This article was published in

M. Aronniemi, J. Sainio, J. Lahtinen, Aspects of using the factor analysis for XPS data interpretation, Surface Science 601, 479-489 (2007).

© 2007 Elsevier Science

Reprinted with permission.

Available online at www.sciencedirect.com
ScienceDirect
Surface Science 601 (2007) 479–489
SURFACE SCIENCE
www.elsevier.com/locate/susc

Aspects of using the factor analysis for XPS data interpretation

M. Aronniemi *, J. Sainio, J. Lahtinen

Laboratory of Physics, Helsinki University of Technology, P.O. Box 1100, FI-02015 TKK, Finland

Received 22 June 2006; accepted for publication 10 October 2006
Available online 3 November 2006

Abstract

In this work, the use of factor analysis to chemical state quantification of XPS data is studied. First, the theory of the method is reviewed with a special emphasis on the issues related to XPS data analysis. In particular, we concentrate on the transformation of the abstract components into physically meaningful ones in the case where reliable reference spectra are not available. We have observed that in the commonly used iterative target transformation factor analysis (ITTFA), in which a delta peak serves as the initial guess, the shape of the obtained component depends strongly on the position of the delta peak and on the minimum allowed intensity level. We propose an approach in which these parameters are varied in order to generate different representations for each component of the data. With simulated model data we show that if the variation is done with a sufficiently small step size, the correct representation will be generated. We also show that in the case of two-component data the iteration of the components is not necessary because a position can be found where a delta peak directly transforms into the correct component without unphysical features. Besides the model data, the proposed method is applied to experimental 2p photoelectron spectra of iron and chromium oxides.
© 2006 Elsevier B.V. All rights reserved.

Keywords: XPS; Factor analysis; Iron oxide; Chromium oxide; Chemical state

1. Introduction

When XPS (X-ray photoelectron spectroscopy) is used to study changes on a sample surface, each of the interesting peak regions is typically recorded successively. This produces series of spectra for the analysis. Examples of this kind of surface studies are depth profiling and oxidation/reduction treatments. Qualitative behavior, such as oxidation or reduction, can in most cases be deduced by visual comparison of the spectra, but in order to obtain quantitative information, such as the proportion of each oxidation state through the series, mathematical methods become necessary. Non-linear curve fitting and factor analysis are typically used for this purpose.

Factor analysis is a multivariate statistical method which can be used to analyze a set of spectra if all the spectra can be expressed as a linear combination of a few components (also called principal components or factors). Considering XPS data, the components correspond typically to different chemical states of the element under study. The advantages of factor analysis over curve fitting are that a complete set of spectra can be analyzed at once and that there is no need for a mathematical function describing the lineshape. Since the early 1980s, application of factor analysis to XPS data has been reported by several authors, see e.g., Refs. [1–20].

The objectives of this article are to illustrate the applicability of factor analysis to quantitative determination of chemical states in XPS data, to point out aspects which affect the accuracy of the results, and to propose modifications to the iterative target transformation factor analysis (ITTFA) technique which can be used when no reference spectra are available for the analyzed chemical states.

The data analyzed in this paper consist of simulated model spectra and experimental data on 2p regions of iron and chromium oxides. These oxides were chosen because they present a large intrinsic background and complicated 2p spectra and because only few articles [3,5,18,20] have

* Corresponding author. Tel.: +358 9 451 3132.
E-mail addresses: Mikko.Aronniemi@hut.fi, aro@fyslab.hut.fi (M. Aronniemi).

0039-6028/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.susc.2006.10.010

been published regarding the application of factor analysis to them. These oxides are also widely studied in the field of surface science due to their various applications in, e.g., catalysis, steel fabricating, and gas sensors. Our interest has been in the growth of iron oxide thin films [21] and in the growth of chromium oxide islands on alumina [22–24].

2. Experimental

The data analyzed in this study were collected with a Surface Science Instruments SSX-100 ESCA spectrometer using monochromatic Al K_α X-rays and an electrostatic hemispherical analyzer. The spectra were recorded with a pass energy of 50 eV, X-ray spot size of 600 μm, and step size of 0.1 eV. Base pressure in the analysis chamber was around 10^{-9} mbar.

The algorithms used in this work were programmed with Matlab version 6.5.1. The calculations were done with a 2-GHz desktop computer.

2.1. Samples

The studied iron oxide sample was commercial FeO (wustite) powder (99.5%, Alfa Aesar), crushed into an indium foil under ambient atmosphere. In stoichiometric FeO iron ions are in the Fe^{2+} state, but due to the atmospheric exposure, the surface of the sample was oxidized to Fe^{3+} and contained some adventitious carbon. The data set, consisting of seven spectra, was collected by sputtering the sample with 4-keV argon ions and recording the Fe 2p photoelectron spectrum as a function of the cumulative sputtering time. In addition, O 1s, In 3d, and C 1s spectra were recorded after each sputtering period. Because no significant shifting of O 1s, In 3d, or C 1s peak positions was observed, the BE (binding energy) scale of the Fe 2p spectra was not corrected. The position of the oxidic O 1s peak was 530.6 eV.

For chromium a commercial foil sample (99.99%, Goodfellow) was used. It was first sputtered with argon ions in order to reveal the clean metallic surface, and then oxidized at room temperature in UHV with O_2 exposures increasing up to 100 L. Finally, the sample was oxidized in air flow in a reactor cell at 400 °C for 15 min. The oxygen exposures were chosen so that the chromium would be in states Cr^0 and Cr^{3+} [5]. The Cr 2p region was recorded after each oxidizing step, which produced a set of 10 spectra. The BE scale was shifted to set the O 1s peak to 530.5 eV.

2.2. Background subtraction

The fundamental assumption of factor analysis – that the spectra to be analyzed can be represented as a linear combination of component spectra – makes the role of the inelastic background subtraction particularly important. Generally, the background contribution is not a constant component and if a wrong background function is used, each spectrum contains a unique background residual, which violates the assumption and makes the analysis subject to errors. Only in the case of invariable depth distribution of the sample composition the background can be included in the components.

In the reported studies comparing the three common backgrounds, linear, Shirley [25] and Tougaard [26], subtraction of the Tougaard background has been found to produce the true unattenuated spectrum (primary excitation spectrum) most correctly [27,28]. Considering the chemical state analysis of Fe 2p and Cr 2p spectra, the effect of the background subtraction method has been discussed in Ref. [29]. Based on the results reported in these papers, the so called universal Tougaard background [26] was chosen to be used in this work. The background was evaluated at 695–820 eV for the Fe 2p spectra and at 564–770 eV for the Cr 2p spectra. The C parameter was set to the "universal" value 1643 eV^2 and the value of the B parameter was determined so that the background follows the measured intensity in the end of the energy range. In the iron oxide spectra, B decreased from about 3440 eV^2 to 3280 eV^2 with increasing sputtering time. For the chromium oxide, B was about 2900 eV^2 in all the spectra.

For the factor analysis, a shorter BE range was considered useful in order to have more weight for the peak area and to decrease the computing time. Thus, the spectra were truncated after the background subtraction so that the analysis range became 695–760 eV for the Fe 2p region and 564–660 eV for Cr 2p. Finally, all the spectra were normalized to unit area.

3. Factor analysis

3.1. Decomposition of the data

In order to apply factor analysis to XPS data, the s measured spectra, each having p points, are first arranged to columns of a data matrix D. The objective of the analysis is to decompose the data matrix D ($p \times s$) into a product of two matrices

$$D = RC. \tag{1}$$

The columns of R ($p \times n$) are called components or factors and the coefficients for a given component are on the corresponding row of C ($n \times s$). In terms of XPS this means that each measured spectrum will be expressed as a linear combination of n component spectra, typically corresponding to various chemical states, and the relative amount of each component can be calculated directly from the corresponding column of C. The second objective of the factor analysis is to help determine the number of components needed to reproduce the data within the experimental error.

Decomposition of the data matrix is done by calculating the eigenvectors of the covariance matrix D^TD. Several

techniques can be applied to this purpose; in this work we have used singular value decomposition (SVD) which decomposes the data matrix D as

$$D = USV^{\mathrm{T}} \tag{2}$$

so that R and C are obtained as

$$R = US, \tag{3}$$

$$C = V^{\mathrm{T}}. \tag{4}$$

Now the columns of R are components that can be used to reproduce the measured spectra, i.e., R forms a basis in which the columns of D can be expressed. Above, S is a diagonal matrix having the square roots of the eigenvalues in decreasing order on the diagonal. The columns of U are the eigenvectors of DD^{T} and the columns of V are the eigenvectors of $D^{\mathrm{T}}D$.

In an ideal case, R would have exactly as many columns as necessary to reproduce D; this is equal to the rank of D. Experimental data, however, contain noise which increases the number of the columns of R equal to the number of the columns of D. It turns out that only the components corresponding to the largest eigenvalues present significant spectral features whereas those with smaller eigenvalues consist mainly of noise. Thus, when determining the number of necessary components, the analyst typically seeks a distinctive drop in the magnitude of the eigenvalues and compares the structure of the components to the noise. With the chosen number of components, $\bar{n} < n$, the data matrix is reproduced as

$$D \approx \overline{D} = \overline{R}\,\overline{C}, \tag{5}$$

where $\overline{R}$ consists of the $\bar{n}$ first columns of R and $\overline{C}$ of the $\bar{n}$ first rows of C. With a proper number of components, the recorded spectra can be reproduced with decreased noise.

In addition to the comparison of the eigenvalues and component shapes, several mathematical functions have been introduced to assist in the determination of the number of necessary components. One of the most commonly used is the indicator function proposed by Malinowski [30]. It can be formulated as

$$f_{\mathrm{IND}}(\bar{n}) = \Delta_{\mathrm{RMS}}(D - \overline{D})\sqrt{\frac{s}{(s-\bar{n})^5}}, \tag{6}$$

where $\Delta_{\mathrm{RMS}}(D - \overline{D})$ is the RMS difference between the recorded and reproduced data. This empirical function has been designed to minimize at a proper number of components. Obviously, the RMS difference decreases as more components are included, and the expression in the square root is employed as a correction factor to increase the function value.

3.2. Physical interpretation of components

Although capable of reproducing the recorded spectra, the columns of $\overline{R}$ are in general not physically meaningful and are thereby called abstract components. In order to obtain a physically meaningful results, $\overline{R}$ and $\overline{C}$ need to be transformed by multiplying them with an appropriate transformation matrix T:

$$X = \overline{R}T, \tag{7}$$

$$Y = T^{-1}\overline{C}. \tag{8}$$

Now the columns of X are the new components, i.e., the spectra of the different chemical states, and the rows of Y have the corresponding coefficients, i.e., the relative abundances. Mathematically this means that a new basis (columns of X) is formed as a linear combination of the original basis (columns of $\overline{R}$). Because there are infinitely many ways to form the linear combination, the components cannot be determined unambiguously, and the physically meaningful ones are just one combination.

3.2.1. Target transformation

Several methods have been presented in the literature for constructing the transformation matrix, see e.g., Ref. [30] for a review. In a method called target transformation or target testing, the transformation matrix T is obtained by guessing the components and expressing them in the basis formed by $\overline{R}$. Malinowski has shown [30] that for a given test component (vector) $\tilde{x}_l$ a representation, called a predicted component x_l, in the $\overline{R}$ basis is obtained as

$$x_l = \underbrace{\overline{R}(\overline{R}^{\mathrm{T}}\overline{R})^{-1}\overline{R}^{\mathrm{T}}}_{\equiv A}\tilde{x}_l. \tag{9}$$

This transformation minimizes in the least-squares sense the deviation between the test component and the predicted component [30]. It is to be noted that the matrix A is calculated using only the abstract component matrix $\overline{R}$. Thus, each component candidate is transformed independently.

Typically, experimental reference spectra or fits to them are used as test components. If $x_l \approx \tilde{x}_l$ within a specified tolerance or the shape of x_l is otherwise satisfactory, x_l is accepted for being used in the final data interpretation. When a necessary number ($\bar{n}$) of linearly independent components have been found, matrix X is formed by setting the vectors x_l, $l = 1 \ldots \bar{n}$, as columns. The transformation matrix T is then obtained through Eq. (7) as

$$T = \overline{R}^{-1}X, \tag{10}$$

or through Eqs. (7) and (9) as

$$T = (\overline{R}^{\mathrm{T}}\overline{R})^{-1}\overline{R}^{\mathrm{T}}\widetilde{X}, \tag{11}$$

where the test components $\tilde{x}_l$ form the columns of $\widetilde{X}$. Finally, the corresponding coefficients (matrix Y) are calculated with Eq. (8). This method has been used in XPS data analysis, e.g., in Refs. [6–8,18].

Instead of the component spectra, guesses of their relative amounts, i.e., the elements of C, can also be used to calculate the transformation matrix. Examples of this approach can be found in Refs. [5,9,19,31].

3.2.2. Iterative target transformation factor analysis (ITTFA)

In many cases, a reliable reference spectrum or even a reasonable guess may not be available for the components. Gemperline has introduced a method called needle search where a delta peak is used as the test component [32]. The position of the delta peak is scanned through the x-axis of the data and the predicted component is calculated at each position using Eq. (9). The squared difference between the delta peak and the predicted component is then plotted as a function of the delta peak position and the local minima are interpreted as the probable component positions. Gemperline has also proposed an iterative procedure to find physically meaningful components without any reference data [32]. In this method, a delta peak placed at a given local minimum is taken as the initial test component and the predicted component is calculated using Eq. (9). In each iteration round, the predicted component is refined by removing unphysical features and then used as a new test component. Typically, the refining is done by updating points below a certain predetermined intensity level to some positive value. The iteration scheme can be formulated as

$$x_{l,i+1} = Ax'_{l,i}, \tag{12}$$

where $x'_{l,i}$ is the refined version of $x_{l,i}$. The iteration is continued until no refining is needed, i.e., $x'_{l,i} = x_{l,i}$, or if this cannot be achieved in reasonable time, until $x_{l,i+1} \approx x_{l,i}$ within a desired limit. The iteration is done separately for each of the $\bar{n}$ components.

This method is called iterative target testing or iterative target transformation factor analysis (ITTFA). It has been applied to XPS data in, e.g., Refs. [2–4,15–17].

3.2.3. Varying the parameters of ITTFA

Although knowledge of the peak shape or exact position is not needed in ITTFA, the analyst has to set values of two parameters for each component: (i) the position of the delta peak serving as the initial test component and (ii) the minimum allowed value (minimum level) for the intensity of the predicted component. When applying ITTFA to our XPS data, we have observed that in many cases the values of these parameters cannot be determined unambiguously.

Regarding the delta peak position, the needle search output does not always show $\bar{n}$ clear local minima or setting a delta peak at a minimum does not always result in an acceptable component after iteration. In these cases it is necessary to change the delta peak position which in turn will change the predicted component.

The ambiguity of the minimum level results from the noise. The background subtraction is usually performed so that the mean intensity is about zero in the off-peak region, typically in both ends of the analyzed BE range. In the presence of noise this means that the spectra inevitably contain negative values, and the iteration will typically proceed too long if the minimum allowed intensity value is set strictly to zero; this has also been pointed out by Gemperline [33]. Thus, a negative minimum level is needed but it seems to be impossible to determine any correct value for it. Moreover, the value of the level significantly affects the shape of the predicted component and thereby the analysis results.

Because the delta peak positions and the minimum allowed intensity level cannot be determined a priori, it is convenient to vary these parameters in order to generate different representations for each principal component. With synthetic model data we have observed that if the variation is done with a sufficiently small step size, the set of generated components includes the correct representation for each principal component. In the case of experimental data, there is no general way to identify the correct representation, and thus the practical strategy is to reject the unphysical ones based on the available information on the true shape of the component. Here, at least two requirements for XPS spectra can be used: (i) the component must not have features that go below the noise level and (ii) the chosen set of components must be such that their coefficients (elements of Y) are positive. Depending on the analyzed data, it may be possible to formulate additional criteria based, e.g., on the symmetry of the peaks. Typically, the analyst ends up with a set of acceptable representations for each component and this determines the uncertainty of the analysis. It should be noted that different representations of the principal components cannot be compared by observing the deviation between the measured and reconstructed data. This deviation only shows how well the data can be expressed with the chosen number ($\bar{n}$) of components and is independent on the shape of the components as long as the components are expressed in the $\overline{R}$ basis.

We have found that if the data can be explained with two components, the component iteration is not needed. It namely turns out that a position can be found where only one multiplication by A (Eq. (9)) is enough to transform a delta peak into a physically meaningful component. In practice, a delta peak is scanned through the BE axis like in the needle search but instead of minimizing the deviation between the delta peak and the predicted component, the optimal shape of the predicted component is searched.

In many cases it seems that the optimal position for the delta peak would lie between two measured BE values. To reach this, the BE step size of the data matrix can be decreased using linear interpolation; this increases the number of points but does not create artificial features, such as smoothing, to the original spectra. The corresponding $\overline{R}$ and A are calculated with Eqs. (3) and (9), respectively. It turns out that after the optimal predicted component has been generated with a decreased BE step size, it can be expressed equivalently on the original BE axis and it gets an equal proportion in the spectra. Alternatively, a narrow Gaussian peak can be used as the test component instead of a delta peak. This allows setting the peak center between the points of the original BE scale, and decreasing the step size is not needed. For the data analyzed in this

work, a narrow Gaussian with the FWHM (full width at half maximum) equal to the BE step size of the data (0.1 eV) was found to give results which were equal to those obtained with a delta peak.

If three or more components are needed to explain the data, a single multiplication of a delta peak by A is not always able to generate the correct representation for every principal component, and the iteration is needed to remove the unphysical features. In such case, different representations can be generated by varying both the delta peak position and the minimum allowed intensity level individually for each component.

4. Results and discussion

4.1. Model data

Before analyzing the real experimental spectra, synthetic model data was used to verify that the correct components can be generated by varying the ITTFA parameters.

4.1.1. Two-component data

The first example, shown in Fig. 1, is a set of 11 spectra created using two Gaussian peaks with height of 1, FWHM of 2 eV, and centers at 10.0 eV and 11.5 eV. The proportion of the first component, defined here as $A_1/(A_1 + A_2)$ where A_i is the area of component i, increases from 10% to 90% through the set. The BE range of the spectra is 20 eV with a step size of 0.1 eV. Gaussian noise with standard deviation of 0.01 was added to the spectra producing a signal-to-noise ratio (SNR) of ~95.

First, the standard ITTFA was applied to the model data. For the lower BE component, the needle search indicated (correctly) a position of 10.0 eV, and a delta peak, serving as the test component, was placed there. The iteration was done with two different minimum levels, 0 and the minimum of the data. When the level was set to 0, the iteration converged after about 10,000 rounds, whereas only six rounds were required when the data minimum was used. The obtained predicted components, as well as the true one, are shown in Fig. 2. It is observed that the minimum allowed intensity clearly affects the component shape but neither of the obtained components matches with the true one, i.e., that used to create the data. Thereby, it seems reasonable to vary the ITTFA parameters in order to generate a better representation for the component.

Fig. 3a illustrates the idea of varying the position of the delta peak. At each position, the delta peak is multiplied only once by A (see Eq. (9)) and no iteration is performed. The figure clearly indicates that the optimal position for the delta peak can be found around 10.4 eV: setting the delta peak below the optimal position produces components with a negative dip above the peak. Delta peaks above the optimal position result in asymmetric components with extra intensity on the high BE side.

With experimental data, the true component shape is naturally not known, and the next step after generating a set of predicted components is to reject the unphysical ones. In the example above (Fig. 3a), the predicted components with the negative dip can be obviously rejected (requirement (i) in Section 3.2.3). On the other hand, the predicted components with extra intensity on the high BE side have no negative features and thus cannot be rejected unless they get negative coefficients (requirement (ii)). If the maximum proportions of the true components in the analyzed set of spectra are less than 100% (i.e., purely

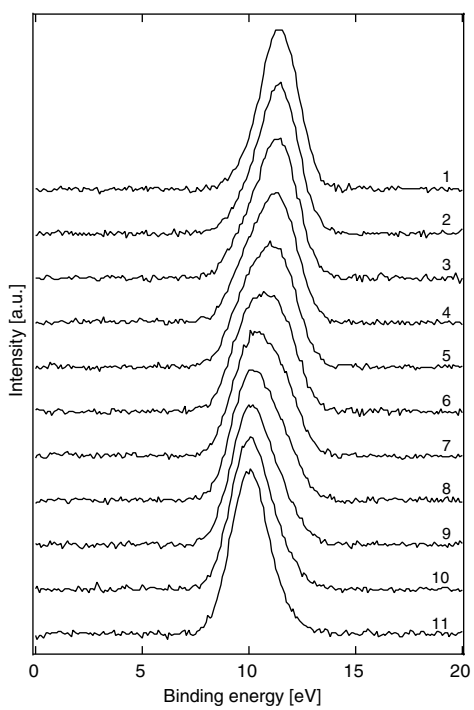

Fig. 1. Example of a model data set created using two Gaussian peaks and noise. The component separation is 1.5 eV and the FWHM of the peaks is 2 eV.

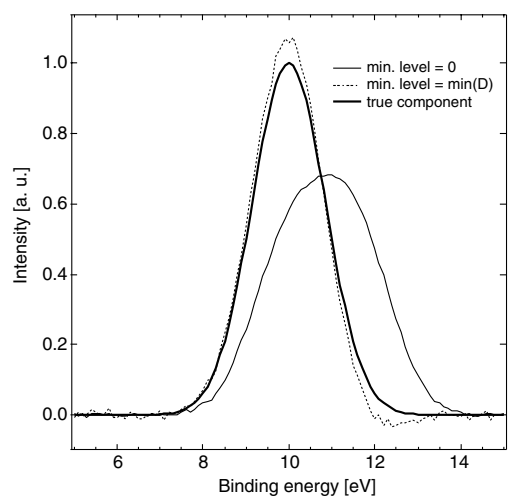

Fig. 2. Lower BE component of the data in Fig. 1 obtained using the standard ITTFA with the minimum allowed intensity level set to 0 and to the minimum of the data. Also shown is the true component.

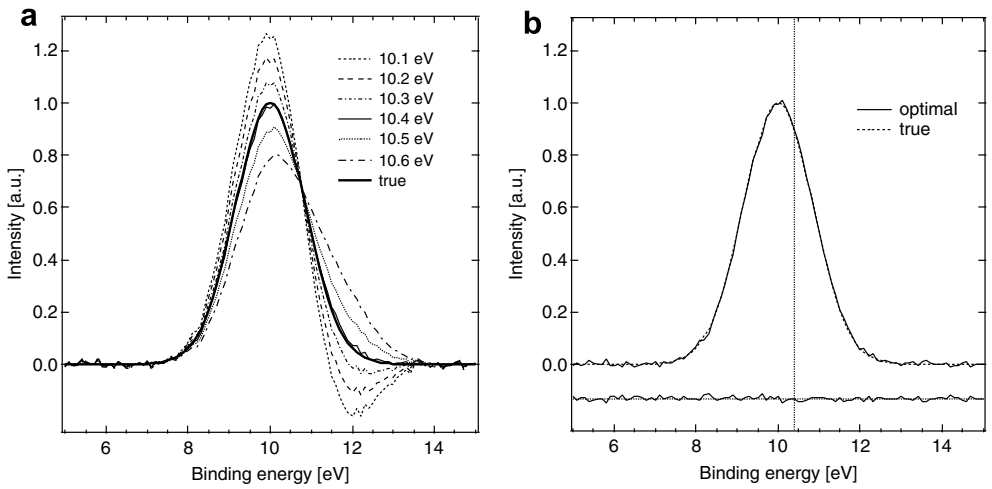

Fig. 3. (a) Searching the lower BE component of the data in Fig. 1: predicted components corresponding to six different delta peak positions with 0.1 eV separation. The true component is shown for comparison. (b) Comparison between the optimal predicted component, generated setting a delta peak at 10.39 eV, and the true lower BE component. The difference is shown below the spectra and the optimal delta peak position is indicated with a dotted vertical line.

single-component spectra are not included in the set), some non-negative mixed components will always get positive coefficients and have to be accepted. This naturally brings uncertainty to the analysis results which can be reduced only by applying some additional criteria for the component shape.

Next, we wanted to find out whether a better component representation can be obtained if the delta peak position is varied with a step size smaller than the BE step size of the data. To this end, predicted components were generated by varying the delta peak position with different step sizes, starting from the BE step size of the data (0.1 eV) and then halving it successively. As a measure of the analysis error E_{FA}, we used the mean absolute deviation of the component proportion from its true value:

$$E_{\mathrm{FA}} = \frac{\sum_{i=1}^{\bar{n}} \sum_{j=1}^{s} |f_{i,j}^{*} - f_{i,j}|}{\bar{n}s}, \tag{13}$$

where $f_{i,j}$ is the obtained proportion of component i in spectrum j and $f_{i,j}^{*}$ is the corresponding true value which was used in creating the data. The random error was characterized by creating and analyzing 100 data sets with an equal noise distribution. The error which would result if the analyst were able to select the optimal (in the least-squares sense) representations for the components is plotted as a function of the step size in Fig. 4. Also shown is the lowest achievable error estimated using the true components as the test components in Eq. (11) which corresponds to an ideal case where a perfect noiseless reference spectrum is available for each component. It is observed that the error made with the optimal predicted components decreases with the step size and finally reaches the ideal level when the delta peak positioning becomes sufficiently accurate.

Fig. 3b shows the optimal predicted component, searched with a step size of 0.01 eV, for the lower BE component of the model data in Fig. 1; also shown is the true component. It is observed that the two spectra are practically identical and the difference is only noise. With a similar procedure the higher BE component can be generated as well.

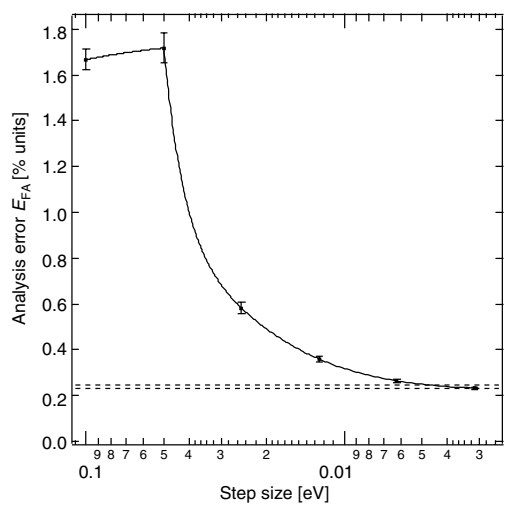

Fig. 4. Dependency of the analysis error on the step size with which the delta peak position is varied. The analyzed data was similar to that shown in Fig. 1. Error bars indicate the standard deviation of the mean. Solid line was drawn to guide the eye. The two dashed lines show the standard deviation of the mean of the lowest achievable error as described in the text.

The effects of two essential properties of the analyzed data, namely the component separation (overlap) and the noise level, on the analysis performance were studied by creating and analyzing two series of data sets where these parameters had different values. The first series consists of seven data sets with the BE separation between the two components (Gaussian peaks) increasing from 0.5 eV to 3.5 eV; the standard deviation of noise was 0.010 (SNR $\sim$ 95). In the second series, having six data sets, the

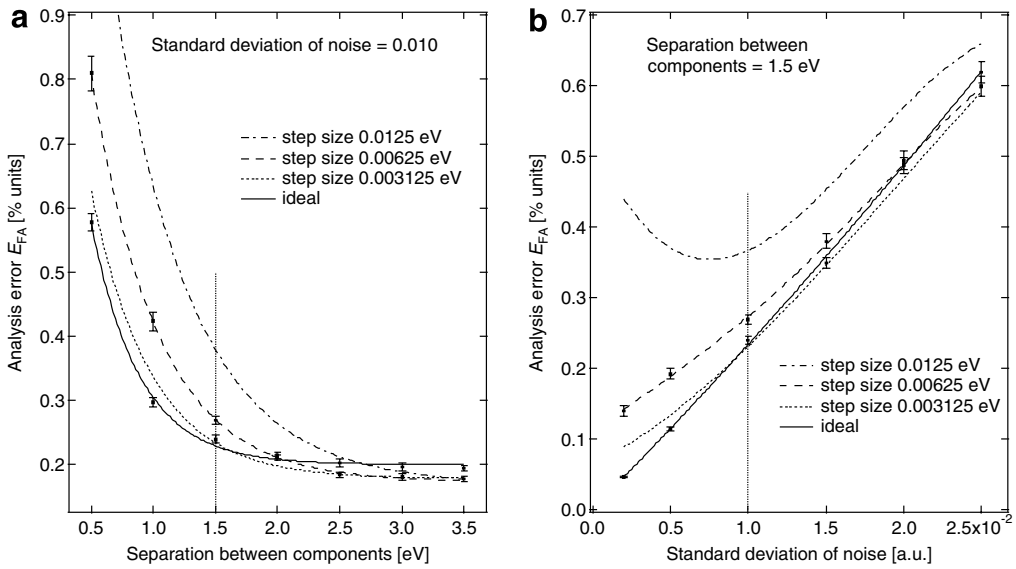

Fig. 5. Error of the analysis as a function of (a) component separation and (b) noise level for different step sizes. The ideal error has been calculated using the true noiseless components as the test components. The error bars show the standard deviation of the mean (not shown for all curves). The smoothed lines were drawn through the points to guide the eye. The dotted vertical lines indicate the parameters corresponding to Fig. 4.

separation was kept fixed at 1.5 eV but the standard deviation of noise was increased from 0.002 to 0.025 (SNR ~ 450–40). Otherwise each data set in these series was similar to the one shown in Fig. 1. The random error was, as previously, estimated by generating each data set 100 times with an equal noise distribution. In Fig. 5 the performance of the analysis is described by plotting the analysis error, obtained as above by selecting the optimal components, as a function of the component overlap and noise level for three different step sizes. Also shown is the error resulting from the use of the ideal reference spectra. It is observed that the analysis error stays below 1% unit in a wide range of component separation and noise level and that the ideal level is reached when the delta peak position is varied with a sufficiently small step size.

Table 1 compares the optimal delta peak positions to the true component positions for the series where the component separation was increased (Fig. 5a). The reported values were obtained with noiseless data, varying the delta peak position with a step size of 0.01 eV. It is interesting to note that although the positions of the optimal predicted components (not reported in the table) coincided in all cases with the true component positions, the deviation between the optimal delta peak position and the true component position increases systematically with increasing component separation. This raises a question whether the optimal positions could be deduced directly from the data matrix. This would eliminate the need for generating several component representations and make the analysis results unambiguous.

Table 1 includes also the component positions indicated by the needle search (see Section 3.2.2). In the literature these positions are sometimes interpreted as the true positions of the components. Table 1 shows, however, that for a small component separation (i.e., strong overlap), the positions indicated by the needle search are too wide apart; thus, e.g., the shift between two chemical states should not be determined using only the needle search results.

Table 1
Comparison between the true component positions, the positions of the delta peaks yielding the best predicted components, and the positions indicated by the needle search

True component positions (eV)	Optimal delta peak positions (eV)	Needle search positions (eV)
10.0 and 10.5	10.13 and 10.37	9.6 and 10.9
10.0 and 11.0	10.25 and 10.75	9.8 and 11.2
10.0 and 11.5	10.37 and 11.13	9.9 and 11.6
10.0 and 12.0	10.50 and 11.50	10.0 and 12.0
10.0 and 12.5	10.62 and 11.88	10.0 and 12.5
10.0 and 13.0	10.75 and 12.25	10.0 and 13.0
10.0 and 13.5	10.87 and 12.63	10.0 and 13.5

The reported values have been determined for a series of noiseless 2-component model data corresponding to Fig. 5a.

4.1.2. Three-component data

As an example of three-component data, a set of 10 spectra, shown in Fig. 6, was created using three asymmetric peaks, with proportions varying between 10% and 80%, and Gaussian noise with a standard deviation of 0.01. The BE range was 0–20 eV with a step size of 0.1 eV.

As previously, a set of predicted components were first generated by scanning the delta peak position. For the outermost components (numbers 1 and 3) the optimal predicted components were obtained without iteration as described in the case of two-component model data. For the middle component (number 2), however, all representations contained negative features and the iteration

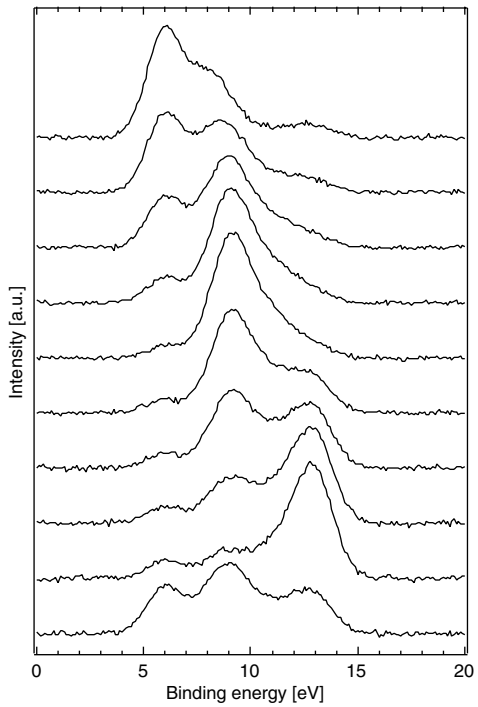

Fig. 6. Three-component model data created using three asymmetric peaks and Gaussian noise.

(Eq. (12)) became necessary. Because the minimum allowed intensity level affects the result of the iteration, it was varied along with the delta peak position. Fig. 7 shows the optimal components and the true ones; here the delta peak position was varied with a step size of 0.01 eV and the minimum allowed intensity level with $0.01 \cdot D_{\min}$, where $D_{\min}$ is the minimum intensity of the analyzed data. The optimal delta peak positions are shown with dotted lines in the figure; the optimal minimum level for the middle component was −0.0225 which is 2.25 times the standard deviation of the generated noise. With these components the analysis error E_{FA} was 0.17% which equals the error obtained with the true components.

Concerning the iteration, it was found out that the intensity value to which the negative points are updated in each iteration round affects the convergence behavior: increasing this value makes the iteration converge more rapidly but too large values lead to distorted component shapes. Here, we set this value to 1% of the component maximum intensity.

In conclusion, the analysis of the model data has shown that in ITTFA it is useful to generate several representations for each principal component by varying the position of the delta peak and the minimum allowed intensity level. If this variation is done with a sufficiently small step size, the set of generated components includes the correct representation. In the case of two-component data, the components can be generated without iteration simply by multiplying the delta peak at each position once by A.

4.2. Experimental data

4.2.1. Iron oxide

The recorded Fe 2p spectra of the iron oxide sample and an example of the Tougaard background are shown in Fig. 8. The indicator function (Eq. (6)) minimized at two suggesting that it is reasonable to represent the data with two components. Using two components is also well justified from the chemical standpoint because oxidized iron is normally encountered in the Fe^{2+} and Fe^{3+} states.

When predicted components were generated by varying the delta peak position with a step size of 0.01 eV, delta peaks around 711.14 eV and 711.70 eV were found to yield spectra that could be interpreted as the Fe^{2+} and Fe^{3+} states, respectively. The $2p_{3/2}$ maximum was at ~710.2 eV in the Fe^{2+}-like spectra and at ~711.4 eV in the Fe^{3+}-like

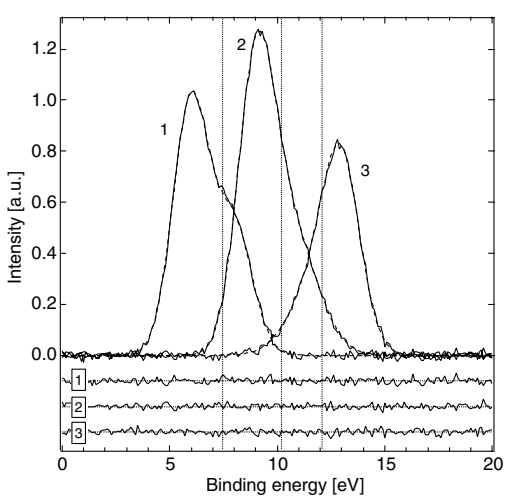

Fig. 7. Comparison between the optimal predicted components (solid line) and true components (dashed line) corresponding to the three-component model data in Fig. 6. The differences are shown below the spectra and the optimal delta peak positions are indicated with dotted vertical lines.

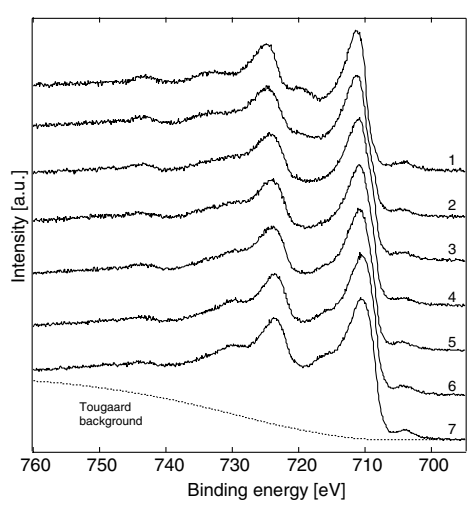

Fig. 8. Recorded Fe 2p photoelectron spectra and an example of the used Tougaard background. The cumulative sputtering time increases with the spectrum index. The small peak at ~704 eV comes from the indium foil used as the substrate.

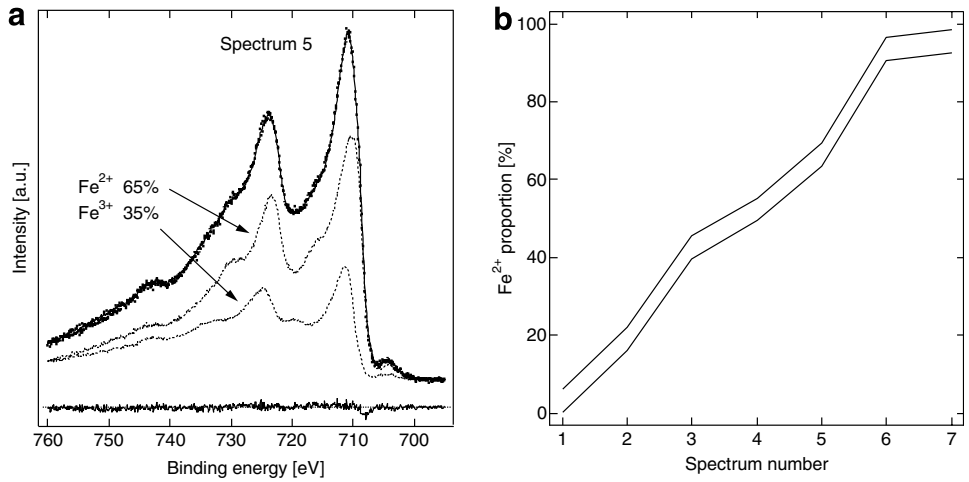

Fig. 9. Results of the factor analysis of Fe 2p spectra. (a) Spectrum number 5 reproduced using a pair of accepted component spectra. The difference between the measured and reproduced spectrum is shown at the bottom. (b) Proportion of the Fe^{2+} state in the analyzed set of spectra. The two lines correspond to the lowest and highest proportions obtained by combining the accepted component spectra.

spectra. These positions as well as the shape of the component spectra agree well with those reported in the literature.

Because the true shape of the component spectra was not known and the variation step size was so small, a group of slightly different predicted components, each with an acceptable shape and no negative features (requirement (i) in Section 3.2.3), was obtained for both states. Proportions of the Fe^{2+} and Fe^{3+} states were then evaluated corresponding to all possible combinations of these predicted components and only the combinations yielding positive proportions were accepted (requirement (ii) in Section 3.2.3). Fig. 9a shows an example of a pair of such accepted components and a reproduction of one of the measured spectra using them. The proportion of the Fe^{2+} state through the analyzed set of spectra is plotted in Fig. 9b. The two curves in the figure are the lowest and highest proportions obtained with the accepted components; thus, they represent the uncertainty that results from the lack of reference spectra or other exact knowledge on the component shape. Here, the difference between the lowest and highest Fe^{2+} proportion for a given measured spectrum is about 6% units.

For the sake of comparison, the Fe 2p data was analyzed also with the Shirley background subtracted. Also in this case the data could be explained with two components. The obtained Fe^{2+} proportion in the analyzed spectra deviated 0–20% units from the results obtained with the Tougaard background (Fig. 9b). Thus, the difference in the state proportions caused by the choice of the background subtraction method was in some spectra larger than that resulting from the uncertainty in the component shape. Such high sensitivity of the Fe 2p analysis to the background subtraction method was found also in our earlier study in which the spectra were analyzed with curve fitting [29].

As the third background alternative, the factor analysis was performed after subtracting only a constant background. This produced reasonable component spectra with proportions close to those obtained above with the Tougaard background.

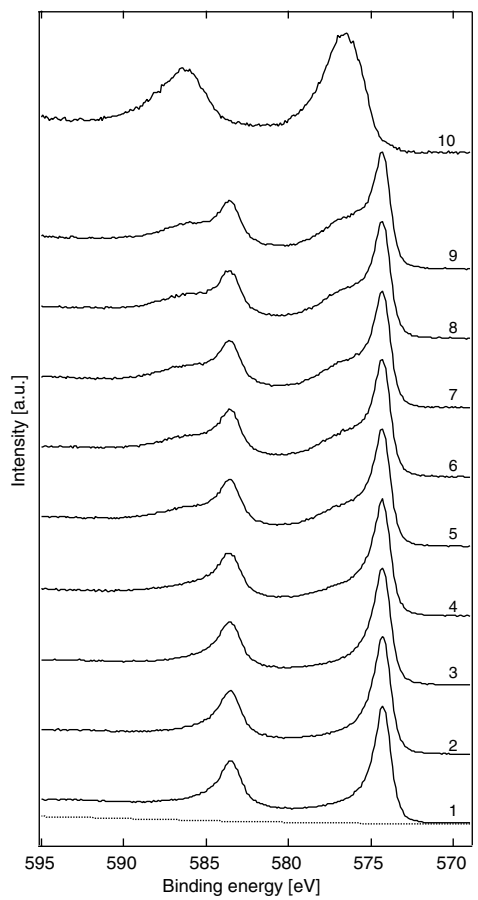

Fig. 10. Cr 2p region of the recorded chromium oxide spectra and an example of the Tougaard background (dotted line). Spectra 1–9 were recorded after O_2 exposures of 0–100 L at the room temperature and spectrum 10 after oxidizing at the atmospheric pressure at 400 °C. The BE range used in the background subtraction and factor analysis was wider than shown here, see Section 2.2.

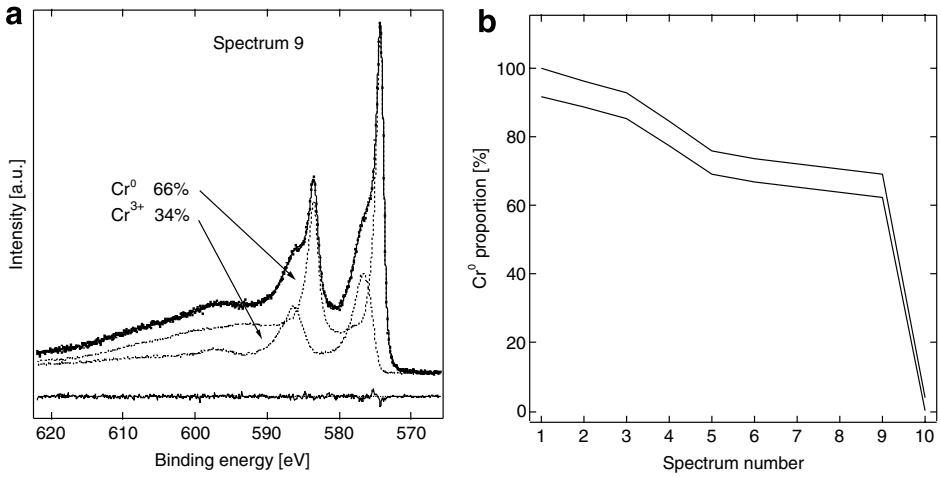

Fig. 11. Results of the factor analysis of Cr 2p spectra. (a) Spectrum number 9 reproduced using the obtained component spectra. The difference between the measured and reproduced spectrum is shown at the bottom. (b) Proportion of the Cr^0 state in the analyzed set of spectra. The two lines correspond to the lowest and highest proportions obtained by combining the accepted component spectra.

4.2.2. Chromium oxide

Fig. 10 shows the Cr 2p region of the recorded chromium oxide spectra and an example of the Tougaard background. Also for this set, the indicator function (Eq. (6)) minimized at 2.

Varying the delta peak position with a step size of 0.01 eV resulted in physically meaningful components around 575.17 eV and 575.65 eV. As in the case of iron oxide, several acceptable predicted components could be found for both states. However, the positions of the $2p_{3/2}$ main peak were ~574.3 eV (Cr^0) and ~576.6 (Cr^{3+}) eV in all acceptable component spectra; these as well as the shape of the components are well in accordance with the values reported in the literature. An example of a two-component reproduction of one of the spectra is shown in Fig. 11a and the maximum and minimum proportions of Cr^0 in the analyzed spectra are shown in Fig. 11b. Also in this case the obtained results seem reasonable.

5. Conclusions

In this work, we have studied the use of factor analysis to XPS data interpretation. First, the theory of the method was reviewed with a special emphasis on the issues related to XPS data. The most critical step in the factor analysis is the transformation of the abstract components to physically meaningful ones. In the XPS studies reported in the literature, this has been performed either by using recorded reference spectra or, if these are not available, by iterating a delta peak at a position indicated by the needle search (ITTFA). Typically, only one representation for each principal component (e.g., chemical state) is reported without any discussion of other possible alternatives and, thereby, the uncertainty in the component proportions.

We have observed that in ITTFA the delta peak position and the minimum allowed intensity level affect the shape of the obtained components. Values for these parameters cannot be determined before the analysis, but on the other hand, they provide a straightforward way to generate different representations for the components. The examples of simulated model data have shown that if these parameters are varied with a sufficiently small step size, the obtained set of predicted components includes the correct representation for each component. Typically this requires that the position of the delta peak is varied with a step size smaller than the BE step size of the data. This can be achieved either by increasing the number of data points by linear interpolation or using a narrow Gaussian peak instead of a delta peak. In the special case of two-component data, the correct components can be generated without iteration. Several acceptable representations are usually obtained for each principal component and this determines the uncertainty of the analysis.

Two sets of experimental XPS data, iron and chromium oxide 2p spectra, were analyzed with the presented method, and the obtained results seem reasonable.

Acknowledgements

The authors wish to acknowledge Ms. S. Ristiluoma for her assistance with the XPS measurements. This work has been financially supported by the Fortum Foundation, the Finnish National Graduate School in Materials Physics, the Academy of Finland, and the Finnish Funding Agency for Technology and Innovation (Tekes) through the Clean Surfaces Technology Program.

References

[1] R.A. Gilbert, J.A. Llewellyn, W.E. Swartz Jr., J.W. Palmer, Appl. Spectrosc. 36 (1982) 428.
[2] J.N. Fiedor, A. Proctor, M. Houalla, D.M. Hercules, Surf. Interface Anal. 20 (1993) 1.
[3] S.J. Scierka, A. Proctor, M. Houalla, J.N. Fiedor, D.M. Hercules, Surf. Interface Anal. 20 (1993) 901.

[4] M.A. Eberhardt, A. Proctor, M. Houalla, D.M. Hercules, J. Catal. 160 (1996) 27.
[5] C. Palacio, H.J. Mathieu, Surf. Interface Anal. 16 (1990) 178.
[6] A.R. Gonzáles-Elipe, J.P. Holgado, R. Alvarez, G. Munuera, J. Phys. Chem. 96 (1992) 3080.
[7] A.R. Gonzáles-Elipe, A. Fernández, J.P. Holgado, A. Caballero, G. Munuera, J. Vac. Sci. Technol. A 11 (1993) 58.
[8] J.P. Holgado, R. Alvarez, G. Munuera, Appl. Surf. Sci. 161 (2000) 301.
[9] R. Reiche, J.P. Holgado, F. Yubero, J.P. Espinos, A.R. Gonzáles-Elipe, Surf. Interface Anal. 35 (2003) 256.
[10] S. Oswald, S. Baunack, Surf. Interface Anal. 25 (1997) 942.
[11] S. Baunack, S. Oswald, D. Scharnweber, Surf. Interface Anal. 26 (1998) 471.
[12] S. Oswald, S. Baunack, Fresenius J. Anal. Chem. 365 (1999) 59.
[13] R. Reiche, R. Thielsch, S. Oswald, K. Wetzig, J. Electron. Spectrosc. Relat. Phenom. 104 (1999) 161.
[14] R. Reiche, S. Oswald, K. Wetzig, Appl. Surf. Sci. 179 (2001) 316.
[15] A. Arranz, Surf. Sci. 563 (2004) 1.
[16] C. Palacio, A. Arranz, Surf. Sci. 578 (2005) 71.
[17] A. Arranz, C. Palacio, Appl. Phys. A 81 (2005) 1405.
[18] T. Yamamura, N. Okuyama, Y. Shiokawa, M. Oku, H. Tomiyasu, W. Sugiyama, J. Electrochem. Soc. 152 (2005) B540.
[19] M.L. Trudeau, A. Tschöpe, J.Y. Ying, Surf. Interface Anal. 23 (1995) 219.
[20] W.F. Stickle, P.E. Sobol, Surf. Interface Anal. 19 (1992) 165.
[21] M. Aronniemi, J. Lahtinen, P. Hautojärvi, Surf. Interface Anal. 36 (2004) 1004.
[22] M. Eriksson, J. Sainio, J. Lahtinen, J. Chem. Phys. 116 (2002) 3870.
[23] J. Sainio, M. Eriksson, J. Lahtinen, Surf. Sci. 532–535 (2003) 396.
[24] J. Sainio, M. Aronniemi, O. Pakarinen, K. Kauraala, S. Airaksinen, O. Krause, J. Lahtinen, Appl. Surf. Sci. 252 (2005) 1076.
[25] D.A. Shirley, Phys. Rev. B 5 (1972) 4709.
[26] S. Tougaard, Surf. Interface Anal. 11 (1988) 453.
[27] S. Tougaard, C. Jansson, Surf. Interface Anal. 20 (1993) 1013.
[28] C. Jansson, S. Tougaard, G. Beamson, D. Briggs, S.F. Davies, A. Rossi, R. Hauert, G. Hobi, N.M.D. Brown, B.J. Meenan, C.A. Anderson, M. Repoux, C. Malitesta, L. Sabbatini, Surf. Interface Anal. 23 (1995) 484.
[29] M. Aronniemi, J. Sainio, J. Lahtinen, Surf. Sci. 578 (2005) 108.
[30] E.R. Malinowski, Factor Analysis in Chemistry, second ed., John Wiley & Sons, Inc., 1991.
[31] S.W. Gaarenstroom, J. Vac. Sci. Technol. 20 (1982) 458.
[32] P.J. Gemperline, J. Chem. Inf. Comput. Sci. 24 (1984) 206.
[33] P.J. Gemperline, Anal. Chem. 58 (1986) 2656.