



HAL
open science

Box plots: A simple graphical tool for visualizing overfitting in peak fitting as demonstrated with X-ray photoelectron spectroscopy data

Behnam Moeini, Hyrum Haack, Neal Fairley, Vincent Fernandez, Thomas Gengenbach, Christopher Easton, Matthew Linford

► **To cite this version:**

Behnam Moeini, Hyrum Haack, Neal Fairley, Vincent Fernandez, Thomas Gengenbach, et al.. Box plots: A simple graphical tool for visualizing overfitting in peak fitting as demonstrated with X-ray photoelectron spectroscopy data. *Journal of Electron Spectroscopy and Related Phenomena*, 2021, 250, pp.147094. 10.1016/j.elspec.2021.147094 . hal-03298446

HAL Id: hal-03298446

<https://hal.science/hal-03298446v1>

Submitted on 26 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Box Plots: A Simple Graphical Tool for Visualizing Overfitting in peak fitting as demonstrated with X-ray Photoelectron Spectroscopy data

Behnam Moeini,¹ Hyrum Haack,¹ Neal Fairley,² Vincent Fernandez,³ Thomas R. Gengenbach,⁴
Christopher D Easton,⁴ and Matthew R. Linford¹

¹Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT 84602, USA

²Casa Software Ltd., Bay House, Teignmouth, United Kingdom

³CNRS, Institut des Matériaux Jean Rouxel, IMN, Université de Nantes, F-44000 Nantes, France

⁴Commonwealth Scientific and Industrial Research Organization (CSIRO) Manufacturing, Clayton,
Victoria 3168, Australia

Abstract

Peak fitting is frequently performed in X-ray photoelectron spectroscopy (XPS). However, recent reports suggest that the current quality of this peak fitting is often inadequate in the scientific literature. Various statistical methods may be advantageously applied to an XPS peak fit to help determine the quality and validity of a fit. In this paper we describe a new statistical tool, which we believe will be helpful for determining the quality of protocols for fitting XPS data. This tool, box plots of random starting conditions, helps identify multiple local minima in a fit space. That is, ideally, different, reasonable starting conditions for a fit should lead to the same result, i.e., ideally, there should be a single global minimum for a fitting protocol. To determine whether a fit space contains multiple local minima, a series of reasonable, random starting conditions are chosen for the fit. If the boxes in the box plot of the peak areas of these fits are narrow, the different possibilities converge to a single global minimum. Conversely, if the boxes are wide, multiple local minima are present. Our approach is similar to the mathematical concept of ‘disproof by contradiction’. It is demonstrated herein in four- and ten-component fits to a moderately complex C 1s narrow scan. The resulting box plots compare favorably to traditional Monte Carlo analyses and uniqueness plots, although each of these statistical tools performs a different function/probes the fit space differently.

Introduction

X-ray photoelectron spectroscopy (XPS) is a quantitative, sensitive, robust, and informative method for chemically analyzing surfaces.[1–6] XPS identifies the elements at a surface and often the functional groups they are in, i.e., the oxidation states of the elements, which are manifested as changes in the binding energies of the elements (chemical shifts).[7] XPS can also provide information about film/surface structure and morphology.[8] XPS is widely used in many areas of science and technology, including in the analysis of thin films,[9, 10] catalysts,[11] semiconducting materials,[12] nanoparticles,[13] and polymers.[14] Indeed, XPS is the most widely used and important method for chemically analyzing surfaces. XPS data is typically acquired as either wide (survey)[15] or narrow (detail)[16] scans, where the former often provide important information about the elements in a sample (the general elemental composition) and the latter are peak fit to identify and quantify the chemical/oxidation states of the elements present. As a general rule, second row elements (Li – Ne) yield more symmetric peaks on flatter backgrounds than the transition metals, actinides, and lanthanides.[17] Insulators usually yield simpler spectra than conductors.

XPS data analysis often involves peak fitting because the widths of XPS signals are comparable to the chemical shifts found for an element in its different oxidation states. However, much XPS peak fitting is non-trivial. Peak fitting requires the analyst to make a series of educated choices, which include the type of baseline, the number of fit components, and the positions, widths, areas, and functional forms of those fit components. In addition, the analyst must be aware of the possibility of differential charging, shake-up peaks, multiplet signals, asymmetric peaks, and overlapping signals from other photoemission, Auger, and/or valance band peaks. Thus, XPS peak fitting requires a solid understanding of the technique. At least in

part because of its complexity, much of the XPS peak fitting shown in the scientific literature is of questionable quality. Indeed, some of us recently reported that ca. 40 % of the fitted XPS data in the literature is significantly flawed.[18] Other concerns regarding XPS include inappropriate data acquisition and archiving.[19] These issues are part of the so-called reproducibility crisis in science.[20–23] Recently, a series of guides have been published to address many of these issues.[3]

A variety of statistical methods have been developed to help guide XPS peak fitting. Perhaps the simplest of these tools are (i) the sum of the fit components and (ii) a plot of the residuals to a fit. At least one of these should be shown for every peak fit. A failure to do so makes it difficult for the reader to assess the quality of a fit. The authors of this work prefer the plot of the residuals; the sum of the fit components is often close to both the original data and the synthetic fit components, which can make it difficult to judge the quality of a fit in those places. As figures of merit, fitting algorithms and analysts often rely on the standard deviation of the residuals or Chi squared (χ^2). The Abbe criterion is occasionally used to identify fits with poorly distributed residuals – it flags fits with residuals that are consistently above or below their corresponding experimental data. Monte Carlo analysis provides an estimate of the uncertainties in the fit components/synthetic peaks. [We need a ref. or two for this. Can you find one or two? Or can someone recommend 1 – 3 good references here?] [I have not been able to find an appropriate reference so far but I'm trying to find some reference/s for this][Can anyone recommend a reference or two here?] It does so by repeatedly adding noise to a denoised version of the spectrum. It then performs the fits in question, observing the uncertainties in the peaks as the noise is changed. Uniqueness plots, which were borrowed from spectroscopic ellipsometry,[24] identify fit parameter correlation.[25] Finally, we note the uses of wavelets and

the Fourier transform to denoise XPS spectra, [26, 27] and principal component analysis (PCA), multivariate curve resolution (MCR), and pattern recognition entropy (PRE) for analyzing XPS depth profiles, i.e., series of XPS spectra.[28]

Ideally, the initial positions, widths, and areas of the synthetic peaks in an XPS peak fit would have little effect on the final fit; in a robust fit, it would be reasonable to expect that, within reason, different starting conditions would lead to the same final answer/global minimum. However, this is not always the case. Local minima may be present in the parameter space associated with an XPS fit, which become apparent when different initial conditions for a fit are employed. That is, in some cases, different starting conditions may lead to significantly different fitting results. This problem would be expected in the case of overfitting. That is, when too many synthetic peaks and/or adjustable parameters are present in a fit, the results of a fit may be overly sensitive to the starting conditions. Of course, this issue is exacerbated by the fact that one does not generally know, a priori, exactly where to place the components in a fit initially or what parameters should describe them. This situation is analogous to the “disproof by counterexample” method used by mathematicians in which a single counterexample is used to disprove a proposition.[29] In other words, one might question the validity of a fitting protocol if different starting conditions in it lead to significantly different results. The method developed in this work tests for local minima in a fit space. In it, we randomly select a number of different starting conditions for a fit that are within a reasonable range. We find that when the parameter space is appropriately limited, the different starting points for the fit generally converge to the same solution. However, in the case of overfitting/too many fit parameters, the fits converge to multiple local minima. The final results of these fits can be represented with a box plot, where narrow ‘boxes’ correspond to more constrained/reasonable starting conditions for a fit and vice

versa. A box plot, which may also be referred to as a ‘box and whisker’ plot, generally consists of a box with upper, middle, and lower lines, and two horizontal lines and two points above and below the box. The middle line of a box plot corresponds to the median value of a data set. The lower and upper lines correspond to the values at the first and third quartiles, respectively. The ends of the vertical lines coming out of the boxes (the whiskers) generally give the values of the data at the 5% and 95% percentiles, and points above and below the box and whiskers correspond to the highest and lowest values in the data set. Thus, a box plot represents the distribution of the data points in a data set. We illustrate our approach to identifying fits with multiple local minima using a C 1s spectrum from a complex sample (see Figure 1a). Our interest in C 1s narrow scans is a result of a recent report that showed that they are the most frequently presented and researched narrow scans in the scientific literature,[18] and that they are often fit incorrectly.[29, 30]

Experimental

XPS Peak Fitting. Both the four- and ten-component fits in this work employed U3 Tougaard type backgrounds and LA peak shapes (LA(1.53,243)) in CasaXPS (Ver. 2.3.24rev1.1Z), i.e., all the fitting reported in this work was performed with this program.

Calculations for the four-component peak fits. We first created a 32 x 4 matrix of random numbers uniformly distributed between 0 and 1 using the following MATLAB script:

```
>> rng shuffle
```

```
>> x = rand (64,4)
```

This algorithm for generating random numbers, D.E. Knuth's shuffle algorithm, also called the Fisher-Yates algorithm,[31] provides a different seed based on the current clock time of the computer. A four-component fit was chosen because there appear to be four main signals/peaks in the C 1s spectrum in Figure 1a. For the resulting four-component fits, peaks with widths of 1.25, 1.25, 1.25, and 3.5 eV were initially placed at 285.0, 287.0, 288.5, and 291.5 eV, and, respectively. The broader (3.5 eV) peak here corresponds to the shake-up peak/signal in the data. To determine the random starting conditions for the fits, the maximum heights/areas for these peaks were first determined by increasing the height of the peaks until their maxima touched the data envelope. These maximal positions were then multiplied by random numbers (uniformly distributed between zero and one) to obtain the starting conditions for the fits. That is, each of the four random values in each row of the 32 x 4 matrix was multiplied by the maximum areas of the four peaks to create a starting condition for a fit. After entering these starting conditions for the synthetic peaks into CasaXPS, the experimental data were fit. In this fitting, chi squared, not 'RMS', was used as the figure of merit to determine the quality of the fit. The fit button in the software was hit repeatedly until the fit converged. This was mostly done as a precaution – little change in the fit or accompanying figure of merit took place after the first iteration.

Calculations for the ten-component peak fits. Three different ten-component fits were performed. These fits had peaks initially positioned at 284.5, 285.2, 285.9, 286.6, 287.3, 288, 288.7, 289.4, 290.1, and 291.6 eV with initial widths of 1.25 eV, except for the shake-up peak positioned at 290.8, which had a width of 3.5 eV. The initial areas of these peaks were determined randomly in the same way as for the four-component fit. That is, the initial areas for a fit were based on a row of a 32 x 10 matrix of random numbers uniformly distributed between 0 and 1 generated in MATLAB (see above). This ten-component fit was then performed three different ways. First, all

the fitting parameters were allowed to vary with essentially no constraints. That is, the widths and peak positions of all the peaks were allowed to vary over a wide range (from 0.01 to 5 eV, and from 282 to 295 eV, respectively), and there were no other constraints on this fit. Second, all of the peaks, except the highest binding energy (shake-up) peak, had the same initial width (1.25 eV), while the width of the highest binding energy peak was initially set to be 3.5 eV. The third fit involved fixing the peak widths, as in the second approach, while also fixing the peak positions at their initial values.

Uniqueness Plots. Uniqueness plots were generated to identify fit parameter correlation in the fits.[25] In a uniqueness plot, the data are first fit using a specific protocol. The quality of this fit is then probed by fixing a parameter in it, e.g., the position of a peak, and performing the fit again, as it was originally, without allowing that parameter to vary. If this parameter can be varied over a range of values without the overall quality of the fit changing, the fit is not unique, i.e., the fit contains so many fit parameters that it can compensate for the changes in the parameter in question. In contrast, if the error in the fit increases when the parameter is varied the fit is said to be unique.

Monte Carlo Analysis. Monte Carlo analyses were performed in CasaXPS. [Neal, would you like to add any extra details here?]

Results and Discussion

The C 1s narrow scans of different materials range from being very simple to quite complex. For example, C 1s narrow scans of hydrocarbon polymers may only show one chemical state (one peak) and simple baselines, while those of partially oxidized materials that contain both sp^2 (graphitic) and sp^3 (aliphatic) carbon may show two signals for reduced carbon,

multiple peaks corresponding to oxidized carbon, one or more asymmetric signals, multiple overlapping peaks, one or more shake-up peaks, and more complicated baselines. In this study, we chose a moderately complex C 1s spectrum that has previously been fit and discussed in the literature.[29] It consists of a mixture of disodium cromoglycate (DSCG) (48 mol%) and valine (52 mol%). The C 1s spectrum of pure DSCG was previously fit with six components, while that of valine was fit with three. As shown in Figure 1a, this spectrum has three obvious, larger peaks, a smaller shake-up peak, and a small rise in its baseline. Accordingly, to test the concept of disproof by contraction in XPS peak fitting and demonstrate the use of box plots to show the results of different initial conditions for a fit, we first fit this spectrum in a rather restrained fashion with four synthetic components and a U3 Tougaard background, as illustrated in Figure 1b. We then fit it in a less restrained fashion with ten components and the same background type, as illustrated in Figure 1c.

Figure 2a shows a box plot created from a series of random starting conditions used for the four-component fit of Figure 1a – b. Here, we randomly chose four components for the fit 32 times. Each of these components had an intensity that varied uniformly/was assigned randomly between null (at the baseline) and the height of the peak envelope above it. The variation in the heights of the boxes in Figure 2a (C1> C2> C3> Sat) is a result of the differences in the height of the C 1s spectral envelope above the baseline at each peak position. Figure 2b shows the box plot generated after the fits. The much narrower box plots (compared to Figure 2a) reveal that essentially the same fit was obtained from all the starting conditions. A repeat of this experiment with another 32 x 4 array of random numbers yielded the same result. Thus, the box plot in Figure 2b reveals that the space associated with this four-component fit does not have a large number of local minima in it, i.e., almost all the starting conditions that were explored led to the

same result. As an aside, in our early exploration of this data set, we found one set of unstable, random starting conditions that leads to a less reasonable and different result. These initial conditions caused the shake-up signal to move into the main peak envelope (see Figure 2c). This fit had a residual standard deviation of 4.7, while the others had residual standard deviations of approximately unity.

A large number of components/parameters in a fit can lead to fit parameter correlation, i.e., the fit can have little statistical meaning. An additional, related problem with a large number of fit components is the creation of a space in which there are multiple local minima. The validity of such a fit may be questioned if different initial conditions for it lead to significantly different results. To explore the ability of box plots to identify fit spaces in which there are multiple local minima, we fitted the data in Figure 1a with 10 components (see Figure 1c). As described in the Experimental, the main peak envelope of this C 1s spectrum was fit with nine peaks and another, broader signal at higher binding energy was used to model the shake-up signal. Initially, these peaks had equal widths and were evenly spaced. A series of 32 random starting conditions was then generated in the same manner as for the four-component fit. Figure 3a shows the box plot of these starting conditions, where, again, the sizes of these boxes reflect the height of the data envelope above the baseline at the different peak positions. The C 1s spectrum in Figure 1a (see also Figure 1c) was then fit with these 32 random starting conditions, and Figure 3b shows the resulting box plot. Unlike the box plot obtained after the unconstrained four-component fit (Figure 2b), which showed almost identical outcomes for the fits, this box plot shows a considerable amount of variation in the results. That is, the box plot reveals that the fit space for this unconstrained ten-component fit contains multiple local minima.

Fit parameters in XPS peak fitting are often constrained to prevent unreasonable (unchemical and/or unphysical) results. For example, it is not uncommon to constrain the areas of spin-orbit components to their theoretical ratio. Two approaches were taken to constrain the ten-component fit shown in Figure 1c. In the first, all of the peaks, except the one modeling the shake-up signal, were constrained to have the same width. While this restriction should increase the likelihood that physically reasonable results will be obtained (widely varying peak widths in an XPS peak fit are often a sign of poor fitting[18], the box plot summarizing the results of performing this fit with 32 different/random initial conditions (Figure 3c), suggests that these constraints do not lead to a robust fit. That is, the widths of the boxes in Figure 3c indicate that multiple, local minima are present in this fit space. This fit still appears to have too many adjustable parameters. Our second approach was to fix the positions of the fit components/synthetic peaks to remain at their starting values and to also keep the widths of the nine peaks describing the main peak envelope equal to each other. The results for the 32 random conditions/starting points shown in Figure 3d indicate that there is now a single global minimum for this approach. Interestingly, the areas of the peaks at the global minimum are often fairly far from their average starting positions in Figure 3a.

While the narrow boxes in Figures 2b and 3d indicate that the corresponding fit spaces primarily consist of a single global minimum, we do not claim that the corresponding fits are ‘good’, by which we mean that they appropriately represent the physics and chemistry of the sample. We only claim that they pass ‘the box plot test’ described herein, and that they *may* be good fits, i.e., a failure to pass this test would call into question the fit protocol. To further explore this idea, we turn to Figure 4, which contains representative fits from the four approaches taken herein. Figure 4a shows a representative four-component fit of the data in

Figure 1a. In many ways it is a very good fit. The three main peaks seem to effectively represent the data. They have similar widths. The broader shake-up signal appropriately models the higher binding energy portion of the spectrum. A major ‘plus’ of this fit is that there were no constraints on these fit components. The only problem with this fit is the moderately high residual standard deviation and accompanying structure in the residuals. Nevertheless, without a priori knowledge about the sample, we would consider this fit to be quite reasonable. In many ways, the fit shown in Figure 4b of an unconstrained, ten-component fit to the data seems very good. The peak widths are quite similar, and the residual standard deviation is very low. However, the corresponding box plot of this analysis/approach reveals the presence of multiple local minima in this fit space, which calls into question the validity of this fit. Figure 4c shows the ten-component fit in which the peaks that modeled the main peak envelope were constrained to have the same widths. The same conclusions we drew from Figure 4b apply here – the fit has reasonable looking peaks and a low residual standard deviation, but multiple local minima in the fit space. Figure 4d shows the ten-component fit in which the peak positions were fixed. While the widths of the fit components are similar, there are problems with this approach – the residual standard deviation of this fit is high, and there is significant structure in the residuals. Interestingly, the box plot associated with this fit (Figure 3d) suggests that there is a single global minimum for this fit. It appears that increasing/strengthening the number of constraints on this fit reduced the number of local minima available for it. Nevertheless, in spite of the much-improved box plot here, we conclude that this fitting approach is flawed. The fit components do not align well with the structure in the data envelope, the residual standard deviation is too high, and there is too much structure in the residuals. We learn from this exercise that more than one type of analysis and metric can be helpful in determining the quality of a fit. Overall, without prior knowledge of

the sample, we would consider the four-component fit shown in Figure 4a to be the best of the fits in Figure 4, even with its somewhat higher residual standard deviation.

Because we are proposing a new statistical tool for evaluating the quality of a peak fits/peak fitting protocols in XPS, we believe it is important to compare the results obtained from box plots with those obtained from two other statistical methods that are also used to evaluate the quality of XPS peak fitting: Monte Carlo analysis and uniqueness plots. Figure 5 shows the Monte Carlo results of representative fits of the four approaches considered herein. That is, Figures 5a, 5b, 5c, and 5d show the Monte Carlo analyses of the unconstrained four-component fit (see also Figures 1b, 2a, 2b, and 4a), the unconstrained ten-component fit (see also Figures 1c, 3a, 3b, and 4b), the ten-component fit with constrained (locked) peak widths (see also Figures 1c, 3a, 3c, and 4c), and the ten-component fit with constrained (locked) peak positions (see also Figures 1c, 3a, 3d, and 4d), respectively. The most important, obvious result here is that there is greater uncertainty in the peak positions in Figures 5b and 5c, which came from the ten-component fits with no or weaker constraints on the peaks, as evidenced by the wider bands of the different fit components in the Monte Carlo analysis. In contrast, Figures 5a and 5d, which came from the four-component fit and the ten-component fit in which the peak positions were fixed to their initial positions, respectively, show narrow bands (the least uncertainty in the Monte Carlo results). These results do not appear to be coincidental. That is, the greatest uncertainty in peak positions/areas indicated by the Monte Carlo analysis (Figures 5b and 5c) come from fitting protocols with larger numbers of weakly constrained peaks. Figure 6 shows the use of box plots to represent the Monte Carlo results in Figure 5. As expected, the box plots of the areas of the different fit components in the Monte Carlo analyses/simulations are widest for Figures 5b and 5c, and narrowest for Figures 5a and 5d.

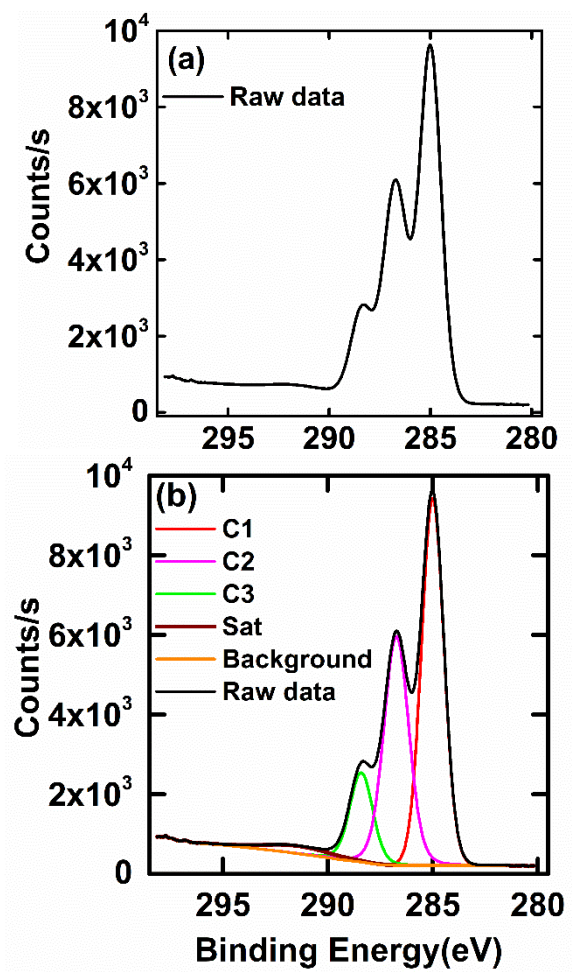
Figure 7 shows uniqueness plots for the fits shown in Figure 4. These plots were generated by systematically changing and fixing the position of one of the peaks in the fit, while allowing the remaining parameters in the fit to change as before. In Figure 7a, we see the uniqueness plot for the unconstrained four-component fit. Its ‘U/V’ shape suggests that at least the parameter that was altered is statistically relevant. That is, the other fit components are unable to compensate for changes made to this parameter. Figure 7b shows the uniqueness plot for the fit with ten unconstrained components. The ‘horizontal line’ generated here indicates that the other parameters can compensate for a change in particular peak position. That is, these parameters are correlated and the fit is not unique. Figures 7c and 7d similarly show the uniqueness plots for the ten-component fits in which the widths of the peaks were constrained to be the same and their positions were fixed, respectively. The first of these two plots (Figure 7c) indicates fit parameter correlation, while the second (Figure 7d) suggests that the fit parameters have statistical significance.

The results from the box plots, Monte Carlo analyses, and uniqueness plots in this work show significant similarities. That is, the higher quality fits, which had either fewer components or more constraints, showed box plots with narrow boxes, little variation in the peaks in the Monte Carlo analyses, and uniqueness plots with ‘U/V’ shapes. In contrast, the lower quality fits, which had more components and fewer constraints, showed box plots with wider boxes, more variation in the peaks in the Monte Carlo analyses, and uniqueness plots that appeared as horizontal lines. These results suggest that both poor peak fitting and more statistically sound peak fitting can be identified by any of these plots. Nevertheless, we emphasize that these different plots show different things. Box plots identify the presence of multiple local minima in a fit space. Monte Carlo analysis provides a measure of the errors of the parameters of a fit, and

uniqueness plots identify fit parameter correlation. Note that in some cases it is advantageous to perform uniqueness plots on more than one of the fit components in a fit – some fit components may have greater uniqueness (be less correlated with the remaining parameters) than others.

Conclusions

We have demonstrated the use of box plots for displaying the results of multiple, random fits to an XPS narrow scan. If the resulting boxes are narrow, the fitting space is believed to contain, or at least be primarily described, by a single global minimum. Conversely, if the boxes are wide, multiple local minima will be present, and the fitting protocol may be suspect.



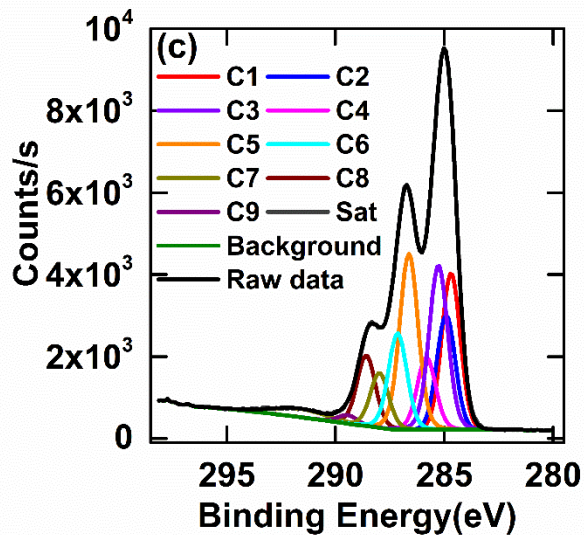
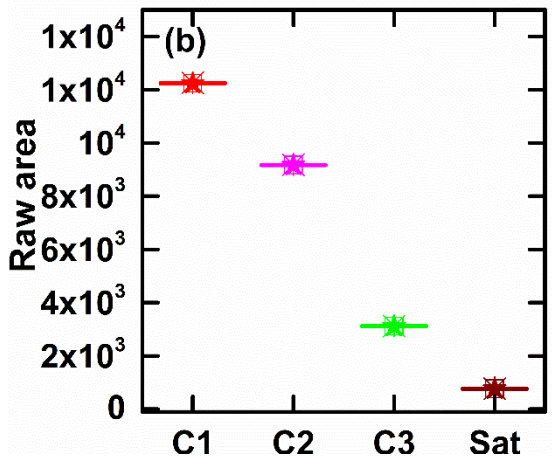
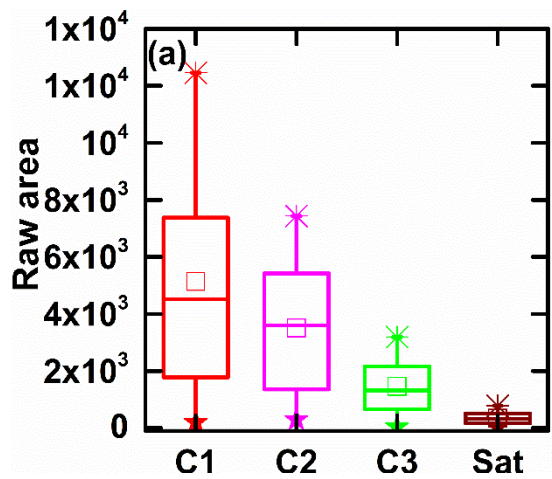


Figure 1. a) C 1s spectrum of a mixture of 48 mol% DSCG – 52 mol% valine,[29] b) a four-component fit to this spectrum, and c) a ten-component fit to



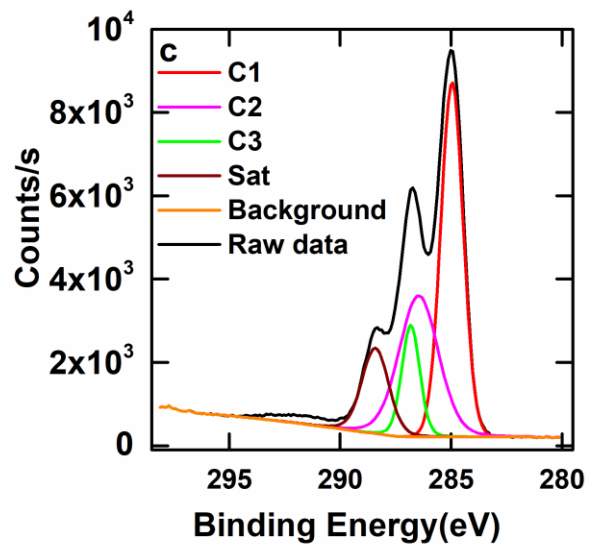


Figure 2. a) Box plot of the initial conditions for 32 four-component fits of the data in Figure 1a (see also Figure 1b), with the conditions for each fit chosen randomly, b) Box plot for the results of the 32 four-component fits represented in Figure 2a, and c) Results of a set of random conditions that led to a poor four-component fit to the data.

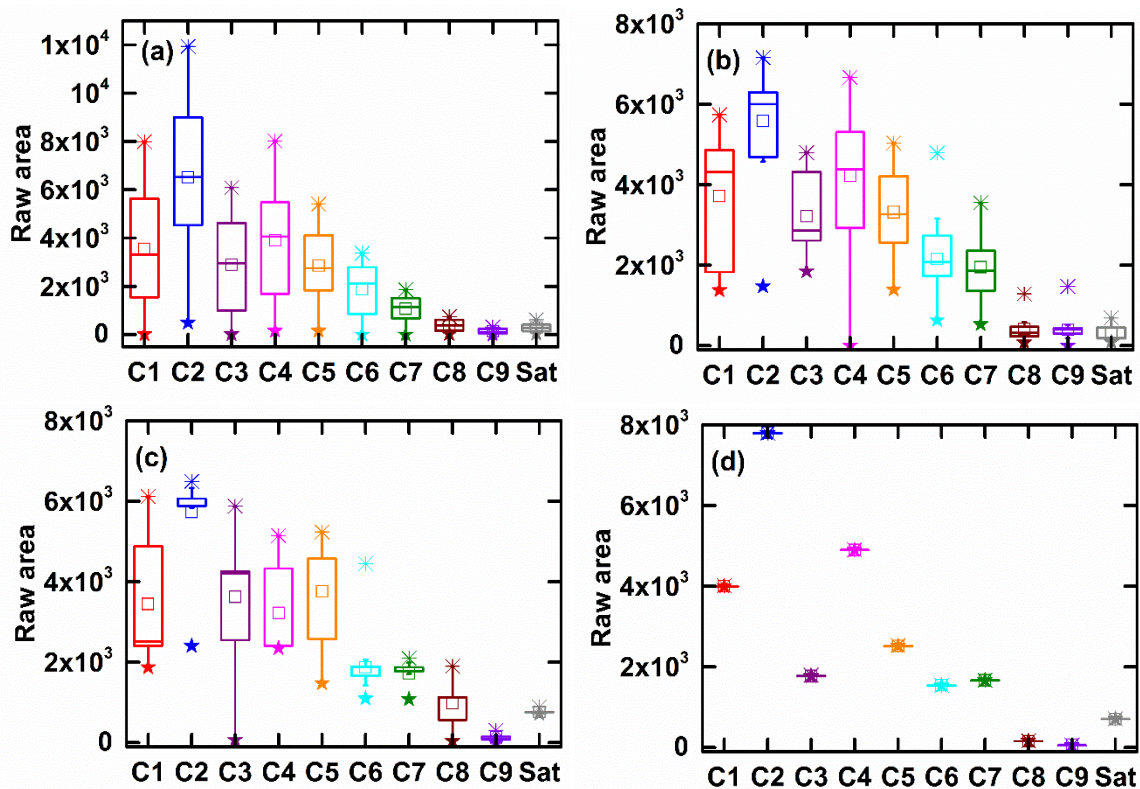


Figure 3. Box plots of the peaks areas of three 10-components fits to the C 1s spectrum in Figure 1a. Box plots: a) Of the random starting conditions for all the fits, b) After fitting with unconstrained peaks, c) After fitting where the only significant constraint was that the peaks modeling the main peak envelope had the same widths, and d) After fitting where the only significant constraints were that all the peaks positions were fixed to their original positions and the widths of the nine peaks describing the main peak envelope were required to be the same.

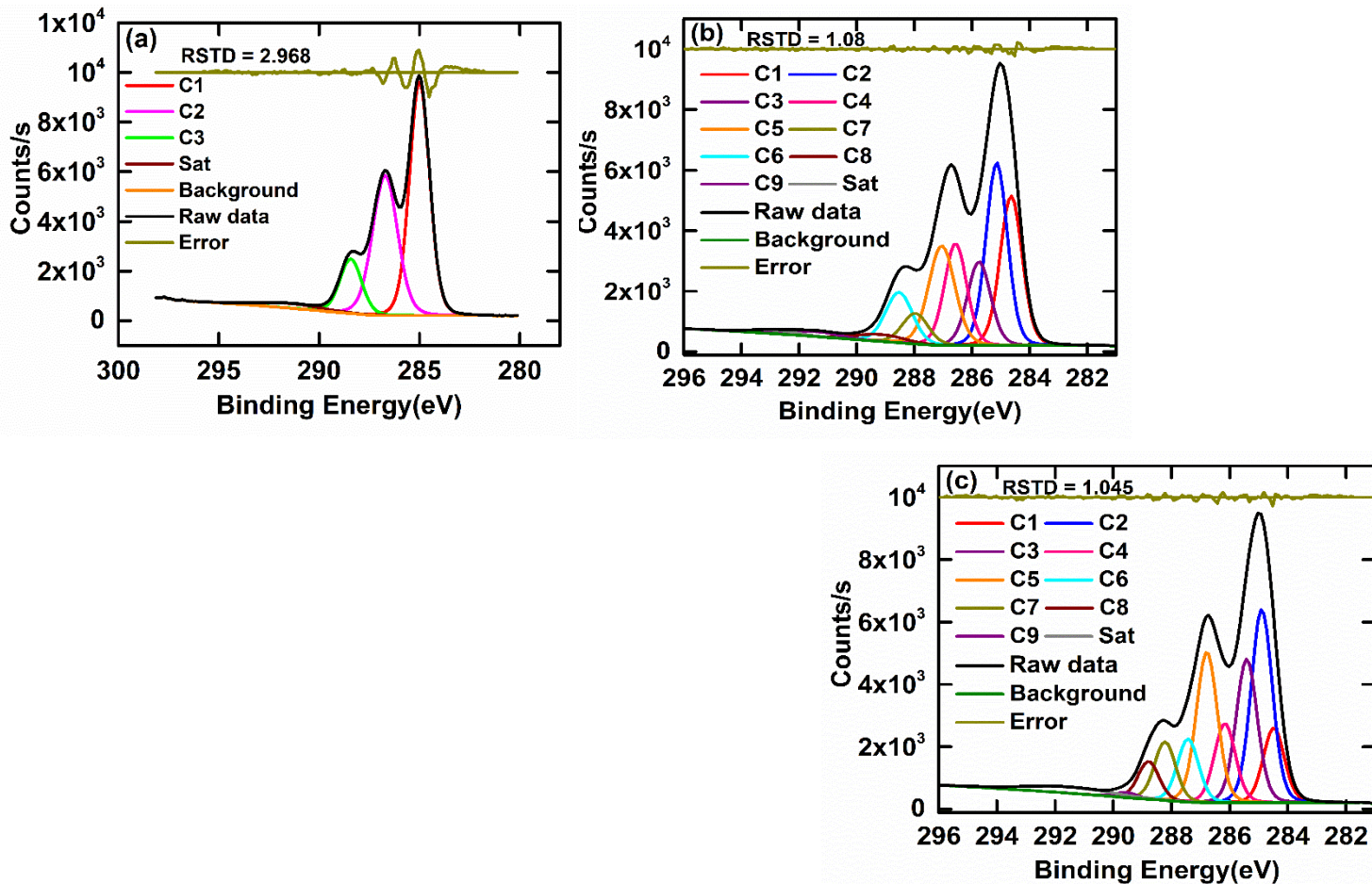


Figure 4. Representative fits of the C 1s spectrum in Figure 1a obtained using the methods described in this work. Please see the text and Figures 1 – 3 for more details. Fits containing: a) Four unconstrained components, b) Ten unconstrained components, c) Ten components in which the fit components modeling the main peak envelope were constrained to have the same widths, d) Ten components in which the peaks were constrained to keep their initial positions, and the nine peaks describing the main peak envelope were constrained to have the same widths.

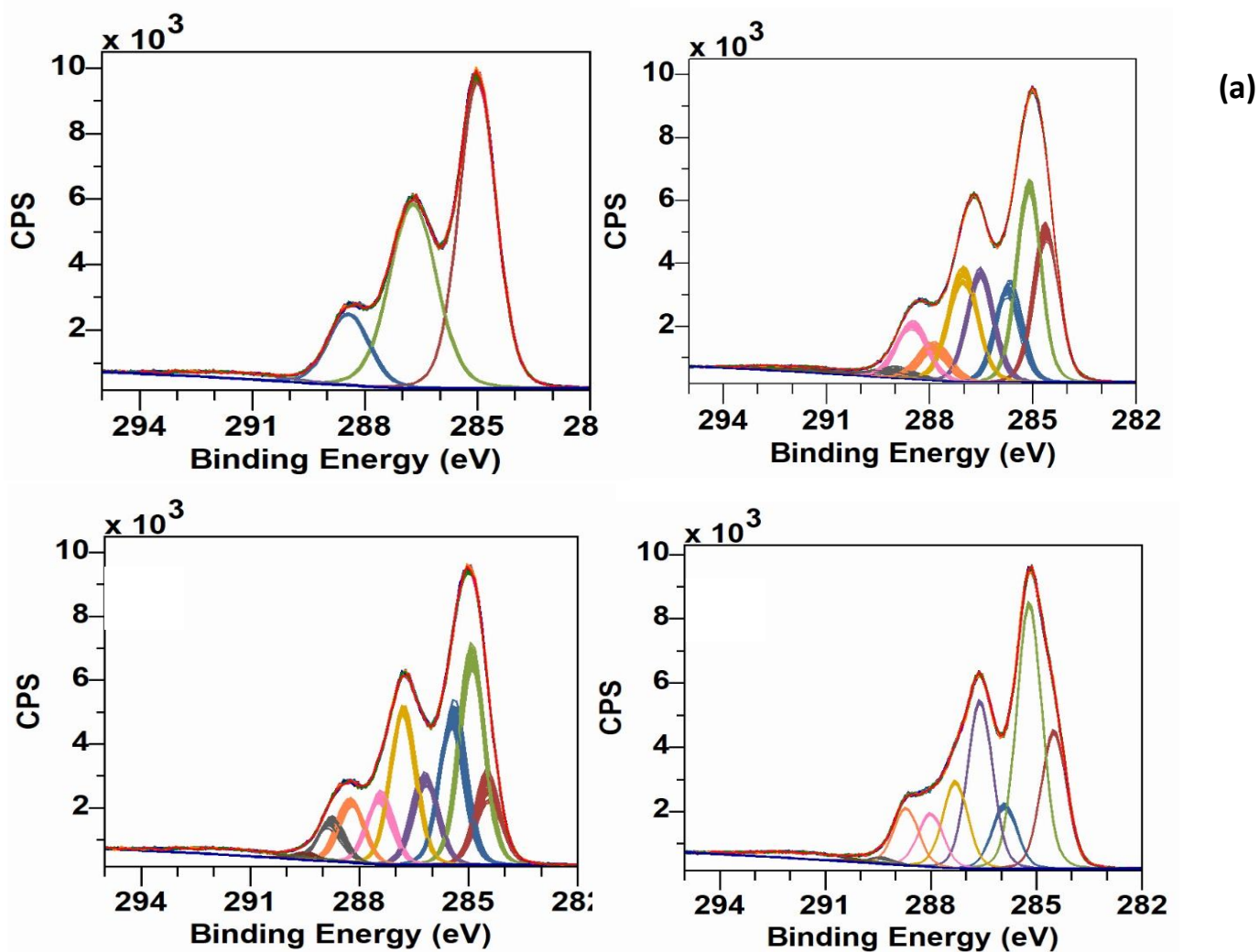


Figure 5. Monte Carlo analysis of the fits to the C 1s narrow scan shown in Figure 4. That is, the panels in Figure 4 correspond to those in this figure. Monte Carlo analysis of a) An unconstrained four-component fit, b) A 10-component fit with no constraints on the fitting parameters, c) A 10-component fit in which the only constraint was that the nine peaks modeling the main peak envelope have the same widths, and d) A 10-component fit in which the peaks

were fixed at their initial positions, and the nine peaks describing the main peak envelope were constrained to have the same widths.

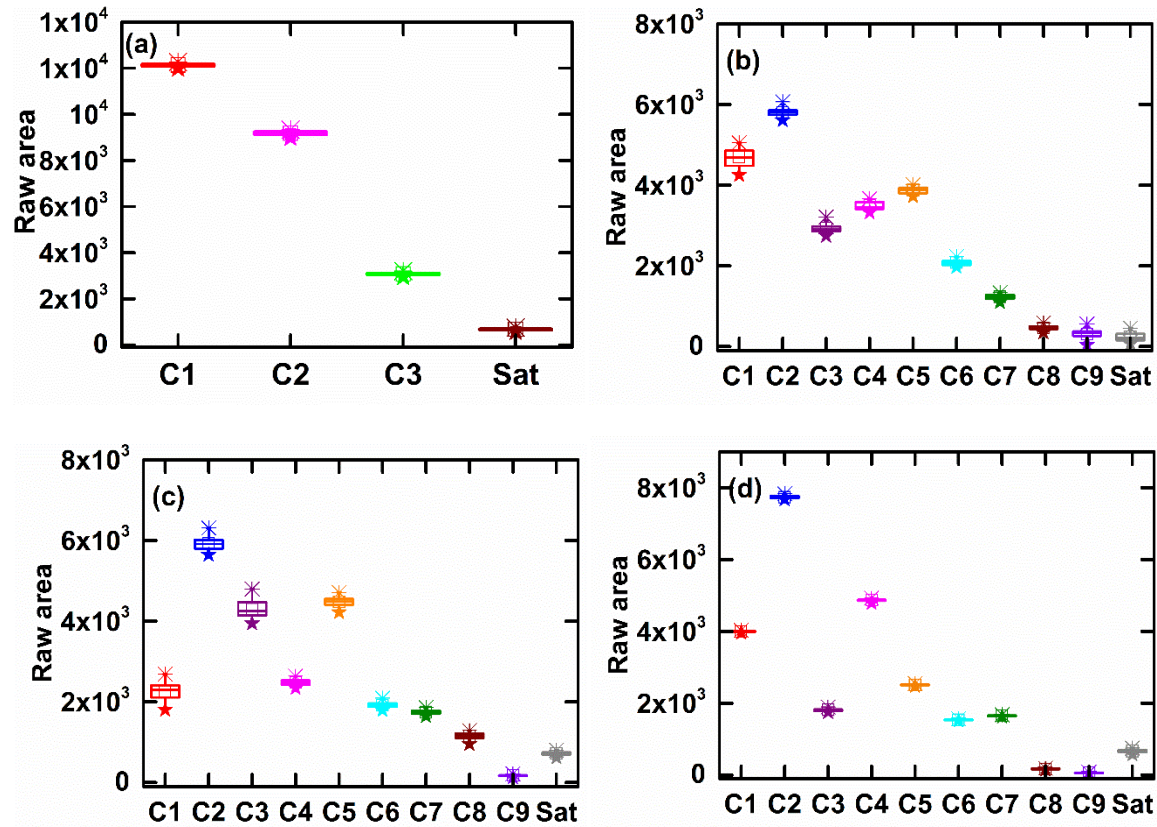


Figure 6. Box plots of the Monte Carlo simulations shown in Figure 5, i.e., of a) An unconstrained four-component fit, b) A 10-component fit with no constraints on the fitting parameters, c) A 10-component fit in which the only constraint was that the nine peaks modeling the main peak envelope have the same widths, and d) A 10-component fit in which the peaks were fixed at their initial positions, and the nine peaks describing the main peak envelope were constrained to have the same widths.

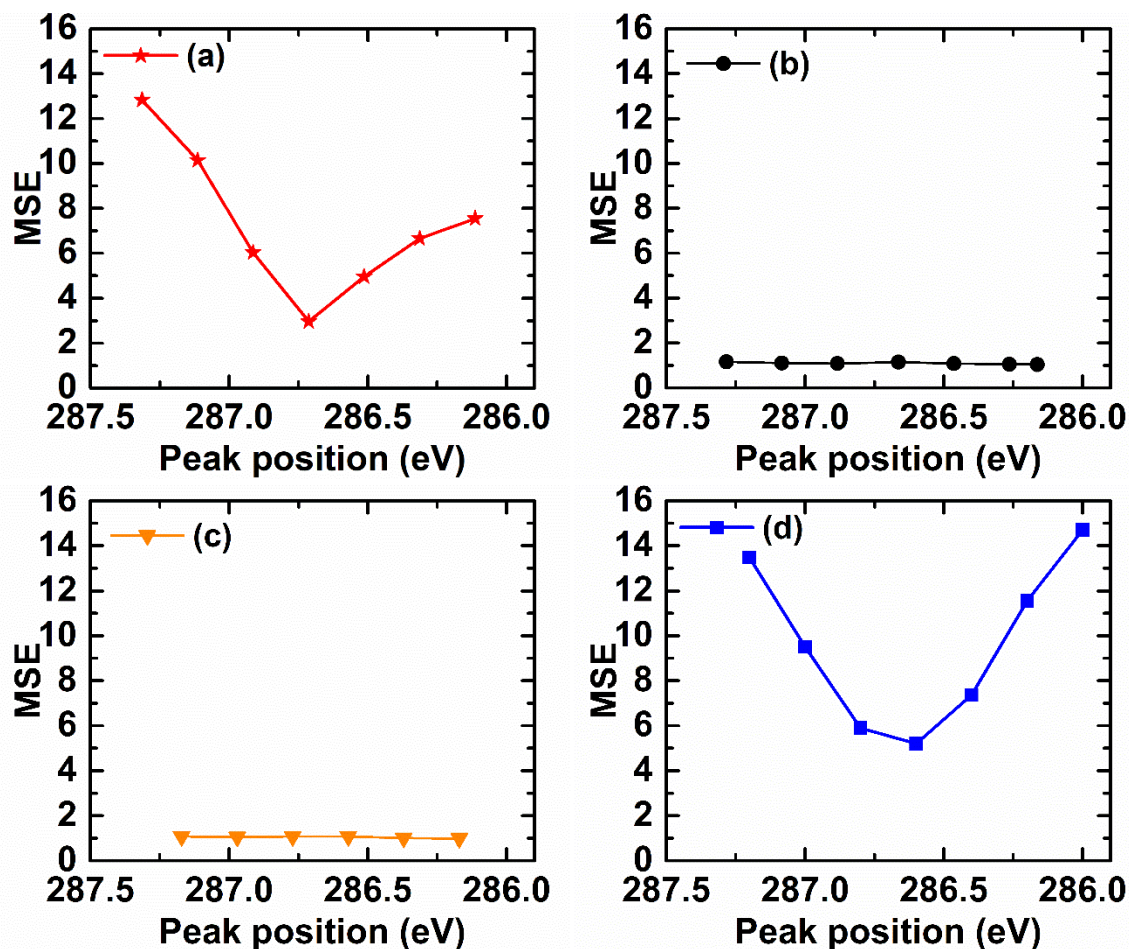


Figure 7. Uniqueness plots of the fits in Figure 4, i.e., of a) An unconstrained four-component fit, b) A 10-component fit with no constraints on the fitting parameters, c) A 10-component fit in which the only constraint was that the nine peaks modeling the main peak envelope have the same widths, and d) A 10-component fit in which the peaks were fixed at their initial positions, and the nine peaks describing the main peak envelope were constrained to have the same widths. [Ben, please provide a little more information here. That is, specify which peak was chosen for each uniqueness plot.]

References [Ben, the last reference doesn't look complete here.]

1. Heide P Van der (2011) X-Ray photoelectron spectroscopy. *An Introduction to Principles and Practices*,
2. Hofmann S (2012) Auger-and X-ray photoelectron spectroscopy in materials science: a user-oriented guide. 49
3. Baer DR, McGuire GE, Artyushkova K, Easton CD, Engelhard MH, Shard AG (2021) Introduction to topical collection: Reproducibility challenges and solutions with a focus on guides to XPS analysis. *Journal of Vacuum Science & Technology A*, 39(2):21601. <https://doi.org/10.1116/6.0000873>
4. Stevie FA, Donley CL (2020) Introduction to x-ray photoelectron spectroscopy. *Journal of Vacuum Science & Technology A*, 38(6):63204. <https://doi.org/10.1116/6.0000412>
5. Baer DR, Artyushkova K, Richard Brundle C, Castle JE, Engelhard MH, Gaskell KJ, Grant JT, Haasch RT, Linford MR, Powell CJ, Shard AG, Sherwood PMA, Smentkowski VS (2019) Practical guides for x-ray photoelectron spectroscopy: First steps in planning, conducting, and reporting XPS measurements. *Journal of Vacuum Science & Technology A*, 37(3):31401. <https://doi.org/10.1116/1.5065501>
6. Baer DR, Artyushkova K, Richard Brundle C, Castle JE, Engelhard MH, Gaskell KJ, Grant JT, Haasch RT, Linford MR, Powell CJ, Shard AG, Sherwood PMA, Smentkowski VS (2020) Erratum: "Practical guides for x-ray photoelectron spectroscopy: First steps in planning, conducting, and reporting XPS measurements" [J. Vac. Sci. Technol. A 37, 031401 (2019)]. *Journal of Vacuum Science & Technology A*, 39(1):17003. <https://doi.org/10.1116/6.0000822>
7. Gupta V, Ganegoda H, Engelhard MH, Terry J, Linford MR (2014) Assigning Oxidation States to Organic Compounds via Predictions from X-ray Photoelectron Spectroscopy: A Discussion of Approaches and Recommended Improvements. *Journal of Chemical Education*, 91(2):232–238. <https://doi.org/10.1021/ed400401c>
8. Tougaard S (2018) Improved XPS analysis by visual inspection of the survey spectrum. *Surface and Interface Analysis*, 50(6):657–666. <https://doi.org/https://doi.org/10.1002/sia.6456>
9. Chambers SA, Wang L, Baer DR (2020) Introductory guide to the application of XPS to epitaxial films and heterostructures. *Journal of Vacuum Science & Technology A*, 38(6):61201. <https://doi.org/10.1116/6.0000465>
10. Sønsteby HH, Yanguas-Gil A, Elam JW (2020) Consistency and reproducibility in atomic layer deposition. *Journal of Vacuum Science & Technology A*, 38(2):20804. <https://doi.org/10.1116/1.5140603>
11. Davies PR, Morgan DJ (2020) Practical guide for x-ray photoelectron spectroscopy: Applications to the study of catalysts. *Journal of Vacuum Science & Technology A*, 38(3):33204. <https://doi.org/10.1116/1.5140747>
12. Conard T, Vanleenhove A, Heide P van der (2020) Achieving reproducible data: Examples

from surface analysis in semiconductor technology. *Journal of Vacuum Science & Technology A*, 38(3):33206. <https://doi.org/10.1116/1.5140746>

13. Baer DR (2020) Guide to making XPS measurements on nanoparticles. *Journal of Vacuum Science & Technology A*, 38(3):31201. <https://doi.org/10.1116/1.5141419>

14. Easton CD, Kinnear C, McArthur SL, Gengenbach TR (2020) Practical guides for x-ray photoelectron spectroscopy: Analysis of polymers. *Journal of Vacuum Science & Technology A*, 38(2):23207. <https://doi.org/10.1116/1.5140587>

15. Shah D, Patel DI, Roychowdhury T, Rayner GB, O'Toole N, Baer DR, Linford MR (2018) Tutorial on interpreting x-ray photoelectron spectroscopy survey spectra: Questions and answers on spectra from the atomic layer deposition of Al₂O₃ on silicon. *Journal of Vacuum Science & Technology B*, 36(6):62902. <https://doi.org/10.1116/1.5043297>

16. Major GH, Fairley N, Sherwood PMA, Linford MR, Terry J, Fernandez V, Artyushkova K (2020) Practical guide for curve fitting in x-ray photoelectron spectroscopy. *Journal of Vacuum Science & Technology A*, 38(6):61203. <https://doi.org/10.1116/6.0000377>

17. Brundle CR, Crist BV (2020) X-ray photoelectron spectroscopy: A perspective on quantitation accuracy for composition analysis of homogeneous materials. *Journal of Vacuum Science & Technology A*, 38(4):41001. <https://doi.org/10.1116/1.5143897>

18. Major GH, Avval TG, Moeini B, Pinto G, Shah D, Jain V, Carver V, Skinner W, Gengenbach TR, Easton CD, Herrera-Gomez A, Nunney TS, Baer DR, Linford MR (2020) Assessment of the frequency and nature of erroneous x-ray photoelectron spectroscopy analyses in the scientific literature. *Journal of Vacuum Science & Technology A*, 38(6):61204. <https://doi.org/10.1116/6.0000685>

19. Suzuki M, Nagao H, Harada Y, Shinotsuka H, Watanabe K, Sasaki A, Matsuda A, Kimoto K, Yoshikawa H (2020) Raw-to-repository characterization data conversion for repeatable, replicable, and reproducible measurements. *Journal of Vacuum Science & Technology A*, 38(2):23204. <https://doi.org/10.1116/1.5128408>

20. Park J, Howe JD, Sholl DS (2017) How Reproducible Are Isotherm Measurements in Metal–Organic Frameworks? *Chemistry of Materials*, 29(24):10487–10495. <https://doi.org/10.1021/acs.chemmater.7b04287>

21. Baer DR, Gilmore IS (2018) Responding to the growing issue of research reproducibility. *Journal of Vacuum Science & Technology A*, 36(6):68502. <https://doi.org/10.1116/1.5049141>

22. Baker M (2016) 1,500 scientists lift the lid on reproducibility. *Nature News*, 533(7604):452–454.

23. Engineering M, on Behavioral B, National Academies of Sciences and Medicine E (2019) Reproducibility and Replicability in Science.

24. N. Hilfiker J, Bungay CL, Synowicki RA, Tiwald TE, Herzinger CM, Johs B, Pribil GK, Woollam JA (2003) Progress in spectroscopic ellipsometry: Applications from vacuum ultraviolet to

infrared. *Journal of Vacuum Science & Technology A*, 21(4):1103–1108.
<https://doi.org/10.1116/1.1569928>

25. Singh B, Diwan A, Jain V, Herrera-Gomez A, Terry J, Linford MR (2016) Uniqueness plots: A simple graphical tool for identifying poor peak fits in X-ray photoelectron spectroscopy. *Applied Surface Science*, 387:155–162. <https://doi.org/https://doi.org/10.1016/j.apsusc.2016.06.070>

26. Charles C, Leclerc G, Pireaux J-J, Rassin J-P (2004) Introduction to wavelet applications in surface spectroscopies. *Surface and Interface Analysis*, 36(1):49–60.
<https://doi.org/https://doi.org/10.1002/sia.1648>

27. Charles C, Leclerc G, Louette P, Rassin J-P, Pireaux J-J (2004) Noise filtering and deconvolution of XPS data by wavelets and Fourier transform. *Surface and Interface Analysis*, 36(1):71–80. <https://doi.org/https://doi.org/10.1002/sia.1650>

28. Chatterjee S, Singh B, Diwan A, Lee ZR, Engelhard MH, Terry J, Tolley HD, Gallagher NB, Linford MR (2018) A perspective on two chemometrics tools: PCA and MCR, and introduction of a new one: Pattern recognition entropy (PRE), as applied to XPS and ToF-SIMS depth profiles of organic and inorganic materials. *Applied Surface Science*, 433:994–1017.
<https://doi.org/https://doi.org/10.1016/j.apsusc.2017.09.210>

29. Gengenbach TR, Major GH, Linford MR, Easton CD (2021) Practical guides for x-ray photoelectron spectroscopy (XPS): Interpreting the carbon 1s spectrum. *Journal of Vacuum Science & Technology A*, 39(1):13204. <https://doi.org/10.1116/6.0000682>

30. Linford MR, Smentkowski VS, Grant JT, Brundle CR, Sherwood PMA, Biesinger MC, Terry J, Artyushkova K, Herrera-Gómez A, Tougaard S, Skinner W, Pireaux J-J, McConville CF, Easton CD, Gengenbach TR, Major GH, Dietrich P, Thissen A, Engelhard M, Powell CJ, Gaskell KJ, Baer DR (2020) Proliferation of Faulty Materials Data Analysis in the Literature. *Microscopy and Microanalysis*, 26(1):1–2. <https://doi.org/DOI: 10.1017/S1431927619015332>

31. Fisher RA, Yates F (1938) Statistical tables: For biological, agricultural and medical research.