

Automated information compression of XPS spectrum using information criteria

Hiroshi Shinotsuka^{a,*}, Hideki Yoshikawa^a, Ryo Murakami^b, Kazuki Nakamura^b, Hiromi Tanaka^b, Kazuhiro Yoshihara^c

^a*Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba 305-0047, Japan*

^b*Department of Electrical and Computer Engineering, National Institute of Technology, Yonago College, Tottori 683-8502, Japan*

^c*Scienta Omicron Inc., Tokyo 140-0013, Japan*

Abstract

We developed and implemented a fully automated method to perform X-ray photoelectron spectroscopy (XPS) spectral analysis based on the active Shirley method and information criteria. Our method searches a large number of initial fitting models by changing the degree of smoothing, and then optimizes the peak parameters and background parameters to obtain a large number of fitting results. The goodness of those optimized models is ranked using information criteria. As a result of applying this algorithm to measured XPS spectra, we found that, using the Bayesian information criterion (BIC), a simple model with reasonably good agreement and a moderate number of peaks was selected. The model selected by the BIC was close to the result of peak fitting performed by XPS analysis experts.

Keywords: X-ray photoelectron spectroscopy (XPS); Active Shirley method; Akaike information criterion (AIC); Bayesian information criterion (BIC)

1. Introduction

High-throughput measurement has become increasingly important for the efficient development of science and technology, and accumulation of large amounts of spectral data is urgently required. Even in X-ray photoelectron spectroscopy (XPS), which is a time-consuming characterization technique, the use of high-intensity synchrotron radiation and a high sensitivity detector enables us to obtain a large amount of spectral data over a short time period. Therefore, high throughput of data processing is also required for efficient spectral data analysis. In the present work, we develop and implement a fully automated method to perform XPS spectral analysis.

In analyzing an XPS spectrum, the background estimation before peak fitting greatly affects the determination of the peak area. It is to be noted that the peak fitting mentioned here refers to the approximate individual peaks constituting the XPS spectrum by Voigt functions. For the XPS background estimation, there are several conventional methods commonly employed, such as the linear method, the

* *Corresponding author.* Tel.: +81-29-851-3354.
E-mail address: SHINOTSUKA.Hiroshi@nims.go.jp

Shirley method [1-2], and the Tougaard method [3], but we adopted a background estimation method known as the active Shirley method proposed by Herrera-Gomez et al. [4]. This method was chosen because it is robust to spectral noises and the end-points of the analysis range of the spectrum selected by the analyst/operator, and it is suitable for automatic analysis of spectra. We have developed an automatic analysis tool for XPS spectra based on the active Shirley method and showed that it can perform automatic peak fitting and film thickness evaluation with excellent reproducibility for SiO₂ thin film samples [5-6].

Because this tool mathematically solves the nonlinear least squares problem by the Marquardt method, the result of the peak fitting can be a local solution that is strongly dependent on the initial parameters of the peak (number of peaks, energy position, height, etc.). This tendency becomes more obvious when the shape of the XPS spectrum has a complicated shoulder structure or satellite structure, or when the statistical noise is large because there are many local solutions. Therefore, the peak fitting result is largely dispersed without converging to the global solution due to the differences in the initial parameters of the peaks.

In this research, to make the active Shirley method applicable to XPS spectra with complicated shapes and a high degree of statistical noise, we developed a heuristic algorithm that systematically sets initial numbers of peaks and automatically extracts a candidate for the global solution. To systematically set the initial numbers of the peak, we adopted a method to systematically change the degree of smoothing applied to the spectrum. By applying the active Shirley method for each initial peak parameter generated in this manner, we methodically obtain many solutions with an automatically estimated background and peaks. Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used as indicators for extracting the desired solution (a candidate for the global solution) from these many solutions.

2. Estimation of optimum solution of XPS peak and background

In this section, we describe the details of our proposed method for the estimation of the optimum solution of the XPS peak and background. Fig. 1 shows a flow chart of our proposed method. Our algorithm is roughly divided into three processes: process 1 creates the initial models, as described in sections 2.1 to 2.3; process 2 optimizes the parameters for all the models, as described in section 2.3; and process 3 chooses the best model using information criteria, as described in section 2.4.

2.1 Systematic search for initial values of XPS peak

To provide a variety of different initial fitting models, we systematically scanned the degree of smoothing applied to a given spectrum. Here, we used the Savitzky-Golay smoothing method [7], the degree of which is tuned by the parameters of smoothing data points and repeat count. We started with a weak smoothing and applied smoothing gradually and strongly but to the extent that the original spectrum

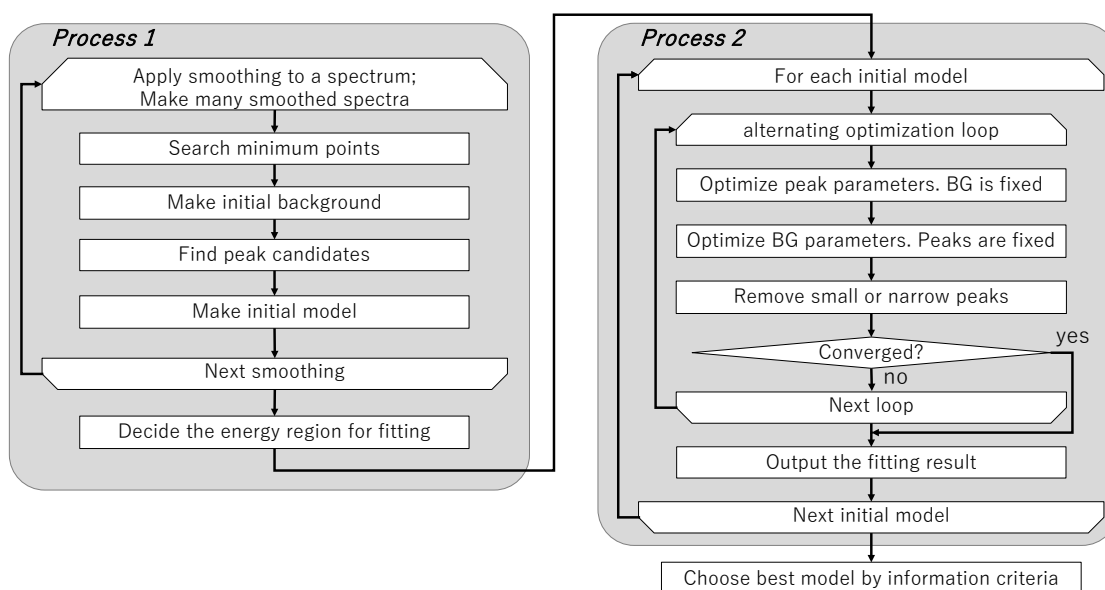


Fig. 1. Flow chart of our proposed method, which is roughly divided into three processes: process 1 creates the initial models, process 2 optimizes the parameters for all the models, and process 3 chooses the best model using information criteria.

shape is not lost. To approach the global solution in a short time calculation, it is necessary to adjust the coarseness/denseness of the search step in the initial fitting models which changed the degree of the smoothing. As a result of studies on various XPS spectra described in section 3, we found a preferable way to apply sequential smoothing. First, 5-points smoothing was performed 480 times, further 7-point smoothing was performed 480 times, and similarly, 9-points, 11-points, 13-points, and 15-points smoothings were sequentially performed 480 times. Eventually, we performed a total of 2,880 smoothings. We started from smoothing with a small width of window in the initial stage where spectral noise is noticeable. We then performed smoothing with a larger window as the spectrum shape became smooth. In that manner, we generated a series of initial fitting models that were evenly distributed for the number of peaks.

By systematically strengthening the degree of smoothing, we generated various initial values of peaks and backgrounds and then peak parameters (number of peaks, energy position, and height, etc.) were optimized for each initial value. We selected the 155 smoothing conditions from the above-mentioned 2,880 smoothings to save the computational time of optimizing the peak parameters at every smoothing stage. Optimization is carried out every two times until the smoothing time is increased to the 40th time, followed by every 4 times up to the 80th time, every 8 times up to the 240th time, every 16 times up to the 960th time, and then optimized for every 32 times up to the 2,880th time. As a result, we obtained 155 optimization models in total.

2.2 Model function describing XPS peaks

The model function, $f(x)$ of an XPS spectrum is given by

$$f(x) = \sum_{k=1}^K A_k V(x; \mu_k, w_k, r_k) + b(x; I_S, I_E) \quad (1)$$

as a function of the binding energy, x . The first term of Eq. (1) is the linear combination of K pseudo-Voigt functions, $V(x; \mu_k, w_k, r_k)$:

$$V(x; \mu, w, r) = rL(x, \mu, w) + (1 - r)G(x, \mu, w), \quad (2)$$

$$L(x, \mu, w) = \frac{1}{1 + \left(\frac{x - \mu}{w}\right)^2}, \quad (3)$$

$$G(x, \mu, w) = \exp\left(-(\log 2) \left(\frac{x - \mu}{w}\right)^2\right) = 2^{-\left(\frac{x - \mu}{w}\right)^2}, \quad (4)$$

where A is the peak height, μ is the peak position, w is the half width at half maximum (HWHM) of the peak, and r is the Lorentz-Gauss mixing ratio. In our framework, we adopted the sum type of the pseudo-Voigt function [8], whose shape is defined by two parameters: a width and a mixing ratio, where the Lorentzian width and Gaussian width are the same. $L(x, \mu, w)$ and $G(x, \mu, w)$ are, respectively, a Lorentzian and a Gaussian function normalized to 1 in height. The mixing ratio, r , is in the range $0 \leq r \leq 1$, and the Lorentzian function and the Gaussian function can be mixed at an arbitrary ratio. The second term of Eq. (1), i.e. $b(x; I_S, I_E)$, corresponds to the background. Based on the approximation of the Shirley method [1-2], the background $b(x; I_S, I_E)$ is expressed by

$$b(x; I_S, I_E) = (I_S - I_E) \frac{Q}{P + Q} + I_E. \quad (5)$$

Here, I_S and I_E are spectral intensities at the end points (starting and ending points) of the higher binding energy side and the lower binding energy side of the analysis range, respectively. P and Q are, respectively, the area of the peak intensity from the higher binding energy end-point to x and the area of the peak intensity from the lower binding energy end-point to x .

2.3 Detection of initial XPS peak and method of shape optimization

We subtracted the initial background from the measured spectrum by the usual iterative Shirley method [2], then calculated the smoothed third-order differential spectrum on the background spectrum using the Savitzky-Golay method [7]. We used that zero point to find the initial peak position and HWHM [5]. The initial value of the Lorentz-Gauss mixing ratio was set to zero to provide a pure Gaussian.

We optimized the peaks and backgrounds after determining their initial values as shown in the flow chart. We used the Levenberg-Marquardt method, which is well known as a solution to the nonlinear optimization problem [9-10]. This method combines the steepest descent method and the Gauss-Newton method, converging quickly to the optimum solution. As a result, all of the parameters included in the model function can be optimized to minimize the sum of the squares of the residual between the measured spectrum (including noise) and the model function described by Eq. (1). However, depending on the XPS spectrum, the parameters may oscillate during the iteration in the Levenberg-Marquardt method. Thus, to

suppress the oscillation, we finally adopted a modified Marquardt method by Fletcher [11] with devised diagonal matrix elements employed when updating the parameters.

After peak and background optimization, we often see tiny peaks with negligible area intensity or extremely sharp peaks. Such peaks are removed by imposing constraint conditions. Specifically, we excluded these peaks that occupy <1% of the total peak area or are sharper than the energy resolution of the XPS analyzer (e.g. full width half maximum <0.2 eV). After peaks were removed, optimization was performed again. We repeated this procedure until there was no need for peak removal.

2.4 Model selection by AIC and BIC

We prepared initial values of various peak numbers, optimized the peaks and background for each initial value, and obtained the final fitting results. Generally, the larger the number of peaks that increase, the smaller becomes the difference (i.e. error function) between the model function and the measured spectrum. However, this causes over-fitting, where the model function becomes too complicated such that physical interpretation is impossible. In order to avoid this over-fitting, it is necessary to take not only the error function but also the trade-off with the model complexity into consideration. As a simple method for doing so, we considered AIC and BIC as model selection criteria [12-13].

The calculation formulae of AIC and BIC are as follows.

$$AIC = -2 \log \hat{L} + 2m, \quad (6)$$

$$BIC = -2 \log \hat{L} + m \log N. \quad (7)$$

\hat{L} is the maximum likelihood calculated from the likelihood between the measured spectrum $\{y_i\}_{i=1}^N$ and the model function $\{f_i\}_{i=1}^N$ obtained as a result of optimization. N is the number of data points of the spectrum, and m is the number of parameters included in the model function. For one peak, there are four parameters: height, position, HWHM, and Lorentz-Gauss mixing ratio. There are also two parameters for determining the background based on the approximation of the Shirley method. Therefore, the number of parameters is calculated as $m = 4K + 2$, where the number of peaks is K . For both AIC and BIC, the first term is the degree of matching between the model function and the measured data, and the second term is the penalty due to the complexity of the model. We calculated AIC and BIC for 155 optimized results, and judged that the best model is obtained from the smallest values of AIC and BIC, respectively. Note that the values of different information criteria (e.g., AIC and BIC) must not be compared because AIC and BIC are different feature quantities.

The first term of AIC and BIC is calculated as follows. For simplicity, we approximated that the noise on the measured spectrum has a uniform Gaussian distribution. It is known that the noise observed in XPS spectra actually follows a Poisson distribution. However, the Gaussian noise is a good approximation when the background intensity is sufficiently larger than the peak intensity. We assumed

that the noise at each spectral data point follows a uniform Gaussian distribution with variance σ^2 . The probability density, p_i , is then given as:

$$p_i = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - f_i)^2}{2\sigma^2}\right\}. \quad (8)$$

The likelihood of the whole spectrum is

$$L = \prod_{i=1}^N p_i. \quad (9)$$

The logarithm of the maximum likelihood \hat{L} can be obtained as:

$$-2 \log \hat{L} = N\{\log(2\pi\hat{\sigma}^2) + 1\}, \quad (10)$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (y_i - f_i)^2. \quad (11)$$

Note that when using AIC or BIC to evaluate the model selection of XPS spectra, we need to align the energy range and the number of spectral data points N to compare all optimized models.

3. Calculation results

We applied this model selection to 143 XPS experimental spectra. The experimental spectral data were obtained from three sources: (1) our measurement data, (2) the database in the MultiPak (ULVAC-PHI) software, and (3) the database in the COMPRO software. We also applied our method to artificial spectrum data with small and large noise, and present the results in Appendix A.

For XPS measurements, we used the AXIS-ULTRA DLD (delay-line detector, Shimadzu/Kratos) equipped with the Al K α (1486.6 eV) monochromatic X-ray source. We applied our method to 12 measurement data of Bi-based high-temperature superconductor Bi₂Sr₂Ca_{n-1}Cu_nO_y and other metal oxides, such as CuO and Fe₂O₃. We used typical core level spectra of 81 elements excited by Al K α from ³Li to ⁹²U in the XPS spectrum database of the MultiPak software (ULVAC-PHI Inc.) [14]. From the XPS spectral database provided in the COMPRO software [15], we selected 50 XPS spectra of oxides, such as SiO₂, Al₂O₃, PbO, NiO, and Dy₂O₃, and organic polymers, such as polyethylene terephthalate (PET) and polyvinyl methyl ketone (PVMK). Among these 143 spectra, we will show the analysis results of C1s of carbon contamination on a copper oxide surface and the valence spectrum of an organic polymer.

The specification of the PC was as follows: the OS was Windows 10 (64bit, desktop type), the processor was Intel Xeon CPU E5-2620 v4 (2.10 GHz), and the RAM was 32.0 GB. Regarding the computational time of our calculation, the case of the C1s spectrum took 5.4 sec in total: 0.3 sec to create the initial models and 5.1 sec to optimize the parameters for all the models. The case of the valence spectrum took 4.4 sec in total: 0.3 sec to create the initial models and 4.1 sec to optimize the parameters for all the models.

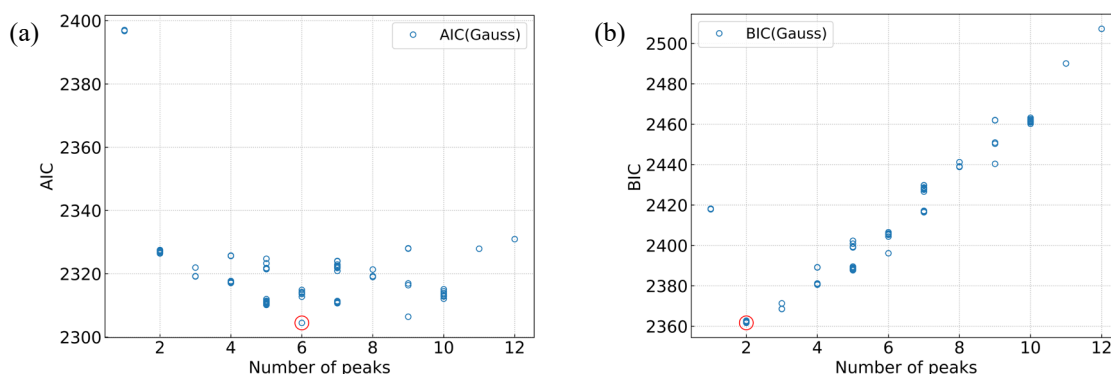


Fig. 2. AIC (a) and BIC (b) values as a function of the number of peaks for the C1s spectrum of carbon contamination of a copper oxide surface.

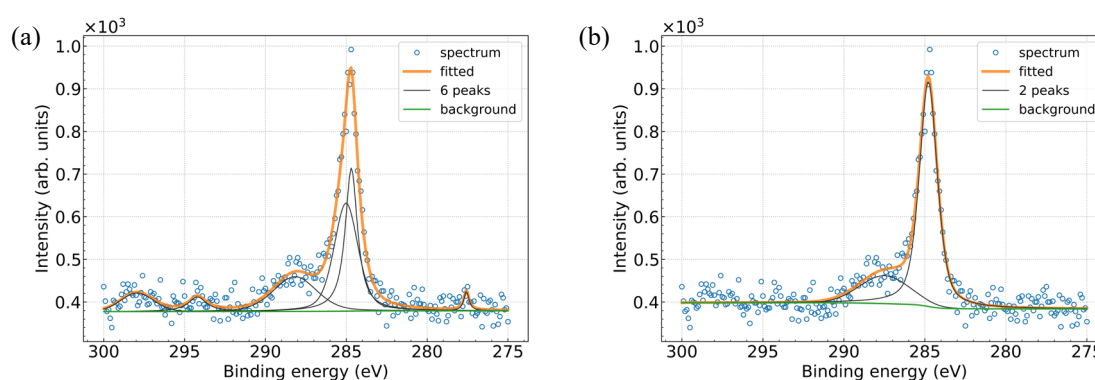


Fig. 3. Fitting results when minimizing AIC (a) and BIC (b) corresponding to the circled models in Fig. 2(a) and Fig. 2(b), respectively.

3.1. C1s spectrum of carbon contamination on copper oxide surface

Application of our method generated 155 different optimization models. The AIC and BIC values as a function of the number of peaks of each model are plotted in Fig. 2.

The peak numbers are systematically varied by changing the degree of the smoothing. One can see that there is a minimal point for each of the AIC and BIC diagrams. This tendency can be explained by the tradeoff between the negative logarithmic likelihood [first term in Eq. (1)] and the penalty term [second term in Eq. (1)]. When the number of peaks is lower, the reproduction of the measured spectrum by the model function becomes worse, and the negative logarithmic likelihood of the first term of the information criterion increases. Conversely, when the number of peaks is higher, the reproduction of the spectrum is improved and the negative log likelihood becomes smaller and approaches a constant value. However, the penalty term will continue to increase as the number of peaks increases, so the information criterion will also become large. As a result, the information criterion diagram reaches a minimum value when the likelihood and penalty terms are moderately sized.

Fig. 2 also shows that there are several or many AIC/BIC values for each peak number. The reason for this is that even with the same number of peaks, there are various local solutions depending on the initial values of peak position, height, width, and Lorentz-Gauss mixing ratio.

In Fig. 2, the data point indicated by a red circle corresponds to the model function that minimizes AIC or BIC for the C1s example spectrum. A model with nine peaks was adopted when we used AIC for model selection, while a model with two peaks was adopted when we used BIC. Looking at the AIC diagram, we can see that the minimum values of AIC at peak numbers five, six and nine are almost the same values. In those cases, the likelihood term and the penalty term in the equation of AIC are roughly balanced in a wide range of number of peaks, and therefore it is difficult to determine the best model functions. However, when using BIC, there is a clear minimum value when the number of peaks is two. That is because the slope of the penalty term in BIC is larger than in AIC. The variance of BIC values is large when the number of peaks is two, which means that many local solutions are searched through our approach. The best model among them is the one with the minimum BIC value.

Fig. 3 shows the fitting results corresponding to the red circles in Fig. 2 when each of AIC and BIC were minimized. Both AIC and BIC detect the major peaks in an energy range of 282 eV to 291 eV. There are also minor peak structures that would be difficult to determine visually by an analyst. In such cases, BIC does not recognize them as peaks, but AIC does. Generally, BIC adopts a model with a smaller number of peaks than AIC. This is because the number of data points of the spectrum in Eq. (7) is sufficiently large, and the penalty term for BIC, $m \log N$, is larger than the AIC penalty term of $2m$.

3.2. Valence spectrum of polyvinyl methyl ketone

Next, we treated the valence spectrum of the organic polymer polyvinyl methyl ketone as an example of a spectrum with intense noise and some structures over a wide energy range. Applying our method to this spectrum, we obtained the results shown in Fig. 4. A point indicated by a red circle corresponds to the model function that minimizes AIC or BIC, and the fitting results of the model function are as shown in Fig. 5. As in the case of Fig. 3, both AIC and BIC detect the major peaks, and BIC adopts a model with a smaller number of peaks than AIC.

3.3. Summary of other calculation results

We confirmed the effectiveness of our method for other spectra. In 137 XPS spectra (96% of the sample set), we find that the model selected by BIC has the following features: All major peaks that can be easily identified visually are detected, and peaks that are difficult to distinguish visually are not detected. That is to say, BIC tends to choose the simplest model for XPS spectra.

If the background of a spectrum is lower on the high binding energy side than on the low binding energy side (because of the matrix background), the Shirley method cannot be applied in principle. Therefore, such a spectrum was excluded from the current analysis.

Another approach might be employed, namely using Markov chain Monte Carlo (MCMC) methods

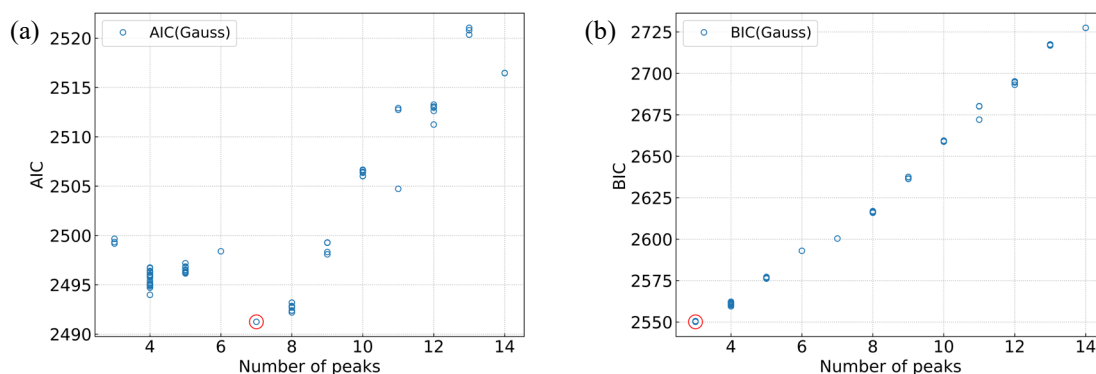


Fig. 4. AIC (a) and BIC (b) values as a function of the number of peaks for the valence spectrum of polyvinyl methyl ketone.

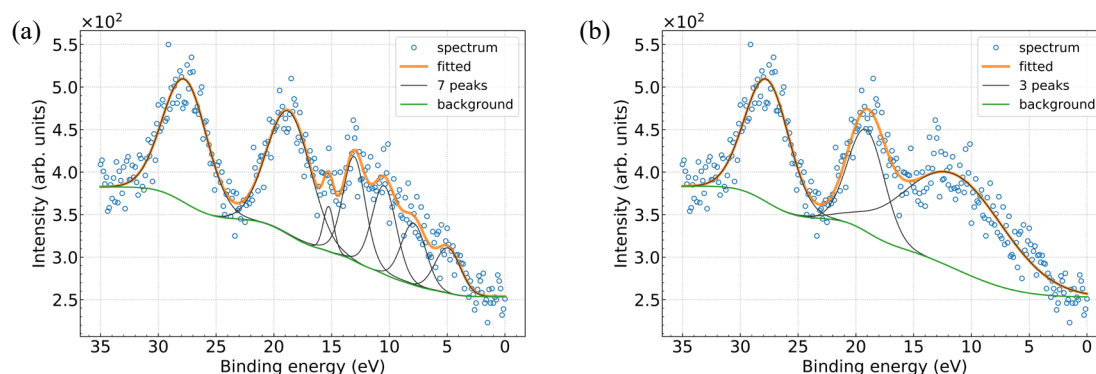


Fig. 5. Fitting results when minimizing AIC (a) and BIC (b) corresponding to the circled models in Fig. 4(a) and Fig. 4(b), respectively.

and Bayesian estimation to perform the model selection more correctly than our methodology. However, such a method is time consuming and our approach is useful because the calculation time is short enough for practical use (<20 s for 87 of the 143 experimental spectra studied).

Note that the current algorithm is useful for optimizing statistically and automatically the model functions but it may simplify them from the view point of physical property knowledge of the electronic structure of the target material and the generation process of photoelectrons. Our method is, therefore, useful to automatically present statistically valid candidate models prior to examining the XPS spectrum based on physical knowledge.

4. Discussion

We have previously reported on an active Shirley algorithm [5]. Using this conventional method, we analyzed the C1s and valence spectra with a poor signal to noise ratio introduced earlier and obtained the results shown in Fig. 6. From this, a large number of sharp peaks, which can be regarded as noise, remained.

By the conventional method, only one smoothing procedure was given, so we obtained only one optimization model. Therefore, there is a disadvantage that the optimization is easily trapped by the local solution including sharp peaks originated from noise. However, in our current method we systematically elevated the smoothing degree and obtained many optimization models. We then applied the information criterion for those optimization models. In a model including sharp peaks, the number of peaks increases, and the penalty term also increases compared with the small degree of improvement of the likelihood term. As a result, an optimized model consisting only of major peaks tends to be selected. However, it should be noted that, depending on the spectral shape and signal to noise ratio, there are rare cases where an optimized model does not fully exclude sharp peaks, not only by AIC but also by BIC.

Regarding the application limits of our method, the Shirley background cannot be drawn when the intensity on the high binding energy side is lower than the intensity on the low binding energy side. In such a case, our program does not work, but generally there is a way to draw a straight line background.

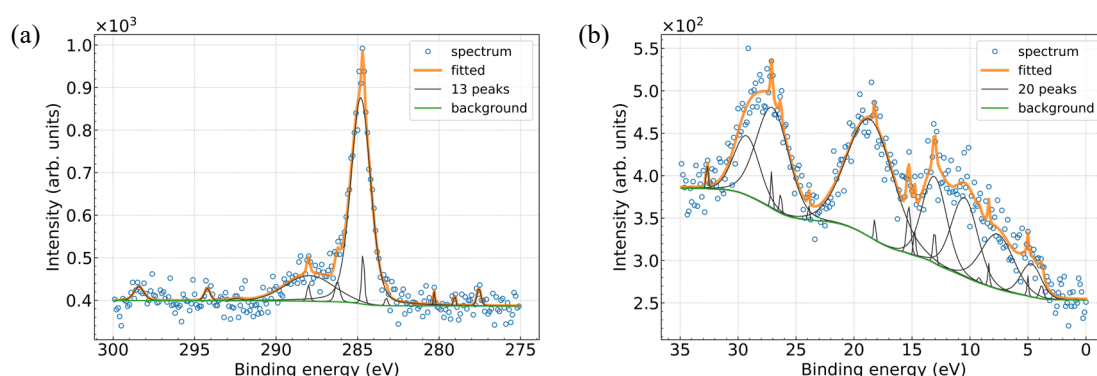


Fig. 6. Fitting results when using the previous active Shirley method for the C1s spectrum of carbon contamination on copper oxide surface (a) and for the valence spectrum of polyvinyl methyl ketone (b).

As shown in section 2.2, we adopted the pseudo-Voigt function in our algorithm. Of course, it is important to use the convoluted Voigt function as a basis function, which will be future work.

5. Conclusions

We have developed an algorithm using the Levenberg-Marquardt method as a mathematical solution based on the active Shirley method for automatic peaks and background separation of XPS spectra. For a spectrum with a complex shape or a spectrum with a poor signal to noise ratio, we found there are cases where the previous active Shirley algorithm can only obtain a local solution and cannot obtain a reasonable solution from the viewpoint of XPS analysis experts. We therefore developed the following new algorithm to overcome this issue. Because the Levenberg-Marquardt method generally obtains local solutions dependent on the initial parameters, we decided to obtain many local solution groups from a large number of initial parameter groups depending on the degree of smoothing. Then, a better or best solution is selected from the local solution groups. Specifically, we systematically perform round robin from weak to strong smoothing on a spectrum to obtain many fitting results. Subsequently, a good model is selected by applying the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) to the fitting results. As a result of applying this algorithm to measured XPS spectra, we found that using the AIC, a model with a large number of peaks and a good agreement with the spectrum was selected. And we found that using the BIC, a simple model with reasonably good agreement and a smaller number of peaks was selected. The model selected by BIC is closer to the result of peak fitting performed by XPS analysis experts.

Acknowledgments

We are grateful to Dr. S. Tanuma from NIMS for his useful comments. We thank Iain Mackie, PhD, and Maxine Garcia, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

References

- [1] D.A. Shirley, High-resolution X-ray photoemission spectrum of the valence bands of gold, *Phys. Rev. B* 5 (1972) 4709.
- [2] A. Proctor, P.M.A. Sherwood, Data analysis techniques in X-ray photoelectron spectroscopy, *Anal. Chem.* 54 (1982) 13–19.
- [3] S. Tougaard, Low energy inelastic electron scattering properties of noble and transition metals, *Solid State Commun.* 61 (1987) 547–549.
- [4] A. Herrera-Gomez, M. Bravo-Sanchez, O. Ceballos-Sanchez, M.O. Vazquez-Lepe, Practical methods

for background subtraction in photoemission spectra, *Surf. Interface Anal.* 46 (2016) 897–905.

[5] R. Matsumoto, Y. Nishizawa, N. Kataoka, H. Tanaka, H. Yoshikawa, S. Tanuma, K. Yoshihara, *J. Electron Spectrosc. Relat. Phenom.* 207 (2016) 55–59.

[6] R. Matsumoto, Y. Nishizawa, N. Kataoka, H. Tanaka, H. Yoshikawa, S. Tanuma, K. Yoshihara, Automatic background estimation and quantitative analysis for XPS spectrum by active Shirley method, *J. Surf. Anal.* 22 (2016) 155–167.

[7] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least-squares procedures, *Anal. Chem.* 36 (1964) 1627–1639.

[8] R. Hesse, P. Streubel, R. Szargan, Product or sum: comparative tests of Voigt, and product or sum of Gaussian and Lorentzian functions in the fitting of synthetic Voigt-based X-ray photoelectron spectra, *Surface and Interface Analysis* 39 (2007) 381–391.

[9] K. Levenberg, A method for the solution of certain non-linear problems in least squares, *Q. Appl. Math.* 2 (1944) 164–168.

[10] D. W. Marquardt, An algorithm for least squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.* 11 (1963) 431–441.

[11] R. Fletcher, A Modified Marquardt Subroutine for Nonlinear Least Squares, AERE-R 6799, Harwell 1971.

[12] H. Akaike, A new look at the statistical model identification, *IEEE T. Automat. Contr.* 19 (1974) 716–723.

[13] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* 6 (1978) 461–464.

[14] ULVAC-PHI Inc., MultiPak, <https://www.ulvac-phi.com/en/products/database-books/phi-multipak/>, 2019 (accessed 17 March 2019).

[15] K. Yoshihara, The introduction of common data processing system version 12, *J. Surf. Anal.* 23 (2017) 138–148.

Appendix. Results for artificial spectrum data

To compare two cases in which the statistical noise is small and large, we show the results of applying our proposed method to the artificial spectrum data. Using parameters similar to the C1s spectrum shown in Fig. 2, we prepared two spectra that consisted of two peaks, a prominent main peak and a sub-peak, that differed only in the magnitude of the noise. The peak parameters of the main peak and sub-peak were $(A_1, \mu_1, w_1, r_1) = (520, 284.8, 0.66, 0.5)$, $(A_2, \mu_2, w_2, r_2) = (60, 287.5, 1.83, 0.0)$, and the background parameter was $(I_S, I_E) = (400.0, 380.0)$, using the symbols defined in section 2.2. The range of binding energy was from 300 eV to 275 eV, the energy step was 0.1 eV, and the number of data points was 251. We applied noise according to the Gaussian distribution, and generated two spectra by selecting two standard deviation values of the Gaussian distribution: 2.0 and 50.0. The corresponding S/N ratios were

500 and 20, respectively. We show the generated artificial spectra in Figs. 7(a) and (b).

Figs. 7(c) to (f) show the results applied to a spectrum with small and large noise. When the noise was small, the number of peaks to be searched was 1 to 6. When the noise was large, the number of peaks to be searched was 1 to 17. We found that the number of peaks searched was limited when the noise was small. Figs. 7(c) and (d) show that when the noise was small, the BIC values at some peak number were concentrated at the same value. By contrast, when the noise was large, we observed that the BIC values were scattered. This means that when the noise was small, it was easy to converge to the global solution, and conversely, when the noise was large, optimization was difficult and trapped in the local solution. Our method allowed model selection based on the BIC. When the noise was small, our method correctly selected the model with two peaks as the best BIC model. By contrast, when the noise was large, it was difficult to detect the sub-peak as a peak, and the model with one peak was selected as the best BIC model.

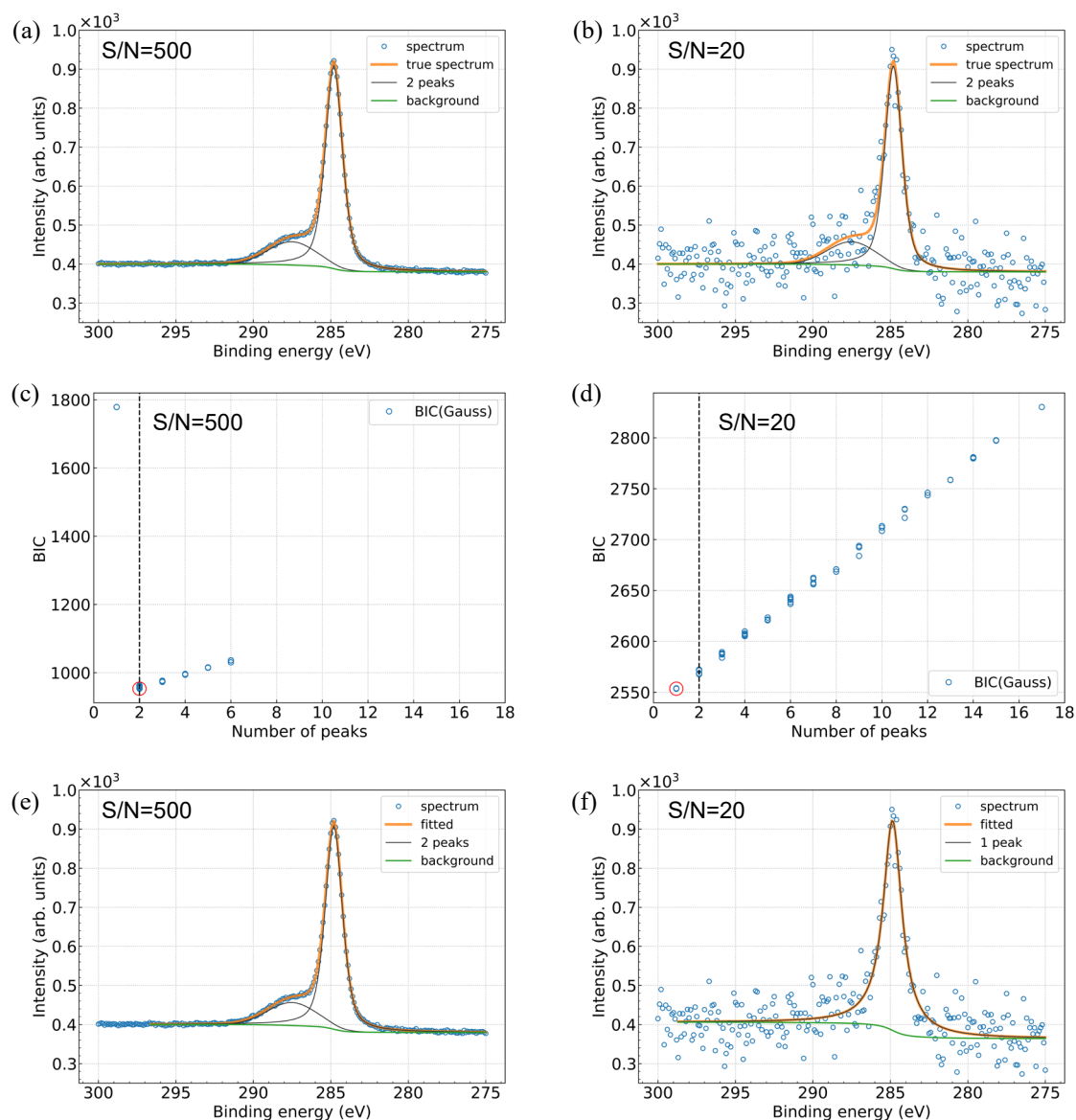


Fig. 7. (a) and (b) are two artificial spectra with small noise (S/N=500) and large noise (S/N=20), respectively. (c) and (d) are the results of BIC values as a function of the number of peaks for each artificial spectrum. The dashed lines show the true number of peaks 2. (e) and (f) are the fitting results when minimizing BIC corresponds to the circled models in (c) and (d), respectively.