HAL
open science

Guide to XPS data analysis: Applying appropriate constraints to synthetic peaks in XPS peak fitting

George Major, Vincent Fernandez, Neal Fairley, Emily Smith, Matthew Linford

To cite this version:

George Major, Vincent Fernandez, Neal Fairley, Emily Smith, Matthew Linford. Guide to XPS data analysis: Applying appropriate constraints to synthetic peaks in XPS peak fitting. Journal of Vacuum Science & Technology A, 2022, 40 (6), pp.063201. 10.1116/6.0001975 . hal-03850951

HAL Id: hal-03850951

https://hal.science/hal-03850951v1

Submitted on 8 Mar 2023

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.

Guide to XPS Data Analysis: Applying Appropriate Constraints to Synthetic Peaks in XPS Peak Fitting

George H. Major,[1] Vincent Fernandez,[2] Neal Fairley,[3] Emily F. Smith,[4] Matthew R. Linford[1]

[1]Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT 84602
[2]Institut des Matériaux Jean Rouxel - Sciences, Université of Nantes, Nantes, France
[3]Casa Software Ltd., Bay House, Teignmouth, UK
[4]School of Chemistry, University of Nottingham, Nottingham, UK

Abstract

Peak fitting of X-ray photoelectron spectroscopy (XPS) data is the primary method for determining the identities and quantities of the chemical states of the atoms near the surface of a sample. Peak fitting is typically based on the minimization of a figure-of-merit, such as the residual standard deviation (RSD). Here we show that optimal XPS peak fitting is obtained when the peak shape (the synthetic mathematical function that represents the chemical states of the material) best matches the physics and chemistry of the underlying data. However, because this ideal peak shape is often unknown, constraints on the components of a fit are usually necessary to obtain good fits to data. These constraints may include fixing the relative full width at half maxima (FWHM) (peak widths), area ratios, and/or the relative positions of the fit components. As shown in multiple examples, while unconstrainted, less-than-optimal peak shapes may produce lower RSDs, they often lead to incorrect results. Thus, the 'suboptimal' results (somewhat higher RSDs) that are obtained when constraints are applied to less-than-perfect peak shapes are often preferable because they prevent a fit from yielding unphysical or unchemical results. XPS peak fitting is best performed when all the information available about a sample is used, including its expected chemical and physical composition, information from other XPS narrow and survey scans from the same material, and information from other analytical techniques.

Introduction

X-ray photoelectron spectroscopy (XPS) is an indispensable tool for understanding surfaces and materials, and its use has greatly increased during the past two decades.[1] XPS is based on the photoelectric effect, where X-radiation leads to the ejection of core electrons. This photoemission generally consists of a series of sharp signals on broad backgrounds that are the result of inelastically scattered electrons. Other peaks/features are also commonly observed in XPS spectra, including Auger and valence band signals. The positions of photoemission peaks are used to identify the elements at a surface and provide information about their oxidation states. The Lorentzian function is the theoretical peak shape for photoemission. However, the sample, instrument, and source of X-rays usually introduce some Gaussian character into it. Thus, XPS signals are often modelled using Voigt functions, which are convolutions of Lorentzian and Gaussian signals – the Voigt function is generally considered to be the most appropriate peak shape for fitting XPS spectra. In special cases, e.g., with conducting materials, asymmetry may need to be introduced into peak shapes.[2] The Gaussian, Lorentzian, and Voigt functions all possess 'bell' shapes. Nevertheless, these functions, and those derived from them, often differ significantly at their edges/wings.[3-6]

The successful use of bell-shaped curves in XPS to interpret and model the specific chemical states of a material depends on two overarching concepts. The first is the extent to which signals from different chemical states overlap. This is a limiting factor in successful peak fitting. The second is the choice of an appropriate model. In general, an XPS model consists of (i) a background, (ii) one or more synthetic fit components (bell-shaped peaks), including parameters that give their widths, positions, and areas, and (iii) constraints that are applied to the different fit components. In this work, we emphasize this third point. The case studies presented below provide examples that demonstrate both the use of constraints as well as some of the limitations of optimisation in XPS peak fitting. These examples show that peak fitting via minimization of a figure-of-merit can be very successful. However, only under certain conditions can the results obtained by optimization be deemed accurate/chemically meaningful. Furthermore, while the minimization of a figure-of-merit is desirable in peak fitting, it is not a necessary condition for the accuracy of a successful chemical state analysis. This assertion is demonstrated below by the peak fitting of spectra of known composition.

In addition to the Voigt function, two pseudo-Voigt functions: the Gaussian-Lorentzian sum and product functions, are commonly used in XPS peak fitting. As their names would suggest, the Gaussian-Lorentzian sum (GLS or SGL) function is produced by summing together different amounts of a Gaussian and a Lorentzian function, while the Gaussian-Lorentzian product (GLP or GL) function is made by multiplying a Gaussian and a Lorentzian function together. In both cases, the relative contributions of the two underlying functions is controlled by a mixing parameter, *m*, as indicated by SGL(m) or GL(m), where *m* takes on values of 0 – 100, or equivalently (with minor adjustments to the mathematics) from 0 – 1.[3-6] Nevertheless, a given value of *m* does *not* mean the same thing in these pseudo-Voigt (or Voigt) functions. That is, the GL(30) and SGL(30) functions are somewhat different, especially at their wings, and they are also different from a comparable Voigt function.[3, 6] A generalized Lorentzian function, the 'asymmetric Lorentzian' or LA function, has also been defined for XPS peak fitting. It employs different exponents on each side of the function to control the tail/introduce asymmetry into it, and it can be convolved with a Gaussian to produce a Voigt-like function. A recently formulated version of the Voigt function is also used in this study. It is a symmetric LA function with a single parameter ($LA(m)$ or $LA(x, m)$), where m varies from 0 – 100.[6] While the similarities and differences between the Voigt, GL, SGL, multiparameter LA, which allows asymmetry, and single parameter LA (a true Voigt) functions may introduce some complexity, and perhaps even confusion, into XPS peak fitting, the availability of different synthetic fit functions provides the XPS data analyst with increased flexibility. For example, it can be helpful to optimize the amount of Gaussian and Lorentzian character in the GL, SGL, and $LA(m)$ synthetic peaks by optimizing the value of *m* in a fit. In this paper, the case for using appropriate synthetic peak shapes in fitting data is made by demonstrating that non-linear optimisation can deliver accurate results for data when the line shape is precisely known. However, when the optimal peak shape is not known, which is the typical case, less-than-perfect peak shapes in a peak model may yield reasonably accurate information about a spectrum when appropriate constraints are applied to them.

The message of this paper is summarized in the following four related observations:

Observation 1. If constraints are not placed on peak shapes, and/or if there is insufficient background knowledge regarding the material being studied, optimization of a peak fit/model using a figure-of-merit alone may lead to suboptimal peak fitting.
Observation 2. It is questionable whether arbitrary peak shapes can be appropriately used in XPS peak fitting unless constraints are placed on them.

Observation 3. Constraints can prevent an optimization from reaching the mathematical minimum of a figure-of-merit. Thus, at least in a mathematical sense, we can define optimal peak fitting as peak fitting performed with unconstrained parameters, and vice versa for suboptimal peak fitting.
Observation 4. For the purposes of determining physically meaningful quantities, suboptimal optimisation is desirable for peak models constructed from components with suboptimal peak shapes.

These observations are illustrated with case studies based on semi-synthetic and real data fitted with synthetic peaks. The consequences of these observations, including the failure to heed the advice in them, have been previously noted – there are many examples in the scientific literature of peak fits that do not employ proper peak shapes or constraints. These concerns are part of the more general reproducibility crisis in science.[7-10] Recently, to address some of these concerns, various XPS experts have published a series of guides on this topic.[1] These guides cover many aspects of XPS, including a general introduction to the technique,[11] the first steps in planning XPS measurements,[12, 13] sample handling,[14] instrument calibration,[15] the use of backgrounds,[16] the C1s spectrum,[17] analysis of survey scans,[18] and a general introduction to XPS peak fitting.[13, 19] This guide here is part of a second series of papers that address the reproducibility crisis in both XPS and other surface analytical techniques. In it, we emphasize the importance of XPS analysts using correct synthetic peaks in their peak fitting. However, when this is not possible, which is often the case because of a lack of knowledge about a sample (for the most part there would be little point in characterizing a sample if we knew everything about it), appropriate constraints should be applied to the synthetic fit components. However, this latter statement should not be taken as a license to be careless in the choice of the peak shapes used in XPS peak fitting. For example, XPS signals with obvious Gaussian or Lorentzian character should be fitted with synthetic fit components that match these shapes.

Theory

This section describes the basic peak shapes/functions, backgrounds, and statistical tools that are commonly used in XPS peak fitting. More complete treatments of these subjects are available in the literature.[3, 20-23]

The Gaussian and Lorentzian Functions

A synthetic peak shape is defined here as a mathematical function used to fit XPS data. The Gaussian (Eq. 1) and Lorentzian (Eq. 2) functions form the basis of most XPS fit functions:

$$Lorentzian: l(x) = \frac{1}{1+4x^2} \quad (1)$$

$$Gaussian: g(x) = e^{-4ln2x^2} \quad (2)$$

Of course, there are other mathematically similar/essentially identical definitions of these functions. As shown in Figure 1, the Gaussian and Lorentzian functions are similar at their peaks/apexes (although the Lorentzian is slightly narrower), but quite different on their edges/wings – the Lorentzian drops to zero relatively slowly; the Gaussian is more compact. An XPS analyst should be able to distinguish between these two functions and estimate whether a given experimental lines shape is more Gaussian or Lorentzian in nature.[3, 6]

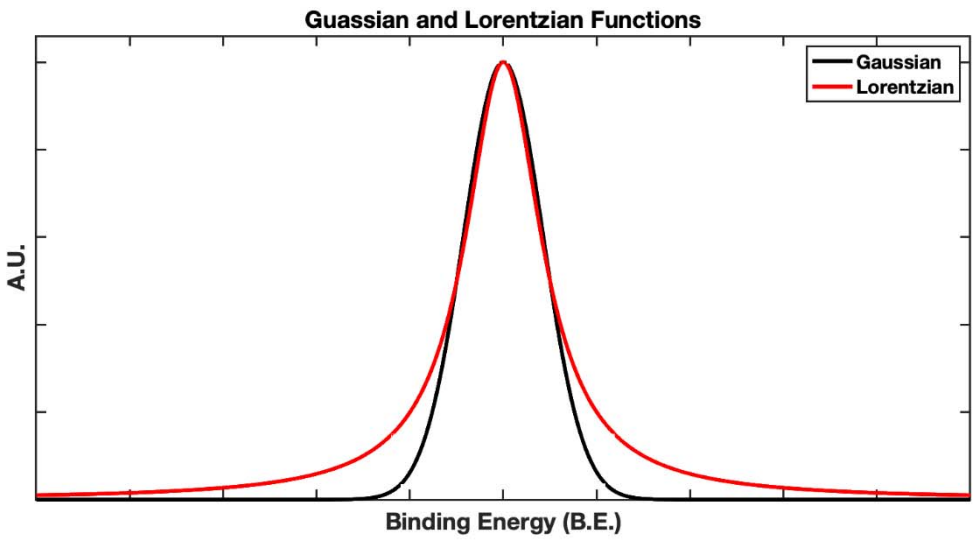

Figure 1. Gaussian (black) and Lorentzian (red) functions showing the more compact nature of the Gaussian function.

The Voigt Function

The Voigt function is often the best representation of photoemission signals in XPS. The Voigt function is the convolution of a Gaussian and a Lorentzian. In general, the underlying shape of photoemission signals in XPS is Lorentzian, which, as noted above, is then modified by several factors. The Voigt function *V(x',y)* can be expressed as:[5]

$$V(x',y) = \frac{y}{\pi}\int_{-\infty}^{\infty}\frac{e^{-t^2}}{(x'-t)^2+y^2}dt \tag{3}$$

where

$$x' = 2\sqrt{log_e 2}\frac{(x-x_0)}{ghm} \tag{4}$$

and

$$y = 2\sqrt{log_e 2}\frac{lhm}{ghm} \tag{5}$$

where *ghm* and *lhm* are defined as the full width at half maxima of the Gaussian and Lorentzian functions, respectively. Finally, a mixing parameter, *M*, with a value between 0.0 and 1.0, with 0.0 being a pure Gaussian and 1.0 a pure Lorentzian, is defined as:

$$M = \frac{lhm}{lhm+ghm}, \tag{6}$$

where *M* a numerical representation of the mixing that is done in Eq. 3 (the amounts of Gaussian and Lorentzian character in the Voigt function are controlled by the widths of its Gaussian and Lorentzian components). Eq. 3 cannot be evaluated analytically, only

numerically. Because of this, true convolution algorithms are computationally expensive,[4, 5] only becoming accessible on personal computers relatively recently. Prior to this time, pseudo-Voigt functions (see discussion below) were often used.

Gaussian-Lorentzian Sum and Product Pseudo-Voigt Peak Shapes

The Gaussian-Lorentzian product (GL) pseudo-Voigt peak shape, Eq. 7, is formed from the product of Gaussian and Lorentzian functions (Eqs. 1 and 2),[3]

$$GL(x,m) = l\left(\frac{x}{f_L}\right) \times g\left(\frac{x}{f_G}\right), \tag{7}$$

where the FWHM for each function, $f_L$and f_G, Eq. 8, vary according to the parameter, m, as follows:

$$f_L{}^2 = \frac{1}{\frac{m}{100}} \text{ and } f_G{}^2 = \frac{1}{1-\frac{m}{100}} \tag{8}$$

Similarly, the Gaussian-Lorentzian sum function (GLS) (Eq. 9) is a pseudo-Voigt peak shape that is created by summing Gaussian and Lorentzian functions (Equation 9). For this peak shape, m again takes on values from 0 to 100.

$$GLS(x,m) = l\left(\frac{x}{f_L}\right) + g\left(\frac{x}{f_G}\right) \tag{9}$$

The Generalized Lorentzian Peak Shape

Eq. 10 defines a generalized Lorentzian peak shape, l_g or LA.[21] Here, each side of a Lorentzian is raised to a different power: α or β, which has the effect of introducing asymmetry into the function.

$$l_g(x{:}\,\alpha,\beta) = \begin{cases} [(l(x)]^{\alpha} & x \le 0 \\ [l(x)]^{\beta} & x > 0 \end{cases} \tag{10}$$

This generalized Lorentzian can then be convolved with a Gaussian function, g, to produce another pseudo-Voigt function (Eq. 11).

$$LA(x{:}\,\alpha,\beta,n) = N\int_{-\infty}^{\infty} l_g(\tau{:}\,\alpha,\beta)g(x-\tau{:}\,f_G(n))d\tau, \tag{11}$$

where N is a normalization factor. The $LA(x,m)$, or just $LA(m)$, peak shape (Equation 12) is constructed from Equation 10 with a mixing parameter, m, that ranges from 0, where the function has the greatest possible amount of Gaussian character, to 100 where it has the greatest possible amount of Lorentzian character, as follows:

$$LA(x,m) = LA\left(x{:}\,1{,}1{,}(1401-\left(\frac{m}{100}\right)1401)\right). \tag{12}$$

This single-parameter function has been described in more detail in a recent publication.[6]

Linear Background

The simplest background that is regularly used in XPS peak fitting is the linear background. It consists of a horizontal or sloped straight line. This background often effectively models signals for which there is no change in the baseline. These signals often come from materials with wide band gaps.

Shirley Background[16]

The Shirley background is a common, semi-empirical background used in XPS peak fitting that is often used to model the large changes, e.g., steps, in the background that are observed for metallic materials. In general, this background is iterated until it converges, where this convergence is usually rapid. The Shirley calculation/iteration is based on the assumption that the background at any binding energy within the peak envelope is proportional to the integrated intensity of the peak at higher kinetic energies. In essence, the Shirley background is a Tougaard background (see below) with $F(E)$ equal to a step function. Another variation of the iterative Shirley background is the so-called 'Smart background' approach, which prevents the background from being above the actual data. The original Shirley background was not iterated – it was applied in one pass.

U 4 Tougaard Background[16]

The peak models applied to the Si 2s spectra shown herein include background curves defined for SiO_2 by Tougaard.[24] The Tougaard background, $T(E)$, is computed from the measured spectrum, $J(E)$, representing the photoelectron peak, as it is multiplied by a shifted function, the Tougaard Universal Cross-Section, $F(E)$, that accounts for electron energy losses/inelastic scattering in the material, as follows:

$$T(E) = \int_E^{\infty} F(E' - E) J(E') dE', \tag{13}$$

where $F(E)$ is defined as:

$$F(x) = U(x: B, C, D, T_0) = \begin{cases} \frac{Bx}{(C-x^2)^2 + Dx^2} & x > T_0 \\ 0 & x \leq T_0 \end{cases} \tag{14}$$

When used with the SiO_2 spectra herein, the universal cross-section uses the parameters B = 299, C = 542, D = 275, and T_0 = 9.3, as calculated for SiO_2 by Tougaard.[24]

Residual Plots and Statistics

XPS experimental data is approximated in a peak model by summing the intensities of one or more synthetic peaks over a chosen background. As defined in CasaXPS,[25] the residual plot gives the differences between the experimental data and the synthetic envelope. It may be presented directly or in normalised form. The residual plot is usually accompanied by one of two figure-of-merit statistics: the residual standard deviation (RSD) or the root mean square, which are defined as follows:

$$Residual\ StDev = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{d_i - e_i}{\sqrt{d_i}}\right)^2} \quad (15)$$

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - e_i)^2}, \quad (16)$$

where the d_i are the original data at a given bin i in units of counts per bin, and the e_i are the corresponding values of the synthetic/model envelope. The RSD is more accurately described as the normalised RSD since the deviation between the data and the approximation when squared are normalised with respect to the value for one standard deviation in counts. The underlying assumption when computing the residual in normalised form is that the data correspond to pulse counted signal obeying Poisson Statistics. The act of normalising as performed in Eq. 15 implies that the expected value for the RSD statistic is unity for Poisson distributed intensities.

Results and Discussion

We now present four case studies that illustrate the four observations made in the Introduction of this paper.

Case 1. Fit to a Semi-Synthetic N 1s Spectrum

In this section, we show that results obtained from peak fitting can be accurate for models constructed without constraints, provided that both the correct number of component curves and the appropriate functional forms for these curves are used. To illustrate this, we synthesized an N 1s spectrum by copying, shifting, scaling, and adding together the single N 1s signal from the spectrum of nylon 6 to generate a spectrum with three underlying components with a known area ratio of 2:5:1 (see Figure 2a). The lack of baseline rise in this composite spectrum (and therefore the linear baseline that was used) is consistent with the wide band gap (HOMO-LUMO energy difference) of nylon 6. We first attempted to fit this semi-synthetic spectrum using GL line shapes, which are widely employed in XPS peak fitting. We introduced three GL peaks into our model and then varied the mixing parameter, *m*, from 0 to 100, in the peak fitting/optimization, where the same mixing parameter was used for all three peaks. No other constraints were applied to this model. As shown in Figure 2b, the minimum in the plot of the error of the fit (chi square) vs. *m* occurs at 55, which suggests that GL(55) peaks best fit this data. However, when our semi-synthetic spectrum is fit to three GL(55) peaks (see Figure 2c), area ratios of 1.8:5.5:1 are obtained, i.e., we do not obtain the correct (previously known) ratios of 2:5:1. Without this previous knowledge of the expected ratios, it would have been tempting to consider this an appropriate fit. Certainly, these results are not exceedingly far from the true results – we are off by ca. 10%. Nevertheless, because of our previous knowledge of the spectrum, we know to modify these peaks or place constraints on them to force the proper area ratios, which will hopefully also maintain the low figure-of-merit error for the fit. As shown in Figure 2d, changing the peak shape to a GL(45) produces the correct area ratios (see also the plot of the ratio of peaks B and C as a function of m in Figure 2e). However, the RSD for the fit with the GL(45) peaks (1.632) is higher than the RSD for the fit with the GL(55) peaks (1.427). Thus, the optimal mathematical fit (at m = 55) here does not match the optimal chemical fit/interpretation of the data (at m = 45). The errors in the peak areas were introduced by a lack of knowledge of the underlying peak shape. We end this paragraph with two

additional observations about these fits. First, as judged by their edges, the underlying, raw signals in Figure 2a are rather compact, which suggests that peaks with a reasonable amount of Gaussian character would be appropriate for fitting them, and both the GL(45) and GL(55) are quite 'Gaussian-like'. Second, it is possible to use an experimental peak as a fit component. When three copies of the raw N 1s data signal used to synthesize the spectrum in Figure 2a were used to fit the data, a perfect fit to the data was obtained (see Figure 2f). Of course, this was expected. In summary, these results underscore the fact that there is some uncertainty associated with XPS peak fitting when the underlying peak shape(s) for the components in a fit are unknown. Unfortunately, this type of situation is more the norm than the exception.

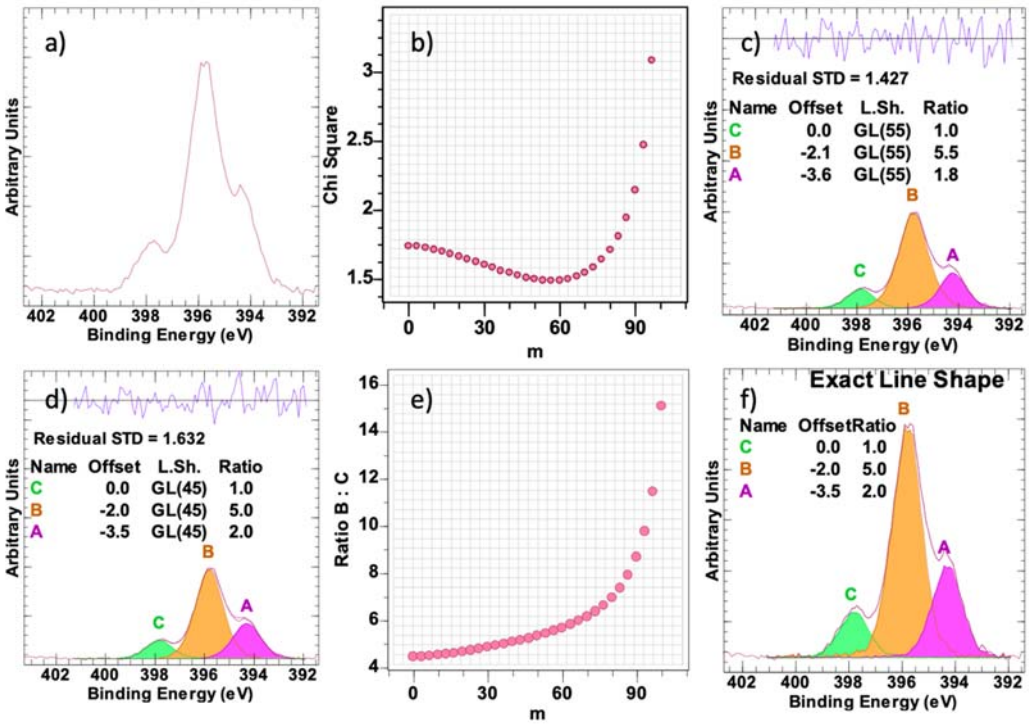

Figure 2. (a) Semi-synthetic N 1s spectrum made by copying, shifting, and scaling the single N 1s peak from a sample of nylon 6. (b) Plot of the error of the fit to the semi-synthetic spectrum in (a) using GL(m) line shapes vs. *m*. (c) Fit to the semi-synthetic spectrum in (a) fit with GL(55) peaks. (d) Fit to the semi-synthetic spectrum in (a) with GL(45) peaks. (e) Plot of the ratio of Peak B to Peak C in the fit to the semi-synthetic spectrum in (a) vs. *m*. (f) Fit of the semi-synthetic spectrum in (a) using the original N 1s peak from the sample of nylon 6 to fit the data in (a).

Case 2. Fit to the C 1s spectrum of cellulose

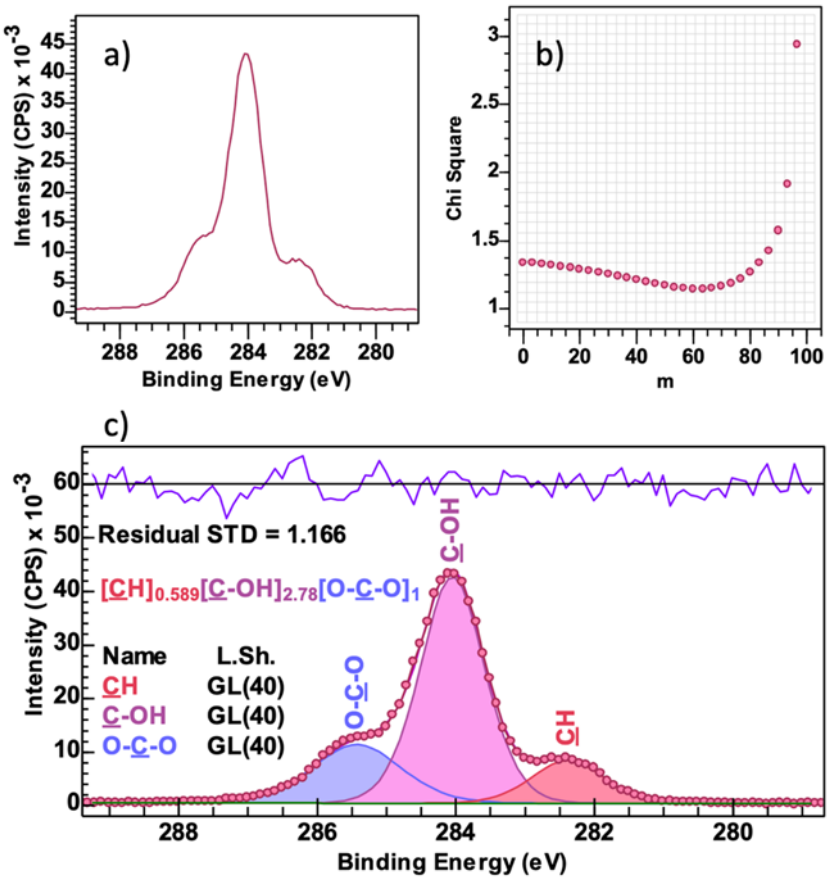

Figure 3. (a) C 1s spectrum of coffee filter paper (mostly cellulose). (b) Plot of the error of the mixing parameter *m* for the GL(m) function. (c) The spectrum in (a) fitted using three GL(40) component peaks. Because the ratio of the C-OH and O-C-O peaks is 2.72:1, instead of the theoretical 5:1 ratio for cellulose, the GL(40) is an inappropriate peak shape for this fit.

The C 1s narrow scan in Figure 3 is of coffee filter paper, which is mostly pure cellulose. Using the approach above for Figure 2, this spectrum was fit to unconstrained GL(m) peaks in Figure 3, where the optimal value of *m* was determined to be 40. Based only on the RSD and distribution of the residuals in this fit, one might conclude that this fit is good. However, similar to the results from Figure 2, the fit gives a 2.72:1 ratio for the C-OH:O-C-O peaks, which is far from the ratio of these two components in cellulose (5:1). Additionally, the fit is inappropriate because the widths of the fit components vary noticeably. (Recall that a common error in XPS peak fitting is for fit components to have significantly different widths when there is no good chemical or physical reason for this to happen.[19]) Again, we see a case where unconstrained components with an 'optimal' value of *m* fit the data well/tightly, but do not produce a physically reasonable result. As we will see throughout this article, because we often do not know what the best fit function should be for a spectrum, constraints are usually necessary to give physically and chemically meaningful results in XPS peak fitting.

Case 3. Fit to a Semi-Synthetic Si 2s Spectrum

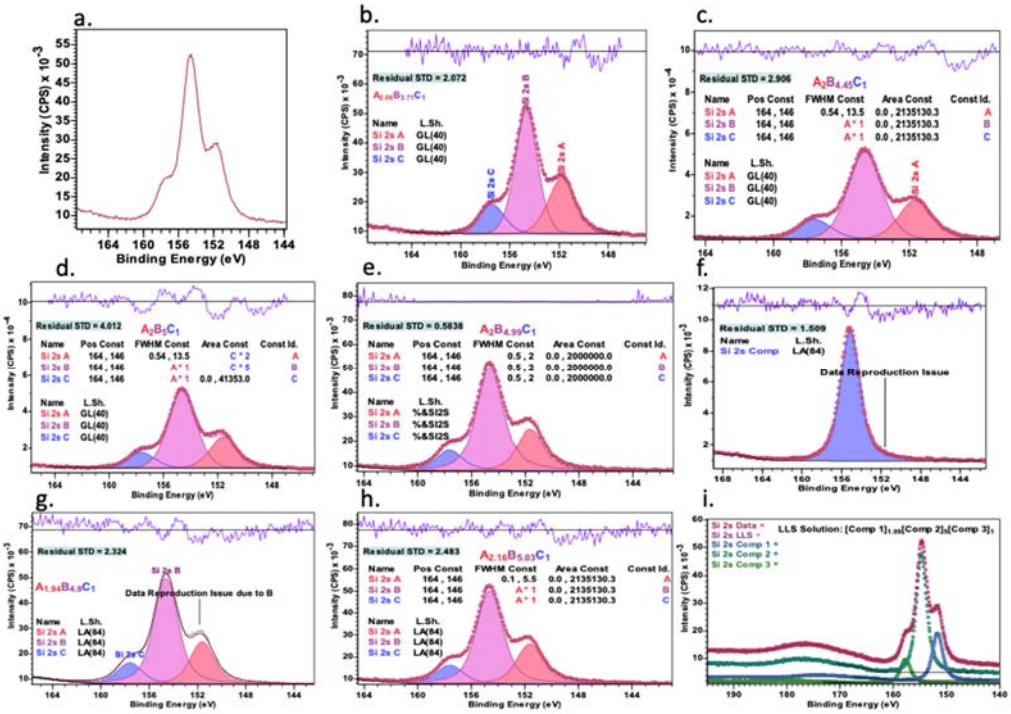

Figure 4. (a) Semi-synthetic spectrum made from real Si 2s peaks. (b) Fit of the semi-synthetic spectrum in (a) with unconstrained GL(40) line shapes, which results in improper peak area ratios. Also, components A and C having different widths. (c) Fit of the semi-synthetic spectrum in (a) with GL(40) line shapes with width constraints, which gives better peak area ratios, but result in a suboptimal mathematical fit, i.e., a higher RSD. (d) Fit of the semi-synthetic spectrum in (a) with GL(40) line shapes with the peak widths constrained to have the same values and the relative peak areas in the fit constrained to the theoretical ratio of 2:5:1. This fit produces a poor RSD value and residuals. (e) Fit of the semi-synthetic spectrum in (a) with the exact, but unconstrained, underlying line shapes used to make the semi-synthetic spectrum, which returns the correct area ratios. (f) Original spectrum (single peak) used to create the semi-synthetic spectrum in (a) fitted using a Voigt function that was optimized to determine the mixing parameter. The LA(84) gave the best fit, with minor discrepancies on the high binding energy side of the peak. (g) Fit of the semi-synthetic spectrum in (a) to unconstrained LA(84) Voigt functions. (h) Fit of the semi-synthetic spectrum in (a) with LA(84) Voigt functions with constrained (equal) FWHM values. (i) The copied, shifted, and/or scaled peaks (as in (f)) that were summed to create the semi-synthetic spectrum in (a).

We next created another semi-synthetic spectrum (Figure 4a) by copying, shifting, scaling, and adding together an experimental Si 2s peak to create a spectrum with three overlapping components with area ratios of 2:5:1. The Si 2s peak used to make Figure 4a has strong Lorentzian character. The purpose in making this semi-synthetic data was, again, to have a composite spectrum with well-known provenance and composition that could be used to test peak models. The semi-synthetic spectrum in Figure 4a was first fit with three unconstrained GL(40) peaks. As shown in Figure 4b, nonlinear least squares returns a 2.06:3.71:1 ratio, which is quite far from the true answer. However, the residuals of this fit are rather well distributed around zero, so from this limited (mathematical) perspective the fit is good. Next, these GL(40) peaks were constrained to have the same FWHM values (Figure 4c), which is often a reasonable assumption/approximation in XPS peak fitting. When this was done, the areas of the fit

components have a 2:4.45:1 ratio. This result is closer to the correct answer, even though the residuals to this fit are not as well distributed as in Figure 4b, and the RSD of this fit is higher. An additional constraint was then placed on the fit components, which is to both constrain them to have the same FWHM values *and* to fix their areas to the correct 2:5:1 ratio (Figure 4d). The point of this exercise was to see how these constraints would affect the residuals. Interestingly (and perhaps ironically), although we have enforced both the correct sample chemistry (area ratios) and the reasonable constraint of equal FWHMs, the residuals and the RSD for this fit are rather poor. Finally, to again test the ability of nonlinear least squares to fit this data under ideal conditions, the semi-synthetic spectrum in Figure 4a was fit with the peak that was used to make it. This approach produced an essentially perfect fit, where the area ratio of the three resulting fit components is 2:4.99:1, and the RSD is 0.5838 (Figure 4e).

In general, s orbitals produce broader XPS peaks with greater Lorentzian character than other orbitals. Figure 4f shows the fit of the Si 2s peak used to make the semi-synthetic spectrum in Figure 4a with an $LA(x, m)$ function, which is a true Voigt function. In the $LA(x, m)$ function, the mixing parameter, m, can take values from 0 – 100, where 0 and 100 correspond to highest degrees of Gaussian and Lorentzian character for the function, respectively. The mixing parameter, m, was a parameter in this fit, which yielded m = 84 (a high degree of Lorentzian character) as the optimal value. However, while the fit is quite good, its residuals reveal that it is not perfect – the fit component and the data neither agree exactly around the peak maximum nor on the lower binding energy side of this peak. This lack of fit may be due to the presence of a plasmon loss signal from the lower binding energy Si 2p peak. We justify the use of this imperfect signal to make Figure 4a because (i) the fit of the raw data to a peak with strong Lorentzian character is quite good, and (ii) many XPS spectra that are fit will similarly have imperfections in them, which makes constraints in peak models even more important. The semi-synthetic spectrum in Figure 4a was next fit both with (i) unconstrained LA(84) peaks (Figure 4g), which resulted in good area ratios for the peaks of 1.94:4.9:1, and (ii) LA(84) peaks that were constrained to have the same widths (Figure 4h), which also resulted in good area ratios for the peaks of 2.16:5.03:1.

In summary, these results suggest that unconstrained, arbitrary peak shapes cannot be used to appropriately fit XPS data. In this particular case, unconstrained GL(40) peaks did not fit the semi-synthetic spectrum in Figure 4a well. However, unsurprisingly, an unconstrained peak that closely matches the original data, i.e., the LA(84) signal, does fit its corresponding semi-synthetic spectrum well.

Case 4. Fit to the C 1s Spectrum from an Ionic Liquid

In this example, we show that even with reasonable peak shapes, an unconstrained fit may produce unphysical results. Figure 5 shows the structure and C 1s spectrum of an ionic liquid. As their name implies, ionic liquids are ionic species that are liquids around room temperature. Many ionic liquids have extremely low vapor pressures and are thus compatible with conventional (high vacuum) XPS analysis. (Materials with higher vapor pressures can often be analyzed by near ambient pressure (NAP)-XPS.[26]) Before discussing attempts to fit this spectrum, note that the expected ratio of signal emanating from carbon bonded to carbon and hydrogen, carbon bonded to nitrogen, and carbon bonded to fluorine is 5:4:2. An initial attempt at fitting was with unconstrained GL(40) peak shapes (Figure 5b). The peak ratios obtained (4:4.26:1.72) were somewhat problematic. Accordingly, a more ideal peak shape (an unconstrained Voigt function with some asymmetry) was used next. (See the next paragraph

for a discussion of the small amount of asymmetry introduced into this peak shape.) In this fit (see Figure 5c), the RSD value has decreased by about 40%, indicating a more appropriate fit. However, the area ratios of the components are still problematic (4:4.69:1.92). Finally, using the same peak shape, but constraining the area ratio of the carbon-nitrogen signal to the C-C/C-H signal, we achieve both a good RSD value (only slightly higher than in the previous fit) as well as the expected area ratios (4:5:1.99). Note that the final, correct ratios among three components were achieved by constraining two of them to each other. These observations demonstrate that fitting XPS data without proper use of peak shapes *and* an appropriate number of relational constraints may result in less than fully accurate intensity ratios that do not fully elucidate the chemistry of the sample.

The primary objective in creating an asymmetric peak shape is to match the shape of experimental XPS data, and not necessarily to model the underlying physical processes leading to the observed asymmetry. Indeed, in some cases, the asymmetry in a signal may be an artefact of the measurement process. For example, a slow change in surface potential during data acquisition may manifest itself as asymmetry in peaks. The apparent binding energies of the peaks from the ionic liquid tended to drift with time in repeated measurements. The asymmetry in the peak shapes in Figures 5c-d was generated using an adjustment to the symmetric LA peak shapes LA(20) and LA(70), which consisted of multiplication of these symmetric Voigt line shapes by a sigmoidal shape with its inflexion point at the maximum of these Voigt functions. The amount of asymmetry introduced into these signals was not large.

We now discuss the reasons for using two different peak shapes (LA(20) and LA(70)) in the C 1s fit in Figure 5. As discussed above, (i) the mixing parameter, m, of the $LA(x, m)$ Voigt function determines the amount of Lorentzian character in the peak shape, with 100 being the highest (most Lorentzian shape), and (ii) the Lorentzian peak shape best represents the underlying shape of photoemission, with Gaussian character introduced into a peak shape because of, say, sample heterogeneity, or the instrument. Therefore, an LA(70) peak shape has quite a bit more Lorentzian character than the LA(20) peak shape, which is more Gaussian. The need for these different peak shapes in the fit in Figure 5 appears to follow from the chemistry/structure of the ionic liquid. At first glance, the C-N carbons in the cation of the ionic liquid appear to be in identical chemical environments – each carbon is bonded once to nitrogen. However, there are subtle differences between these carbons. These carbons appear as a methyl (CH_3) and three methylene (CH_2) units, where two of the methylene units are in a ring and the other is in a straight alkyl chain. Thus, the greater degree of Gaussian character of the LA(20) peak that describes this chemical state appears to be a result of these similar, but still distinct, chemical environments. That is, the single LA(20) peak shape can be seen as modeling three closely-spaced/overlapping signals. The C-C/C-H type carbons in the ionic liquid are similarly modeled with LA(20) peak shapes in this model – these carbons are also slightly different from each other in the molecule. In contrast, the C 1s peak at ca. 293.5 eV attributed to carbon in the 'CF_3' groups in the anion of the ionic liquid only has one chemical form here, which is consistent with the more Lorentzian LA(70) peaks used to describe it. The other photoemission peaks from the ionic liquid of the N 1s, F 1s, S 2p, and O 1s regions similarly suggest the validity of the model proposed in Figure 5d. We therefore present it with a high degree of confidence that these components relate well to the material as specified.

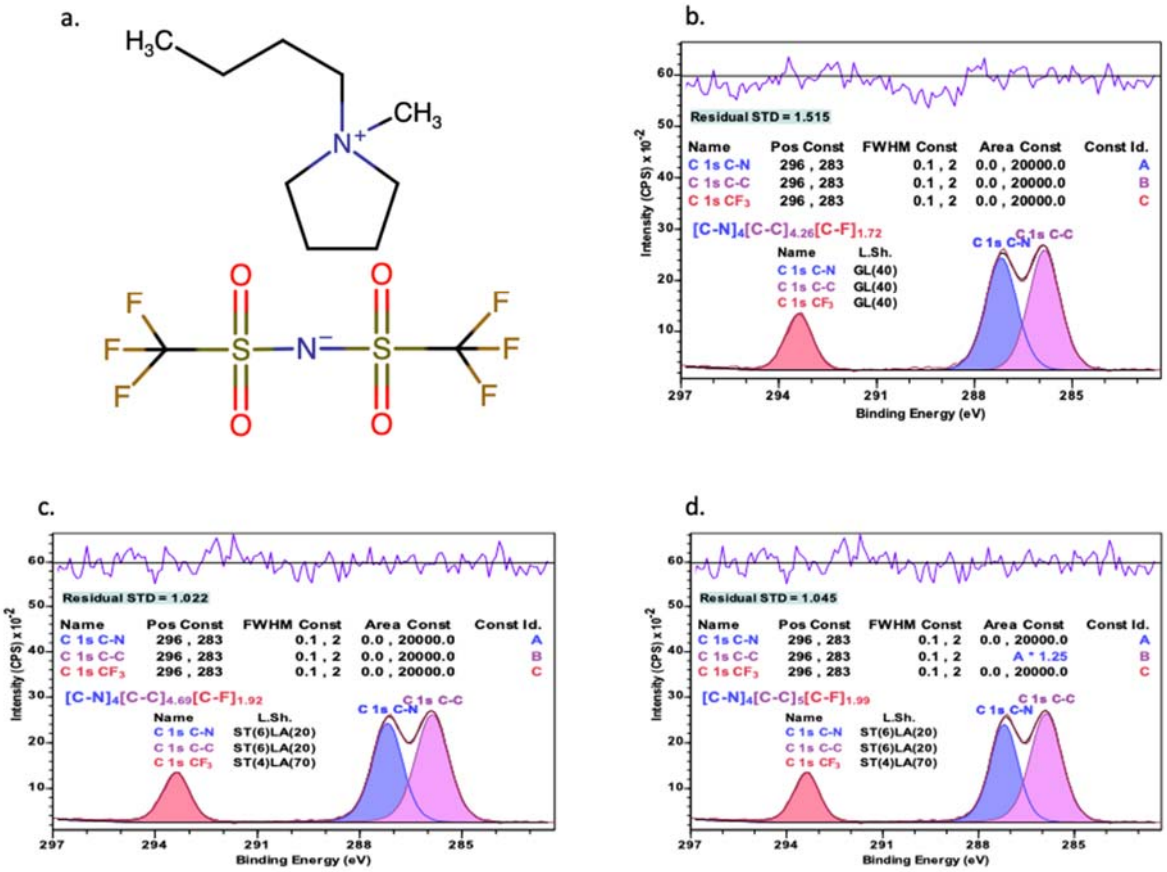

Figure 5. (a) Top: Ionic liquid cation: $C_4C_1Pyrr^+$. Bottom: Ionic liquid anion: TF_2N^-. (b) C 1s XPS narrow scan of the ionic Liquid $[TF_2N][C_4C_1Pyrr]$ fit with a three-component peak model using unconstrained GL(40) peak shapes. (c) Same spectrum as in (b) fitted with unconstrained Voigt (LA(20) and LA(70)) functions. (d) Same spectrum as in (b) fitted as in (c), but with an area constraint, which provides a slightly less optimal fit (mathematically) compared to the fit in (c) that used unconstrained Voigt functions, but the correct ratios of chemical states.

Conclusion

We have presented four case studies of fitting bell-shaped curves to XPS data that illustrate the importance of both estimating the best possible peak shapes and using constraints in developing XPS peak models. Two of these examples were based on semi-synthetic spectra that had known component/peak ratios. Failure to obtain a good approximation for the bell-shaped curves (and the background shapes) in a fit creates uncertainties in the results. That is, fits based on unconstrained, less-than-ideal peak shapes generally produce fits with moderate to large errors in their predicted peak ratios that neither support nor rule out many chemical structures. Furthermore, relying on optimization to find the best parameters for peak shapes can lead to errors in peak ratios. In other words, the mathematically best fit is not always the most chemically reasonable fit. In summary, successful XPS chemical state analysis based on non-linear optimization depends on developing peak shapes specific to given problems and applying appropriate constraints.

Author Declaration

One of the authors, NF, writes, sells, and promotes the fitting software (CasaXPS) used in this work.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. D. R. Baer, G. E. McGuire, K. Artyushkova, C. D. Easton, M. H. Engelhard and A. G. Shard, J Vac Sci Technol A **39**, 021601 (2021).
2. B. Moeini, M. R. Linford, N. Fairley, A. Barlow, P. Cumpson, D. Morgan, V. Fernandez and J. Baltrusaitis, Surf. Interface Anal. **54**, 67-77 (2022).
3. V. Jain, M. C. Biesinger and M. R. Linford, Appl. Surf. Sci. **447**, 548-553 (2018).
4. P. M. A. Sherwood, Surf. Interface Anal. **51**, 589-610 (2019).
5. P. M. A. Sherwood, Surf. Interface Anal. **51**, 254-274 (2019).
6. G. H. Major, V. Fernandez, N. Fairley and M. R. Linford, Surf. Sci. Interface **54**, 262-269 (2022).
7. M. Baker, Nature **533**, 452-454 (2016).
8. H. V. Fineberg, *Reproducibility and replicability in science*. (National Academies Press, 2019).
9. R. Harris, Chem. Eng. News **95**, 2 (2017).
10. J. Park, J. D. Howe and D. S. Sholl, Chem. Mater. **29**, 10487-10495 (2017).
11. F. A. Stevie and C. L. Donley, J. Vac. Sci. Technol. A **38**, 063204 (2020).
12. D. R. Baer, K. Artyushkova, C. Richard Brundle, J. E. Castle, M. H. Engelhard, K. J. Gaskell, J. T. Grant, R. T. Haasch, M. R. Linford and C. J. Powell, J. Vac. Sci. Technol. A **39**, 017003 (2021).
13. D. R. Baer, K. Artyushkova, C. Richard Brundle, J. E. Castle, M. H. Engelhard, K. J. Gaskell, J. T. Grant, R. T. Haasch, M. R. Linford, C. J. Powell, A. G. Shard, P. M. A. Sherwood and V. S. Smentkowski, J. Vac. Sci. Technol. A **37**, 031401 (2019).
14. F. A. Stevie, R. Garcia, J. Shallenberger, J. G. Newman and C. L. Donley, J Vac Sci Technol A **38**, 063202 (2020).
15. J. Wolstenholme, J Vac Sci Technol A **38**, 043206 (2020).
16. S. Tougaard, J. Vac. Sci. Technol. **39**, 011201 (2021).
17. T. R. Gengenbach, G. H. Major, M. R. Linford and C. D. Easton, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films **39**, 013204 (2021).
18. D. Shah, D. I. Patel, T. Roychowdhury, G. B. Rayner, N. O'Toole, D. R. Baer and M. R. Linford, J. Vac. Sci. Technol. B **36**, 062902 (2018).
19. G. H. Major, N. Fairley, P. M. Sherwood, M. R. Linford, J. Terry, V. Fernandez and K. Artyushkova, J Vac Sci Technol A **38**, 061203 (2020).
20. P. Van der Heide, *X-ray photoelectron spectroscopy: an introduction to principles and practices*. (John Wiley & Sons, 2011).
21. G. H. Major, D. Shah, V. Fernandez, N. Fairley and M. R. Linford, Vac. Tech. Coat. **21**, 43-46 (March, 2020).
22. G. H. Major, D. Shah, T. G. Avval, T. Roychowdhury, A. J. Barlow, M. Greiner, V. Fernandez, N. Fairley and M. R. Linford, Vac. Tech. Coat. **21**, 36-40 (May, 2020).

23. G. H. Major, D. Shah, V. Fernandez, N. Fairley and M. R. Linford, Vac. Tech. Coat **21**, 35-39 (April, 2020).
24. S. Tougaard, Surf Sci **216**, 343-360 (1989).
25. N. Fairley, V. Fernandez, M. Richard-Plouet, C. Guillot-Deudon, J. Walton, E. Smith, D. Flahaut, M. Greiner, M. Biesinger and S. Tougaard, Applied Surface Science Advances **5**, 100112 (2021).
26. D. I. Patel, T. Roychowdhury, V. Jain, D. Shah, T. G. Avval, S. Chatterjee, S. Bahr, P. Dietrich, M. Meyer, A. Thißen and M. R. Linford, Surf. Sci. Spectra **26**, 016801 (2019).