts on data analysis.

The presence of noise is not the only effect that might introduce uncertainty in the data analysis. Most spectra will undergo data pre-processing, such as baseline subtraction, before any further data analysis. Such processes rely on a manually (i.e., picked by researcher) or automatically (i.e. calculated by a pre-defined algorithm) assigned baseline. Certainly, introducing an automatic mechanism in the process could lead to better reproducibility – if not necessarily accuracy – in data analysis. However, some deviation is still unavoidable due to various reasons, such as e.g., changes in sample alignment, fluctuation of the excitation source, or the presence of scattering sample surface.

To study the impact of such deviations, a linear background was added into the test data. Firstly, a constant value (C) is introduced to a test data:

$$ZC = ZP + C \qquad (8)$$

The original test data T4 before and after adding a constant background (i.e., 10%, 20%, 40%, and 80% of the maximum intensity of the test data) is shown in Fig. 2. The chemical composition predicted by GRM and LSF is given in the Table 3. The GRM is evidently not sensitive to the added baseline-induced uncertainty at all, because the deviation merely changes from 1.0% for no background addition to 1.2% for an addition of 80% of maximum test data intensity. However, adding different constant backgrounds of different intensity to the data has a dramatic impact on the LSF-derived composition and related uncertainty, which is in the worst case (i.e., the largest background added) more than two orders of magnitude larger than that using GRM. In this simple case, the increased error in the LSF procedure is of course easily remedied (as shown in Supplementary document S5) by including a variable constant background in the fitting procedure (as would generally be the case in real analytical setting). Nevertheless, the comparison demonstrates that the GRM procedure is less sensitive to deviation in selecting background function.

To better illustrate this point, a linear function was used as a background (i.e., sloping baseline):

$$ZL(j) = ZP(j) + a + b \times j \quad j = 1, 2, \cdots N \qquad (9)$$

where a and b are the interception and slope of the linear function, respectively. The original test data T4 is combined with a linear baseline with different slopes (b = 0.0005, 0.0010, 0.0020, and 0.0040), shown in Fig. 3. The error of the chemical compositions predicted by GRM, shown in the Table 4, increases with increasing slope (e.g., for b = 0.0040 the error is around 39%), but is always lower than that of LSF. Note that, in this LSF procedure, a constant is also included, as discussed in Supplementary document S5, to simulate the real analytical practice. Hence, in contrast to adding a constant background (which has no effect on GRM), here the GRM results are significantly influenced. It is easy to
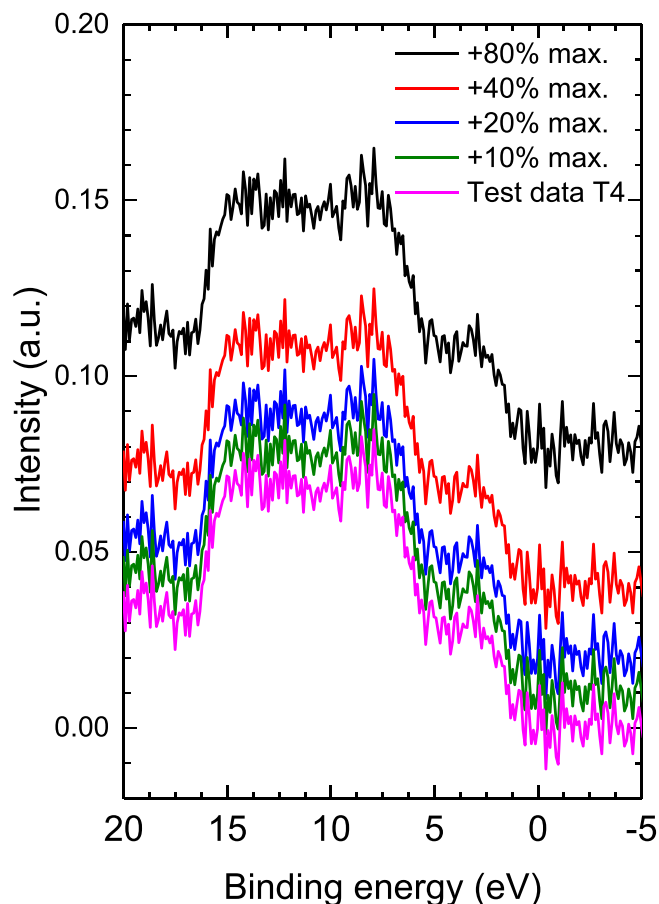
Fig. 2. Impact of the addition of constant backgrounds to the spectral shape of the calculated HAXPES valence band spectra (test data T4). The added background is a constant, being proportional (e.g., 10%) to the maximum intensity of test data T4.

Table 3

Chemical composition derived by GRM for test data T4 to which different constant baselines have been added, comparing with the results obtained by LSF. The added constant is proportional (e.g., 10%) to the maximum intensity of test data T4.

Test data	Set k'_1	GRM		LSF	
		Calc. k'_1	δ	Calc. k'_1	δ
T4	0.44	0.45	1.0%	0.44	0.4%
+10% max	0.44	0.45	1.0%	0.65	21%
+20% max	0.44	0.45	1.0%	0.85	41%
+40% max	0.44	0.45	1.0%	1.26	82%
+80% max	0.44	0.45	1.2%	2.07	163%

notice miscalculation of the baseline when the slope is larger than 0.0005. In this circumstance, the error induced by the GRM is 4.1%, while the LSF leads to an error of 8.1%.
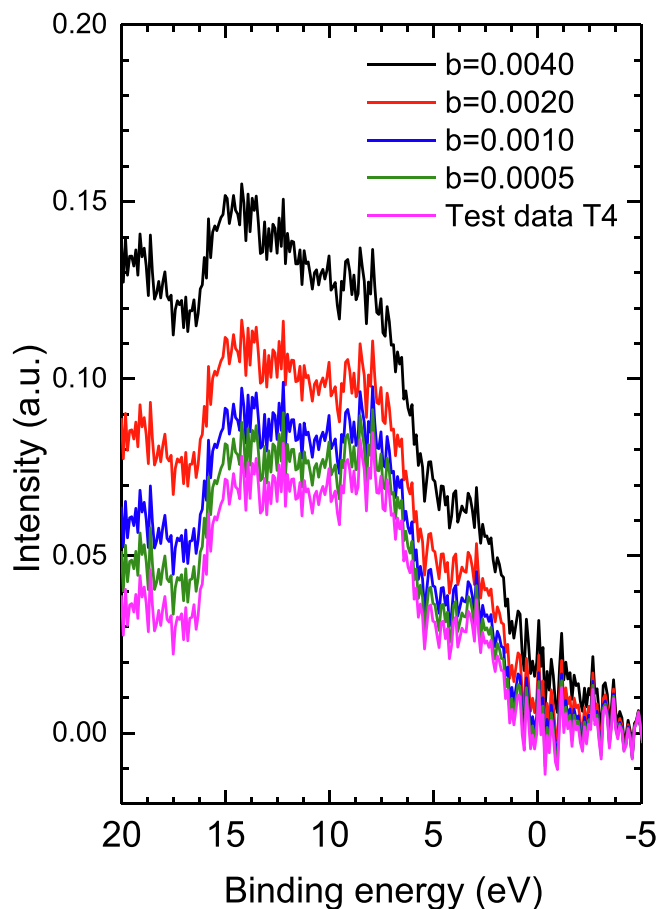
Evidently, if the data sequence Z incorporates a function L, the change of GRGs with and without L is

$$\Delta \gamma_{0i} = 2\left[\overline{L}\left(\overline{Z} - \overline{Z_i}\right) - \overline{L(Z - Z_i)}\right] + \left(\overline{L}^2 - \overline{L^2}\right) \qquad (10)$$

which is elaborated in the Supplementary document S6. Hence, the change of difference between two GRGs is

$$\Delta_{mi} = 2\left[\overline{L}\left(\overline{Z_i} - \overline{Z_m}\right) - \overline{L(Z_i - Z_m)}\right] \qquad (11)$$

If the function L is a constant, theoretically, the changes of both GRGs

Fig. 3. Impact of the addition of linear backgrounds with different slopes (b) on the spectral shape of the calculated HAXPES valence band spectra (test data T4).

Table 4
Chemical composition derived by GRM for test data T4 to which linear baselines with different slopes (b) have been added, comparing with the results obtained by LSF.

Test data	Set k'_1	GRM		LSF	
		Calc k'_1	δ	Calc. k'_1	δ
T4	0.44	0.45	1.0%	0.46	2.2%
b = 0.0005	0.44	0.40	4.1%	0.52	8.1%
b = 0.0010	0.44	0.35	9.2%	0.58	14%
b = 0.0020	0.44	0.25	19%	0.70	26%
b = 0.0040	0.44	0.05	39%	0.94	50%

and their differences will be zero. This explains why the proposed method yields small and relatively stable error in analyzing data with a shifted baseline. Since the average of a noise signal is close to zero, the GRM is unsurprisingly not sensitive to it (i.e., not correlating to the strength and type of noise). On the other hand, if the function L is not a constant, the changes of GRGs and their differences depend greatly on the function L and the reference sequences. Nevertheless, the proposed GRM method performs better than the LSF in limiting the analysis error, when there is inconsistency in data pre-processing.

Finally, the HAXPES valence band spectra of μc-SiO$_x$:H with different oxygen content, as shown in Fig. 4, are analyzed by GRM. The valence band is composed of contributions from Si and O electrons: feature A (at ~10 eV) is attributed to Si 3s, B (~7 eV) to hybridized Si 3s – O 2p, C (~4 eV) to Si 3p, D (~14 eV) to hybridized O 2p – Si 3p, and E (~9 eV) to hybridized O 2p – Si 3s derived states [25,26]. The relative spectral intensity of the Si-related features A–C in the spectra of the μc-SiO$_x$:H
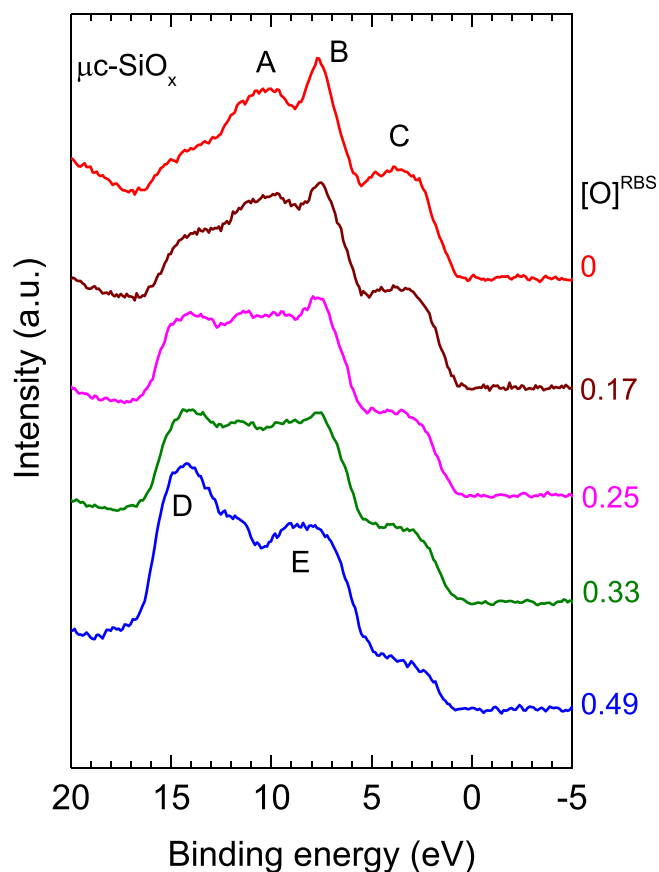
Fig. 4. Experimental HAXPES valence band spectra of μc-SiO$_x$:H samples (with [O]^{RBS} = 0, 0.17, 0.25, 0.33, and 0.49). Note that the spectra are offset vertically for easier comparison.

samples gradually decreases, while the O-related features D and E increase with increasing oxygen content. Even though the spectra of μc-SiO$_x$:H with [O]^{RBS} = 0 and 0.49 are quite distinct, they do not represent pure phases. This is most apparent inspecting the spectra of the Si 2p core level (see Supplementary document S7). The fit analysis reveals that multiple Voigt functions representing different Si oxidation states are required to result in a reasonably description of the Si 2p data. While only 81% of the silicon in the μc-SiO$_x$:H sample with [O]^{RBS} = 0 is present as Si^0, in the μc-SiO$_x$:H sample with [O]^{RBS} = 0.49, 23% of the silicon is present in the form of Si^0 [6] (see also Table 1). The chemical composition parameters (i.e., k'_1 and k'_2) of μc-SiO$_x$:H samples are derived by GRM and LSF, using the HAXPES valence band spectra of μc-SiO$_x$:H samples with [O]^{RBS} = 0 and 0.49 as reference sequences. After taking the composition of reference data (the results of curve fit in Ref. [6]) into account, the Si^0 state fractions of the μc-SiO$_x$:H samples are computed (i.e., 0.81 × k'_1 + 0.23 × k'_2) and shown in Table 5. The results agree with those of curve fit in a reasonable margin of uncertainty.

Table 5
The fraction of Si^0 oxidation state calculated based on the curve fit analysis of the Si 2p HAXPES data (see Ref. [6] for details), as well as the results obtained by GRM and LSF methods, based on the HAXPES valence band spectra.

Curve fit	GRM			LSF		
Si^0	Calc. k'_1	Calc. k'_2	Si^0	Calc. k'_1	Calc. k'_2	Si^0
0.81	1.00	0.00	**0.81**	1.00	0.00	**0.81**
0.68	0.82	0.18	**0.70**	0.82	0.18	**0.71**
0.56	0.57	0.43	**0.56**	0.57	0.43	**0.56**
0.47	0.47	0.53	**0.50**	0.47	0.53	**0.50**
0.23	0.00	1.00	**0.23**	0.00	1.00	**0.23**

Note, curve fitting is a generally time-consuming process, but could yield more information (e.g., different full width at half maximum due to inhomogeneous broadening [6]) from the data analysis.

In the Supplementary document S7, the Si 2p HAXPES spectra of μc-SiO$_x$:H is analyzed using the same method, while the Si 2p HAXPES data of μc-SiO$_x$:H with [O]^{RBS} = 0 and 0.49 are used as references for Si and Si oxides, respectively. There are close linear correlations between the results derived by the GRM, the LSF, the curve fit, and the RBS, as well. For example, the μc-SiO$_x$:H sample with [O]^{RBS} = 0.17 consists of 70% and 67% of Si^0, respectively, determined by GRM using valence band and Si 2p HAXPES data, while the curve fit revealing a 68% of Si^0 oxidation state. This suggests that the proposed method can effectively analyze HAXPES data and reveal the critical information (e.g., chemical composition), without sacrificing much accuracy.

Conclusion

A method based on GRA (i.e., GRM) is developed to analyze the correlation between comparison and reference spectra, and extract critical information (i.e., the chemical composition) from X-ray spectroscopic data. The typical procedures are summarized here: first, GRGs are computed for a set of data with known chemical compositions; second, the differences of the GRGs are used to derive the coefficients (e. g., a$_{11}$, a$_{12}$) for the linear equation, Eq. (3); third, unknown chemical composition is extracted from a spectrum using the linear equations, Eq. (3), after calculating GRGs for the spectrum. A Poisson noise and various linear functions were intentionally introduced to the valence band of μc-SiO$_x$:H to simulate noise and errors commonly observed in analysis of X-ray spectroscopic data. A comparison of the results obtained by GRM and LSF suggests that GRM is reliable in analyzing spectra with poor signal-to-noise ratio, and is much less sensitive to the deviation in data pre-processing, such as inconsistence in baseline shifting and correction. Furthermore, the results acquired by GRM from the valence band agree well with those obtained by curve fit of Si 2p HAXPES spectra and the information gained by other experimental techniques (i.e. RBS). This indicates that the proposed method is effective for analyzing X-ray spectroscopic data, with less demanding in personal experience and great potential for automatic data analysis of complex material systems.

CRediT authorship contribution statement

Rui Mu: Conceptualization, Methodology. **Yujie Zheng:** Data curation. **Andreas Lambertz:** Resources. **Regan G. Wilks:** Investigation, Writing - review & editing. **Marcus Bär:** Supervision, Writing - review & editing. **Yufeng Zhang:** Project administration, Validation, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rinp.2021.104031.

References

[1] Ramanujam J, Singh UP. Copper indium gallium selenide based solar cells – a review. Energy Environ Sci 2017;10(6):1306–19.

[2] Kroemer H. Quasi-electric fields and band offsets: teaching electrons new tricks (nobel lecture). ChemPhysChem 2001;2(8–9):490–9.

[3] Hwang HY, Iwasa Y, Kawasaki M, Keimer B, Nagaosa N, Tokura Y. Emergent phenomena at oxide interfaces. Nat Mater 2012;11(2):103–13.

[4] Nitta N, Wu FX, Lee JT, Yushin G. Li-ion battery materials: present and future. Mater Today 2015;18(5):252–64.

[5] Yang WS, Park BW, Jung EH, Jeon NJ, Kim YC, Lee DU, et al. Iodide management in formamidinium-lead-halide–based perovskite layers for efficient solar cells. Science 2017;356(6345):1376–9.

[6] Smirnov V, Lambertz A, Moll S, Bär M, Starr DE, Wilks RG, et al. Doped microcrystalline silicon oxide alloys for silicon-based photovoltaics: Optoelectronic properties, chemical composition, and structure studied by advanced characterization techniques: doped microcrystalline silicon oxide alloys for silicon-based photovoltaics. Phys Status Solidi A 2016;213:1814–20.

[7] Sarangan A. Nanofabrication. In: Fundamentals and Applications of Nanophotonics. Elsevier; 2016. p. 149–84.

[8] Teixeira V, Carneiro J, Carvalho P, Silva E, Azevedo S, Batista C. High barrier plastics using nanoscale inorganic films. In: Multifunctional and Nanoreinforced Polymers for Food Packaging. Woodhead Publishing; 2011. p. 285–315.

[9] Bagolini A, Gaiardo A, Crivellari M, Demenev E, Bartali R, Picciotto A, et al. Development of MEMS MOS gas sensors with CMOS compatible PECVD inter-metal passivation. Sens Actuators, B 2019;292:225–32.

[10] Aman SGM, Koretomo D, Magari Y, Furuta M. Influence of deposition temperature and source gas in PE-CVD for SiO$_2$ passivation on performance and reliability of In–Ga–Zn–O thin-film transistors. IEEE Trans Electron Devices 2018;65(8):3257–63.

[11] Vermang B, Brammertz G, Meuris M, Schnabel T, Ahlswede E, Choubrac L, et al. Wide band gap kesterite absorbers for thin film solar cells: potential and challenges for their deployment in tandem devices. Sustain Energy Fuels 2019;3(9):2246–59.

[12] Félix R, Witte W, Hariskos D, Paetel S, Powalla M, Lozac'h M, et al. Near-surface [Ga]/([In] + [Ga]) composition in Cu(In, Ga)Se$_2$ thin-film solar cell absorbers: an overlooked material feature. Phys Status Solidi A 2019;216(18):1800856.

[13] Ravel B, Newville M. ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT. J Synchrot Radiat 2005;12:537–41.

[14] Bravo-Sanchez M, Romero-Galarza A, Ramirez J, Gutierrez-Alejandre A, Solis-Casados DA. Quantification of the sulfidation extent of Mo in CoMo HDS catalyst through XPS. Appl Surf Sci 2019;493:587–92.

[15] Gong HN, Shen H, Huang HZ, Wu NZ. Preliminary exploration of Touggard's method used for the background subtraction in XPS. Acta Phys Chim Sin 2002;18(4):326–31.

[16] Miyazato I, Takahashi L, Takahashi K. Automatic oxidation threshold recognition of XAFS data using supervised machine learning. Mol Syst Des Eng 2019;4(5):1014–8.

[17] Kuo Y, Yang T, Huang GW. The use of grey relational analysis in solving multiple attribute decision-making problems. Comput Ind Eng 2008;55(1):80–93.

[18] Wei GW. GRA method for multiple attribute decision making with incomplete weight information in intuitionistic fuzzy setting. Knowl Based Syst 2010;23(3):243–7.

[19] Muthuramalingam T, Akash R, Krishnan S, Phan NH, Pi VN, Elsheikh AH. Surface quality measures analysis and optimization on machining titanium alloy using CO$_2$ based laser beam drilling process. J Manuf Processes 2021;62:1–6.

[20] Thangaraj M, Annamalai R, Moiduddin K, Alkindi M, Ramalingam S, Alghamdi O. Enhancing the surface quality of micro titanium alloy specimen in WEDM process by adopting TGRA-based optimization. Materials 2020;13(6).

[21] Nguyen PH, Banh TL, Mashood KA, Tran DQ, Dong Pham V, Muthuramalingam T, et al. Application of TGRA-based optimisation for machinability of high-chromium tool steel in the EDM process. Arab J Sci Eng 2020;45(7):5555–62.

[22] Li Z, Chen L. A novel evidential FMEA method by integrating fuzzy belief structure and grey relational projection method. Eng Appl Artif Intell 2019;77:136–47.

[23] Li X, Wang Z, Zhang L, Zou C, Dorrell DD. State-of-health estimation for Li-ion batteries by combing the incremental capacity analysis method with grey relational analysis. J Power Sources 2019;410–411:106–14.

[24] Sun G, Guan X, Yi X, Zhou Z. Grey relational analysis between hesitant fuzzy sets with applications to pattern recognition. Expert Syst Appl 2018;92:521–32.

[25] Wiech G, Feldhütter HO, Šimůnek A. Electronic structure of amorphous SiOx:H alloy films studied by X-ray emission spectroscopy: Si K, Si L, and O K emission bands. Phys Rev B 1993;47(12):6981–9.

[26] Zakaznova-Herzog VP, Nesbitt HW, Bancroft GM, Tse JS, Gao X, Skinner W. High-resolution valence-band XPS spectra of the nonconductors quartz and olivine. Phys Rev B 2005;72(20):205113.