Taylor & Francis
Taylor & Francis Group

Development of spectral decomposition based on Bayesian information criterion with estimation of confidence interval

Hiroshi Shinotsuka [a], Kenji Nagata [a], Hideki Yoshikawa [a], Yoh-Ichi Mototake [b], Hayaru Shouno [c] and Masato Okada [a,d]

[a]Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba, Japan;
[b]Research Center for Statistical Machine Learning, The Institute of Statistical Mathematics, Tachikawa, Japan;;
[c]Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu, Japan;;
[d]Graduate School of Frontier Science, The University of Tokyo, Kashiwa, Japan

ABSTRACT

We develop an automatic peak fitting algorithm using the Bayesian information criterion (BIC) fitting method with confidence-interval estimation in spectral decomposition. First, spectral decomposition is carried out by adopting the Bayesian exchange Monte Carlo method for various artificial spectral data, and the confidence interval of fitting parameters is evaluated. From the results, an approximated model formula that expresses the confidence interval of parameters and the relationship between the peak-to-peak distance and the signal-to-noise ratio is derived. Next, for real spectral data, we compare the confidence interval of each peak parameter obtained using the Bayesian exchange Monte Carlo method with the confidence interval obtained from the BIC-fitting with the model selection function and the proposed approximated formula. We thus confirm that the parameter confidence intervals obtained using the two methods agree well. It is therefore possible to not only simply estimate the appropriate number of peaks by BIC-fitting but also obtain the confidence interval of fitting parameters.
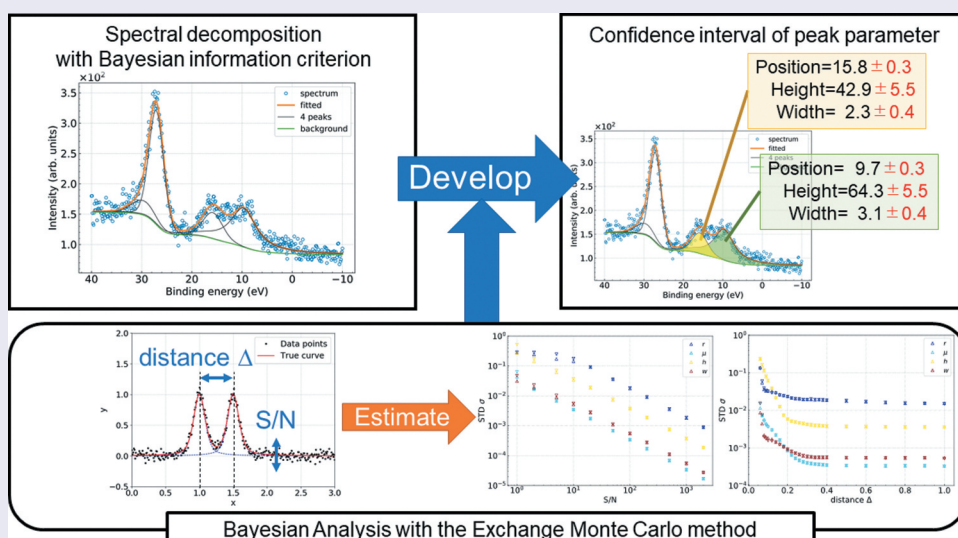
1. Introduction

High-throughput measurements have become increasingly important for the efficient development of science and technology, and there is an urgent need to accumulate large amounts of spectral data. 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 a rapid accumulation of large amounts of spectral data [1–3]. Matsumura et al. performed peak shift analysis of high-throughput XPS spectra using the expectation-maximization algorithm [4]. High-throughput data processing is therefore required for efficient spectral data analysis.

Peak fitting is performed in the analysis of XPS spectra. Such fitting is usually carried out using the gradient method. This technique faces three main problems. The first is that the technique tends to find

a local solution, the second is that the number of peaks cannot be estimated, and the third is that the confidence interval of fitting parameters cannot be evaluated. In the gradient method, the initial value of the parameter must first be given, and the result is readily affected by the initial value. Peak fitting requires the number of peaks to be determined at the beginning, but the gradient method does not show how many peaks are appropriate. The method is also susceptible to spectral noise, and although it is intuitively understood that the confidence interval of the fitting parameter is wide when there is much noise, there is no framework for evaluating the confidence interval.

By incorporating informatics knowledge, we previously developed a low-cost and efficient method of obtaining appropriate models in terms of not only the fitting of parameters but also the number of peaks, even though the developed method is based on the gradient method [5]. Having many initial models allows us to search for pseudo-global solutions, and a Bayesian information criterion (BIC) allows us to obtain a model with an appropriate number of peaks. We refer to this technique as BIC-fitting in the present paper. However, it remains difficult to evaluate the confidence intervals of the fitting parameters with this technique.

A spectrum decomposition technique based on Bayesian estimation has been proposed for quantitative evaluation of the confidence interval of fitting parameters [6]. This technique solves all three problems described above. By carrying out the model selection to a given spectrum on the basis of Bayesian estimation, we may be able to estimate not only peak parameters such as the peak position but also the number of peaks. Furthermore, when adopting this technique, global solutions can be searched for efficiently by performing optimization using algorithms in what is called the exchange Monte Carlo (EMC) method, even in the case that there is a local optimal solution. Through Bayesian estimation, all peak parameters can be optimized, and the confidence intervals of the parameters can also be determined using the standard deviation (STD) of the Bayesian posterior distribution. However, the EMC method has a huge computational cost and is difficult to use when analyzing many spectra.

In this study, therefore, we develop an algorithm to calculate the confidence interval of fitting parameters obtained by the EMC method from the results of BIC-fitting. By generating various spectral data on a computer and applying Bayesian estimation, we obtain the behavior of model selection and the STD of the Bayesian posterior distribution for each peak parameter by computer simulation. In particular, in this paper, we cover the peak-to-peak distance between two peaks and the signal-to-noise (S/N) ratio of spectral data. As a result, we succeed in

deriving an approximated model formula representing the relationship of the STD of the posterior distribution with the peak-to-peak distance and S/N ratio.

We also apply the approximated model formula to real spectra. The confidence interval of the peak parameters obtained using the EMC method is compared with that obtained by applying BIC-fitting to the approximated formula. As a result, it is confirmed that the parameter confidence interval obtained by the EMC method can be reproduced by BIC-fitting and using the approximated formula. This approximated formula is applicable to not only BIC-fitting but also other optimization methods and can be used to estimate the parameter confidence interval to the same extent as when adopting the EMC method.

2. Calculation methods

2.1. Fitting model: pseudo-Voigt function

We first describe the model function used in this study. We consider fitting spectral data $D = \{(x_i, y_i)\}_{i=1}^{n}$, where n is the number of spectral data points, by summing pseudo-Voigt functions $V(x; h, \mu, w, r)$:

$$y = f(x; \theta) = \sum_{k=1}^{K} V(x; h_k, \mu_k, w_k, r_k). \quad (1)$$

The pseudo-Voigt function $V(x; h, \mu, w, r)$ is frequently used in spectral decomposition. We here adopt the sum type of the pseudo-Voigt function, defined as a linear combination of the Gaussian and Lorentzian functions:

$$V(x; h, \mu, w, r) = h\{r\tilde{L}(x; \mu, w) + (1-r)\tilde{G}(x; \mu, w)\}, \quad (2)$$

$$\tilde{G}(x; \mu, w) = e^{-(\log 2)\left(\frac{x-\mu}{w}\right)^2} = 2^{-\left(\frac{x-\mu}{w}\right)^2}, \quad (3)$$

$$\tilde{L}(x; \mu, w) = \frac{1}{1 + \left(\frac{x-\mu}{w}\right)^2}, \quad (4)$$

where K is the number of peaks. The fitting parameters are $\theta = \{h_k, \mu_k, w_k, r_k\}_{k=1}^{K}$, where h_k is the peak height, μ_k is the peak position, w_k is the half width at half maximum (HWHM) of the peak, and r_k is the Lorentz–Gauss mixing ratio of the pseudo-Voigt functions. In the peak fitting of XPS, the appropriate basis function is the Voigt function defined by the convolution of a Lorentzian function derived from the natural width and a Gaussian function derived from a device. Indeed, the pseudo-Voigt function, an approximated form of the Voigt function, is commonly used because of computational difficulty in peak fitting with the Voigt function [7].

A least-squares method is often used to optimize fitting parameters. In this method, parameters are obtained so as to minimize the error function $E(\theta)$

representing the difference between the model function and the spectral data $D = \{(x_i, y_i)\}_{i=1}^n$:

$$E(\theta) = \frac{1}{2n} \sum_{i=1}^n \{y_i - f(x_i; \theta)\}^2. \tag{5}$$

In spectral decomposition, this problem becomes a nonlinear least-squares problem, and it is difficult to derive such an optimum solution analytically. It is therefore common to find the parameter θ that minimizes the error function based on the gradient method. However, there is a problem that the fitting result is easily trapped into a local solution depending on the selection of initial values. In addition, it is impossible to objectively determine the number of peaks K from the data. The gradient method also has a problem that the confidence interval of the fitting parameter cannot be obtained. Bayesian estimation can solve these problems as we will see below [6].

2.2. Bayesian spectral deconvolution

Bayesian estimation is a framework in which the process of generating data in a probabilistic model is formulated and an estimation is made by tracing back the causal relationship using the Bayesian theorem [6,8]. By combining Bayesian estimation with the exchange Monte Carlo (EMC) method, we may be able to not only perform spectral deconvolution but also obtain confidence intervals for fitting parameters through Bayesian posterior probabilities. It is also possible to select a good model by comparing the Bayesian free energies of different models with different numbers of peaks. In this study, we call this method the Bayesian EMC method. Details are shown in Appendix A.

2.3. BIC-fitting

It is usually difficult to analytically evaluate the Bayesian free energy for model selection because it requires multiple integration on the parameter space. The BIC is obtained by approximating the multiple integration under the assumption that the likelihood function can be approximated with a Gaussian distribution for all parameters. The BIC is expressed as the sum of a likelihood term and a penalty term, and the model is selected on the basis of the trade-off between models.

We have developed a low-cost and efficient method of obtaining appropriate models in terms of not only the fitting parameters but also the number of peaks using many initial models and the BIC [5]. The method searches many initial fitting models by changing the degree of smoothing, and then optimizes the peak parameters using the modified Levenberg–Marquardt method [9–11], which is one of the

gradient methods. The goodness of the optimized models is ranked on the basis of the BIC, written as

$$\mathrm{BIC} = -2\log\widehat{L} + m \log n \tag{6}$$

where $\hat{L}$ is the maximum likelihood calculated from the likelihood between the measured spectrum $\{y_i\}_{i=1}^n$ and the model function $\{f(x_i; \theta)\}_{i=1}^n$ obtained as a result of optimization. m is the number of parameters included in the model function. When we ignore the background, $m = 4K$ is obtained using the number of peaks K. The logarithm of the maximum likelihood $\hat{L}$ can be obtained as

$$-2 \log \hat{L} = n\{\log(2\pi\hat{\sigma}^2) + 1\}, \tag{7}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i; \theta))^2. \tag{8}$$

Using the BIC values of optimized models as a criterion for model selection, we can select a simple model with reasonably good agreement and a moderate number of peaks. We hereafter refer to this technique as BIC-fitting in this paper.

BIC-fitting can perform spectral fitting and model selection, but cannot obtain confidence intervals for fitting parameters. The purpose of this study is to extend the BIC-fitting method so that the confidence intervals of the fitting parameters can be obtained at the same time, using the results of simulation by the Bayesian EMC method.

Models used in spectral decomposition are singular models whose parameters and properties do not correspond to each other [12,13]. In this case, the BIC may have penalty terms different from those in the exact evaluation of free energy, and the approximation of the BIC may affect the result of model selection. In Section 4, we compare the model obtained using the Bayesian EMC method with the model obtained from BIC-fitting, targeting the analysis of the measured XPS spectrum, and we discuss the effectiveness of BIC-fitting.

3. Simulation with artificial spectra

In the present study, we use the Bayesian spectral decomposition framework described in Section 2 to clarify the effects of the peak-to-peak distance in the true spectra and the S/N ratio of the data on the confidence interval of the estimated parameters. In this section, we discuss the simulations performed for verification.

3.1. Settings

In the simulation, we use spectral data artificially measured by computer simulation. For the data set $D = \{(x_i, y_i)\}_{i=1}^n$, we take the number of data $n = 301$

and x_i between $[0.0, 3.0]$ in steps of 0.01. Assuming that the number of peaks $K = 2$, we define the true spectral function $f(x; \theta)$ used for data generation as

$$f(x; \theta^*) = \sum_{k=1}^{2} V(x; h_k^*, \mu_k^*, w_k^*, r_k^*). \qquad (9)$$

Here, the true parameters θ^* are $h_1^* = h_2^* = 1.0$, $w_1^* = w_2^* = 0.1$, and $r_1^* = r_2^* = 0.5$. We also fix $\mu_1^* = 1.0$ and set $\mu_2^* = \mu_1^* + \Delta$ at various peak-to-peak distances Δ for discussion. We generate data in 32 patterns for the range $0.0 \leq \Delta \leq 1.0$. We add noise that follows a Gaussian distribution with zero mean and variance σ^2 to the data and prepare 14 patterns of σ values in the range $[0.0005, 10.0]$. Hence, the total number of prepared data sets is $32 \times 14 = 448$. Three examples of artificially measured spectral data are shown in Figure 1, where Figure 1(a–c) present spectral data with settings of $(S/N, \Delta) = (20.0, 0.5)$, $(20.0, 0.1)$, and $(0.2, 0.5)$, respectively.

The S/N ratio of the data can be defined using the value of the noise level σ. The intensity of the true spectrum is $h_1^* = h_2^* = 1.0$, and the signal intensity is thus 1.0. In this study, we define the S/N ratio as $S/N = h_1^*/\sigma = 1.0/\sigma$. The range of the S/N ratios in the simulation is $[0.1, 2000]$.

We perform a Bayesian estimation for all data sets. We assume that the candidate numbers of peaks K are one and two. The prior distribution $p(K)$ for the number of peaks is thus defined as

$$p(K) = \begin{cases} 0.5 & (\text{if } K = 1 \text{ or } 2) \\ 0 & (\text{otherwise}) \end{cases}. \qquad (10)$$

The prior distribution for each parameter is set as

$$p(h_1) = p(h_2) = \text{Gamma}(h_i; 2.0, 1.0), \qquad (11)$$

$$p(\mu_1) = p(\mu_2) = \mathcal{N}(\mu_i; 1.5, 0.2), \qquad (12)$$

$$p(w_1) = p(w_2) = \text{Gamma}(w_i; 2.0, 0.5), \qquad (13)$$

$$p(r_1) = p(r_2) = U(r_i; 0.0, 1.0), \qquad (14)$$

where $\text{Gamma}(x; \eta, \theta), \mathcal{N}(x; \nu, \xi)$, and $U(x; x_{Min}, x_{Max})$ are respectively the gamma distribution, Gaussian distribution, and uniform distribution:

$$\text{Gamma}(x; \eta, \theta) = \frac{1}{\Gamma(\eta)\theta^\eta} x^{\eta-1} e^{-x/\theta}, \qquad (15)$$

$$\mathcal{N}(x; \nu, \xi) = \sqrt{\frac{1}{2\pi\xi^2}} \exp\left(-\frac{(x-\nu)^2}{2\xi^2}\right). \qquad (16)$$

$$U(x; x_{Min}, x_{Max}) = \begin{cases} \frac{1}{x_{Max}-x_{Min}} & (\text{if } x_{Min} \leq x \leq x_{Max}) \\ 0 & (\text{otherwise}) \end{cases}. \qquad (17)$$

The prior distribution of the parameters of interest can be given by the analyst. However, it is possible to
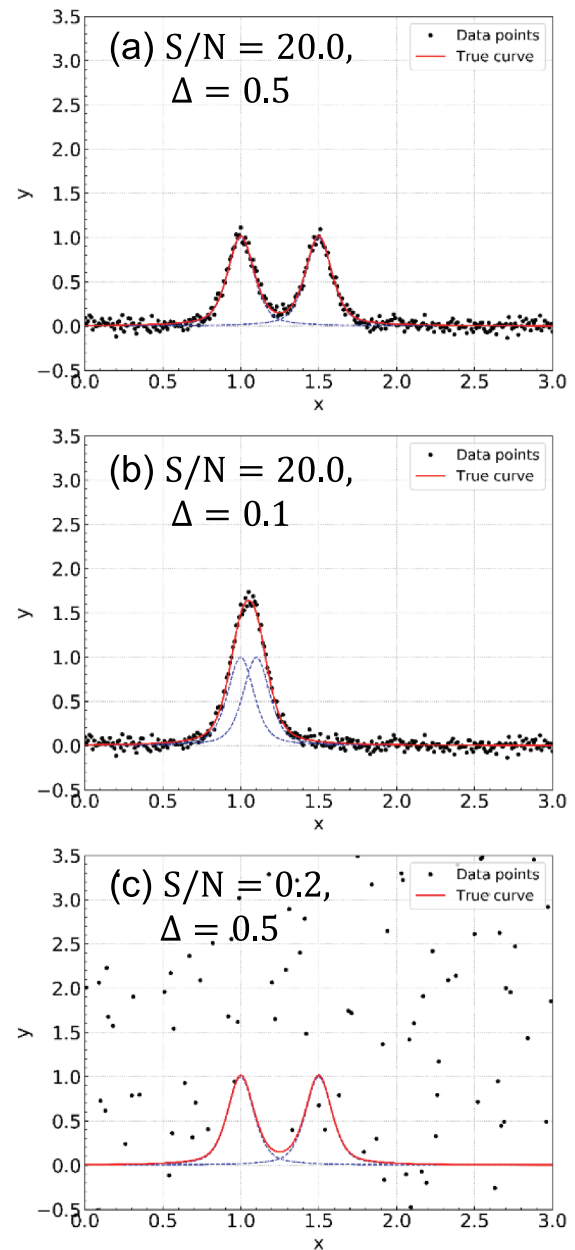

Figure 1. Three examples of artificially measured spectra with Gaussian noise. The solid line is the true curve $f(x; \theta^*)$ and the dots are the artificially measured spectral data: (a) $S/N = 20.0$ and $\Delta = 0.5$, (b) $S/N = 20.0$ and $\Delta = 0.1$, and (c) $S/N = 0.2$ and $\Delta = 0.5$.

predict an appropriate distribution shape by considering the characteristics of the spectrum. For example, the peak height and width must be positive. We can infer the approximate range of peak height and width values by looking at the structure of the spectrum. We have adopted the gamma distribution as the distribution function with those characteristics. On the other hand, the peak position can move either to the positive or negative side. We then adopted the Gaussian distribution without any special boundary in the prior distribution. Since it is clear that the peak position is between 1 and 2, we set a Gaussian distribution with an average of 1.5 and an STD of 0.2 as in Eq. (12). The Lorentz–Gauss mixing ratio can

Table 1. Settings of the inverse temperature $\{\beta_m\}_{m=1}^M$ corresponding to Eq. (18).

σ	0.0005	0.001	0.002	0.005	0.01	0.02	0.05	0.1	0.2	0.5	1	2	5	10
S/N	2000	1000	500	200	100	50	20	10	5	2	1	0.5	0.2	0.1
γ	1.2	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
M	128	64	64	64	44	44	32	32	32	32	32	32	32	32

range from 0 to 1 by definition. In this simulation, we decided not to impose any further constraints and adopted a uniform distribution for the Lorentz–Gauss mixing ratio.

As settings of the Monte Carlo method, we use 50,000 Monte Carlo steps (MCSs) as a burn-in and then use 30,000 MCSs for sampling.

The inverse temperatures $\{\beta_m\}_{m=1}^M$ in the EMC method are defined as [14]

$$\beta_m = \begin{cases} 0 & (m = 1) \\ \gamma^{m-M} & (m \neq 1) \end{cases}. \quad (18)$$

The value of γ and the total number of inverse temperatures M in the equation are set according to the level of noise σ added to the data (Table 1).

3.2. Results of Bayesian EMC method and derivation of approximated formula

We first show the results of model selection through Bayesian estimation. The results of model selection corresponding to the spectral data in Figure 1 are shown in Figure 2. The straight line represents the free energy $F(K)$ and the histogram represents the posterior probability $p(K|D)$. It is seen that the correct number of peaks $K = 2$ is estimated in Figure 2(a), whereas $K = 1$ is estimated in Figure 2(b). In the case of Figure 2(b), the peak-to-peak distance Δ is too small to extract information dividing the spectrum into two peaks at the given S/N ratio. In the case of Figure 2(c), there is no significant difference between $p(K = 1|D)$ and $p(K = 2|D)$, and we cannot estimate the number of peaks because the S/N ratio is too small. Therefore, by performing model selection, we expect that we can determine the peak-to-peak distance Δ and S/N ratio that are necessary for estimating the correct structure from the spectral data.

We then perform model selection for various peak-to-peak distances Δ and S/N ratios. Results are shown in Figure 3. The abscissa represents the S/N ratio and the ordinate represents the peak-to-peak distance Δ. The values in the figure indicate the posterior probability $p(K = 2|D)$ for $K = 2$. It is seen that when S/N < 0.5, the number of peaks cannot be estimated regardless of the peak-to-peak distance. Furthermore, we can clearly classify the region where $K = 1$ or 2. We next discuss the posterior distributions for the parameters θ. The posterior probabilities for each parameter when Bayesian estimation is performed on the spectral data in Figure 1(a) are shown in Figure 4. The dashed lines in the figures indicate the true
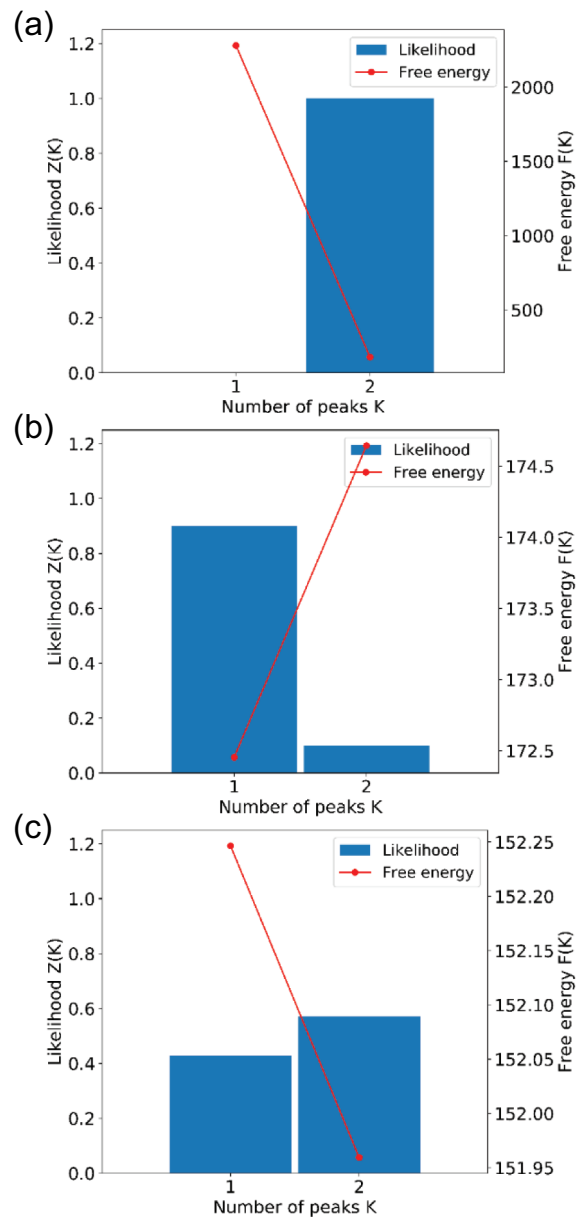

Figure 2. (a)–(c) Results of model selection by Bayesian estimation respectively corresponding to spectral data in Figure 1(a)–(c).

parameter values corresponding to the artificially measured spectral data. In the case of the peak position, the histogram shows the differences in peak positions μ_1 and μ_2 from their true values μ_1^* and μ_2^*, respectively. The results show that each parameter is estimated with good accuracy in that the distribution roughly includes the true value and appears similarly to a Gaussian distribution. As exceptions, the HWHM w_1 and Lorentz–Gauss mixing ratio r_1 for $k = 1$ are distributed away from the true values. This might be due to the properties of the pseudo-Voigt functions
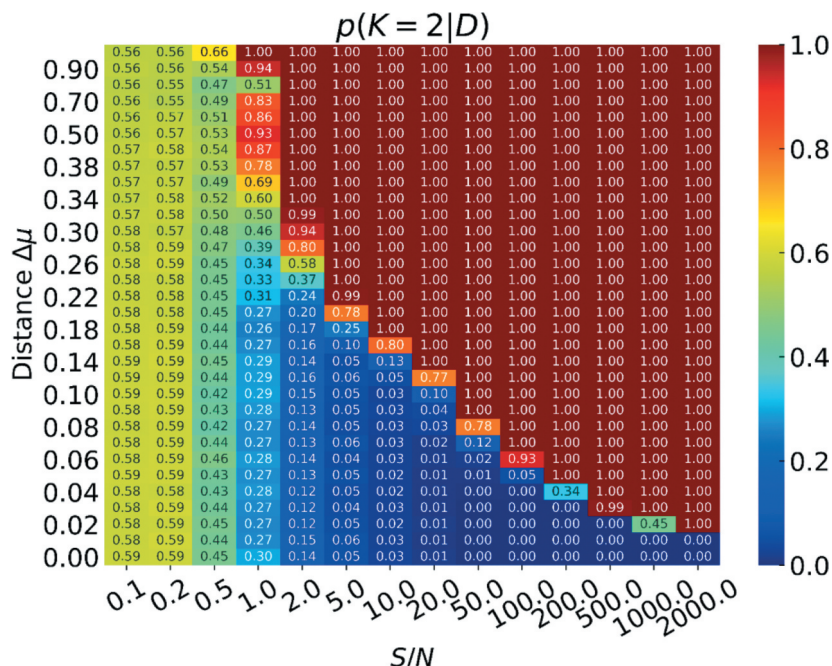
Figure 3. Results of model selection for various peak-to-peak distances Δ and S/N ratios. The values in the figure indicate the posterior probability $p(K = 2|D)$ for $K = 2$.
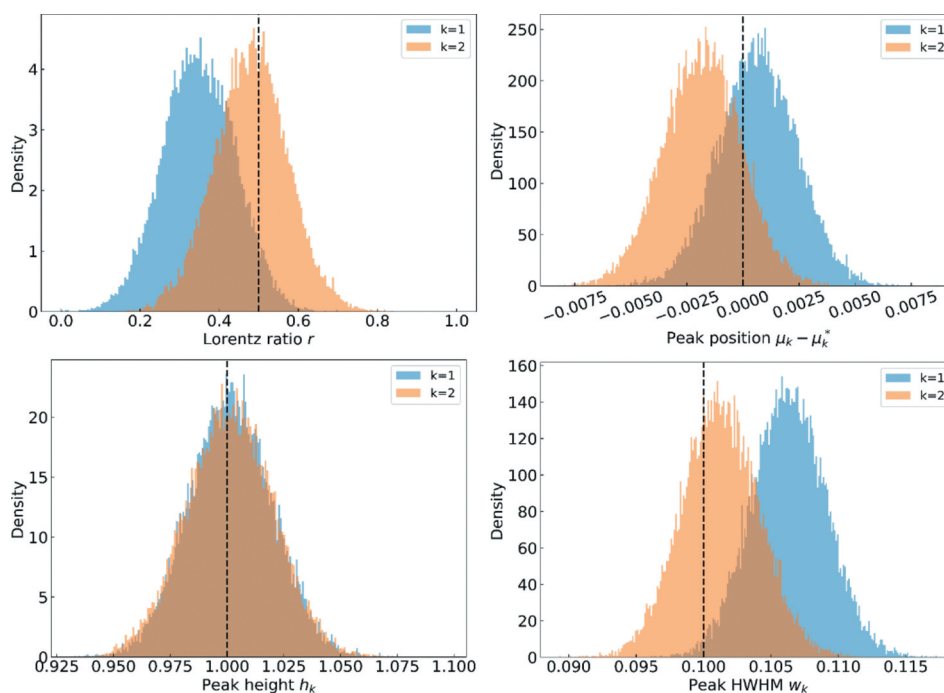

Figure 4. Posterior distribution of each parameter when Bayesian estimation is performed on the spectral data in Figure 1(a). Dashed lines indicate the true parameter values used to generate the spectral data.

and that there is large variability around the true values depending on the artificially measured spectral data. Details are given in Appendix B.

On the basis of the above results, we examine the STDs of the posterior distributions of various Δ values and S/N ratios to obtain the confidence intervals of these parameters. A plot of the relationship between the S/N ratio and the STD of the posterior distribution for Δ = 0.5 is

shown in Figure 5. The results suggest that estimation of the Lorentz–Gauss mixing ratio r is the most difficult because its STD is larger than the STDs of the other parameters. We also find that there is the relationship $s_j \propto \sigma$ between the STD s_j $(j \in \{r, \mu, h, w\})$ and the S/N ratio for any peak parameter. An exception is that the Lorentz–Gauss mixing ratio deviates from the relationship for S/N < 10^1. We believe that this is because the STD
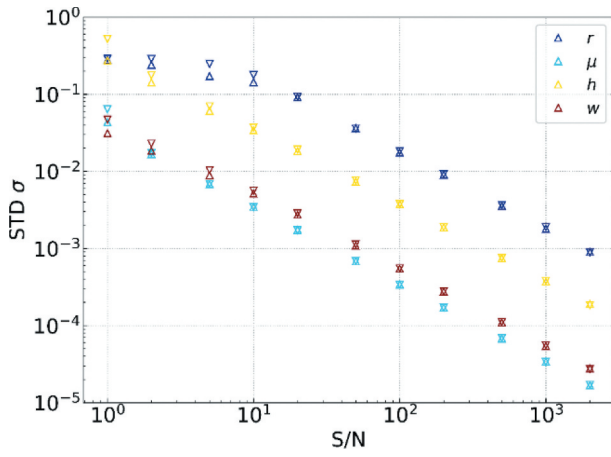
Figure 5. STD of the posterior distribution as a function of the S/N ratio for $\Delta = 0.5$. Triangles and inverted triangles are respectively the STDs of the parameters for the first and second peaks.
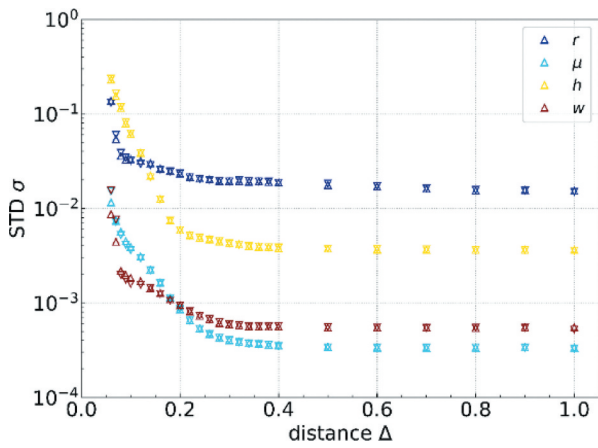

Figure 6. STD as a function of the peak-to-peak distance Δ for S/N = 100.0. Triangles and inverted triangles are respectively the STDs of the parameters for the first and second peaks.

of r in the posterior distribution cannot exceed the STD in the prior distribution defined by Eq. (14) $\left(1/\sqrt{12} = 0.2887\right)$. The STD as a function of the peak-to-peak distance Δ for S/N = 100.0 is shown in Figure 6.
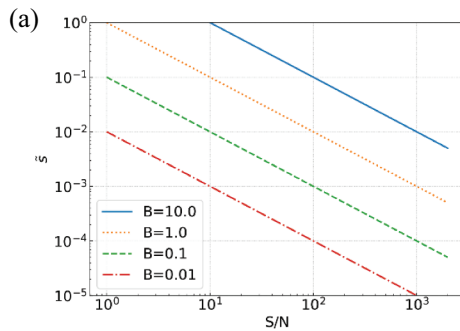
The results indicate that the STD is stable for all parameters when $\Delta > 0.4$. Conversely, when $\Delta \leq 0.4$, the STD is larger for smaller Δ, indicating that the estimation is unstable. The estimation thus becomes unstable when Δ is small because the two peaks overlap. According to this analysis, peak overlap begins to affect parameter estimation when Δ is less than about 4 times the HWHM.

As a result of simulations for various artificial spectral data, we found the following features. First, when the peak-to-peak distance Δ is large (especially $\Delta > 0.4$), the STD of any parameter has the relation $s_j = B_j \sigma$. Next, when $s_j/B_j\sigma$ was plotted against the peak-to-peak distance Δ for each parameter, we found that any curve with an arbitrary noise level σ can be approximated by a single curve. Also, as Δ decreases, s_j diverges to positive infinity. Although a function with such characteristics is not unique, it can be expressed as a power function + baseline as one of the candidates. However, it is necessary to adjust the position of the asymptote where the values diverge, the curvature of the curve, and the position where it reaches the baseline. Considering such requirements, we suggest an approximated formula as follows:

$$s_j\left(\Delta, \frac{\sigma}{h_0}\right) = \frac{\sigma}{h_0}B_j\left\{\left(\frac{\Delta - E_j}{D_j}\right)^{C_j} + 1\right\},$$

$$j = \{r, \mu, h, w\},$$
(19)

where h_0 is the true peak height and $\Delta = \Delta/w_0$ is the peak-to-peak distance scaled by the HWHM of the true peak w_0. Considering the characteristics of each peak parameter $j = \{r, \mu, A, w\}$, we define the scaled STDs as

$$\tilde{s}_r = s_r, \tilde{s}_\mu = \frac{s_\mu}{w_0}, \tilde{s}_h = \frac{s_h}{h_0}, \tilde{s}_w = \frac{s_w}{w_0},$$
(20)

and performed regression using Eq. (19). The fitting parameters in Eq. (19) are $\{B_j, C_j, D_j, E_j\}$. Figure 7 shows schematic diagrams of values obtained using Eq. (19). Figure 7(a) shows $\tilde{s}$ as a function of S/N $(= h_0/\sigma)$ for $B = 10.0, 1.0, 0.1$ and 0.01, where the
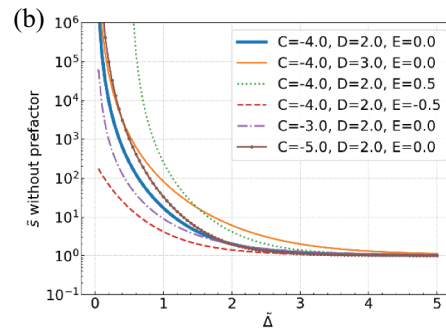
Figure 7. Schematic diagrams of values obtained using Eq. (19). (a) Scaled STD $\tilde{s}$ as a function of S/N $(= h_0/\sigma)$ for $B = 10.0, 1.0, 0.1$ and 0.01, where the peak-to-peak distance Δ is sufficiently large: $\Delta = 10.0, C = -3.0, D = 2.0$ and $E = 0.0$. (b) $\tilde{s}$ as a function of the peak-to-peak distance Δ for several conditions of $\{C, D, E\}$ when we set the prefactor of Eq. (19) $\sigma B/h_0 = 1$.
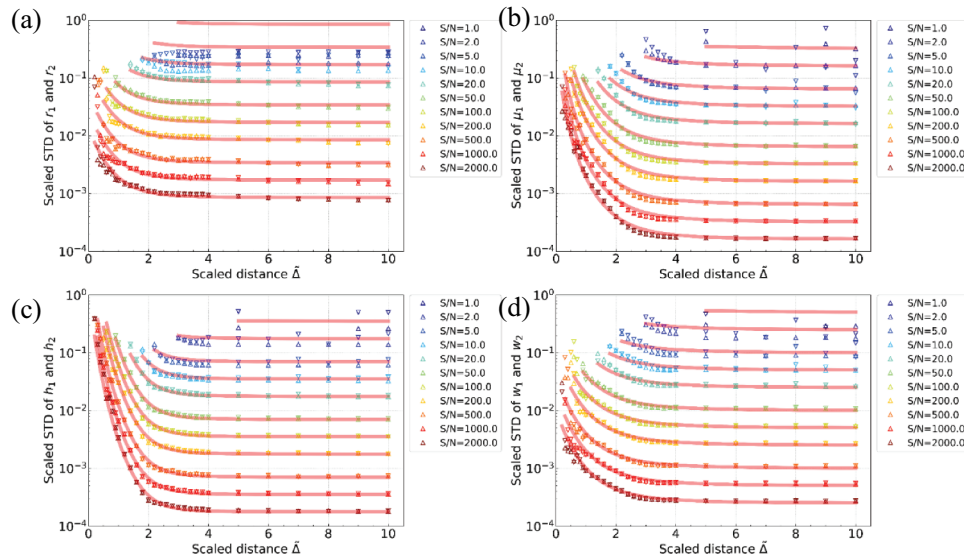
Figure 8. Results of regression using the fitting function in Eq. (19) of the STDs of the posterior distributions for each parameter. Parameters are the (a) Lorentz–Gauss mixing ratio r_k, (b) peak position μ_k, (c) peak height h_k, and (d) HWHMs of the peaks w_k.

peak-to-peak distance Δ is sufficiently large. Figure 7 (b) shows $\tilde{s}$ as a function of the peak-to-peak distance Δ for several conditions of $\{C, D, E\}$, when we set the prefactor of Eq. (19) $\sigma B / h_0 = 1$. By adjusting the parameters B, C, D and E, we can express the STD of any peak parameter. A description of how to optimize the features and parameters of these functions is given in Appendix D. In the present experiments, we fix $D_j = 2.5$. Although there is no reason to fix the value of D_j and it was decided heuristically, we confirmed that the fitting result was very good as shown in Figure 8. Using the same value for all parameters simplifies the formula and improves usability. The remaining three features $B_j, C_j,$ and E_j are used for regression, and we obtained them as shown in Table 2. The regression results are shown in Figure 8. The results show that approximate regression is achieved under all conditions. In addition, the range of operable S/N ratios of this equation differs between peak parameters; the ranges are presented in Appendix D.

As will be described later in detail, even when the Bayesian EMC method is not used, the confidence interval of the peak parameter can be estimated by using Eq. (19), after we obtain a fitted spectrum by an optimization method such as the BIC-fitting method.

As a further use of this formula, when the S/N ratio of the measured spectrum is known, we can estimate the peak-to-peak distance that achieves a certain confidence interval. Alternatively, when the peak-to-peak distance of the measured spectrum is known from the chemical shift, we can estimate the S/N ratio required to obtain the desired confidence interval of the parameter. This makes it possible to use Eq. (19) in experimental planning such as the setting of measurement time and energy resolution for individual measurements.

4. Simulation using a real spectrum

In this section, we analyze real XPS spectra to confirm the practicability of the approximated formula in the previous section. As an example of a real spectrum, we select a valence spectrum of SiO_2 from the XPS spectrum databases provided in COMPRO software [15] (Figure 9). The binding energy E_B varies from −10 to 40 eV with an energy step of 0.1 eV, resulting in 501 data points. There is a strong peak assigned to O2s in the vicinity of $E_B = 27$ eV. There is a peak structure derived from the hybridization of O2p, Si3s, and Si3p at $E_B = 10 - 18$ eV. We apply the Bayesian EMC method and BIC-fitting to this spectrum.

4.1. Model function with background

A real spectrum has a background. In addition to the superposition of the peaks of the pseudo-Voigt function of Eq. (1), the Shirley background $b(x; I_S, I_E)$ is added to the model function [5]:

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

(21)

Table 2. Values of the fitted parameters $2\{B_j, C_j, D_j, E_j\}$ in Eq. (19).

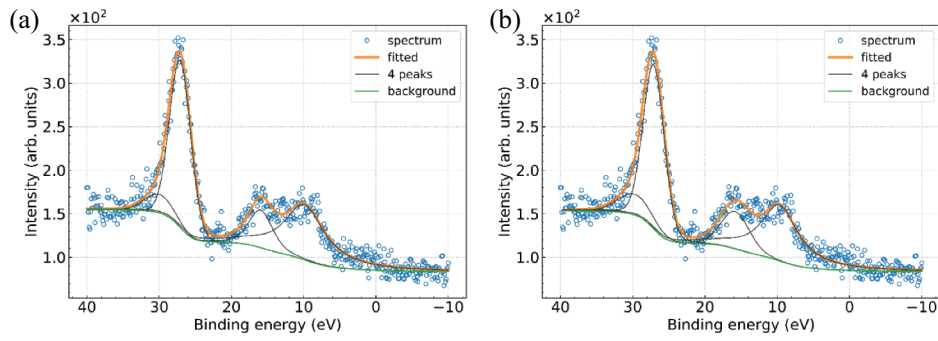| Parameter	B_j	μ_j	D_j (fixed)	E_j
r	1.708	−4.626	2.5	−1.394
μ	0.324	−3.271	2.5	−0.216
h	0.355	−5.756	2.5	−0.540
w	0.504	−3.158	2.5	−0.760

Figure 9. Fitted spectra from Bayesian estimation (a) and BIC-fitting (b) for the experimental valence spectrum of SiO$_2$. Open circles are the experimental spectrum, the orange line is the fitted spectrum, the green line is the background, and the black lines are all peaks above the background.

$$b(x; I_S, I_E) = (I_S - I_E)\frac{Q}{P + Q} + I_E, \quad (22)$$

where I_S and I_E are respectively the intensity of the spectrum on the high-binding-energy side and the low-binding-energy side in the analysis range. P and Q are respectively the areas of peak intensity from the high-binding-energy side to x and the area of the peak intensity from the low-binding-energy side to x.

4.2. Settings of the Bayesian EMC method and BIC-fitting

In carrying out the Bayesian EMC method, we set the candidate number of peaks as $K = 1, 2, \ldots, 8$. We set the prior distributions of each peak parameter as

$$p(h_k) = \text{Gamma}(h_k; 4.0, 70.0), \quad (23)$$

$$p(\mu_k) = U(\mu_k; -10.0, 40.0), \quad (24)$$

$$p(w_k) = \text{Gamma}(w_k; 2.0, 2.0), \quad (25)$$

$$p(r_k) = U(r_k; 0.0, 1.0), \quad (26)$$

$$p(I_S) = \mathcal{N}(I_S; 80.0, 132.25), \quad (27)$$

$$p(I_E) = \mathcal{N}(I_E; 150.0, 132.25). \quad (28)$$

The values of the inverse temperature $\{\beta_m\}_{m=1}^{M}$ in the EMC method are $M = 24$ and $\gamma = 1.5$ in Eq. (18). As a setting of the Bayesian EMC method, 80,000 MCSs are used for the burn-in and a subsequent 80,000 MCSs for the sampling. The computational conditions in BIC-fitting are the same as those in the literature [5].

4.3. Results and discussion

By the Bayesian EMC method, we can estimate the number of peaks as previously mentioned. By plotting the marginal likelihood $Z(K)$ and free energy $F(K)$ as a function of the number of peaks K, as shown in Figure 10(a), we estimated that $K = 4$ with a probability of 97%. BIC-fitting can also be used to estimate the number of peaks. Figure 10(b) shows the BIC values as a function of the number of peaks K, and the model with the smallest BIC is found at $K = 4$. Considering the properties of the singular model, the results of model selection using the BIC and free energy are not always in agreement [12,13], but in the case of this real spectrum, the same number of peaks are selected with the two methods. Note that the computation by the Bayesian EMC method takes
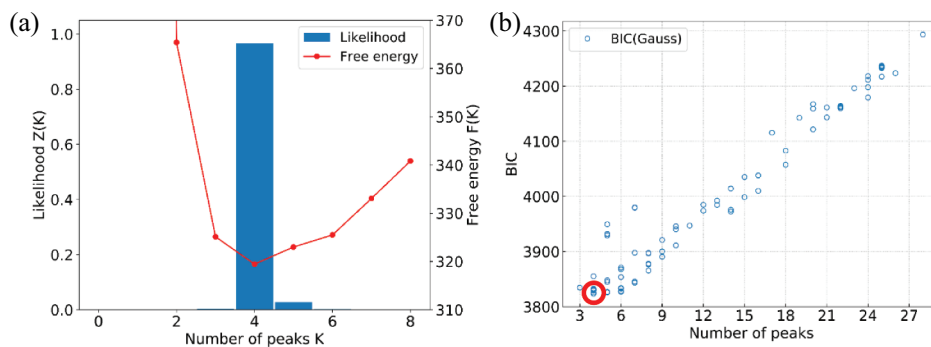

Figure 10. Results of model selection through Bayesian estimation (a) and BIC-fitting (b) for the experimental valence spectrum of SiO$_2$. The red circle in (b) indicates the model with the minimum BIC.
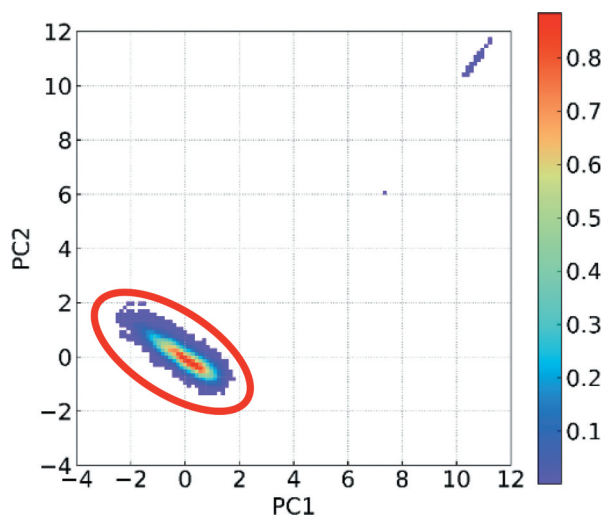
Figure 11. Two-dimensional histogram of PC1 and PC2 obtained by PCA of EMC sampling.

10.5 h, whereas BIC-fitting is completed in less than 3 min.

Figure 9(a) shows the fitted spectrum of the optimum solution obtained from the sampling results for $K = 4$ by the Bayesian EMC method. The spectrum obtained by BIC-fitting is shown in Figure 9(b). We find that both methods reproduce the original spectrum well, and that the positions and shapes of individual peaks are similar. It is note that BIC-fitting derived models equivalent to the optimal solutions obtained using the Bayesian EMC method, despite the limited search space for the solutions.

Using the sampling results of the Bayesian EMC method, we obtain the confidence interval from a posterior probability of each parameter. Principal component analysis (PCA) is performed to find the trend of sampled model groups. Figure 11 shows a two-dimensional heat map of the first and second principal components obtained by PCA with respect to the peak positions of the four peaks. Most models belong to the group enclosed by a red circle in the figure, and optimum solutions are included in this group. Models with features different from those of the optimal solution also appear with a posterior probability as high as 0.2%. We decide to exclude such minority models obtained using the PCA results and then evaluate the posterior probability.

We here focus on two peaks near the binding energies $E_B = 10$ and 16 eV. These two peaks are similar in height and overlap each other. We set the IDs of the two peaks to $k = 1$ and 2. Specifically, Figure 12 shows the posterior probability densities of their peak parameters $\{\mu_k, h_k, w_k, r_k\}(k = 1, 2)$. Looking at Figure 12, the shapes of the distributions are almost Gaussian for the peak position, height, and width, indicating that the sampling was performed appropriately. The distribution of the Lorentz–Gauss mixing ratio is widely scattered within the defined region, which suggests that the ratio is difficult to estimate. The mixing ratio is related to the shape of the tail of the pseudo-Voigt function. Our results suggest that the shape of the tail of the peak is difficult to estimate, because the noise of the target spectrum is relatively high. The results of evaluating the STDs of the
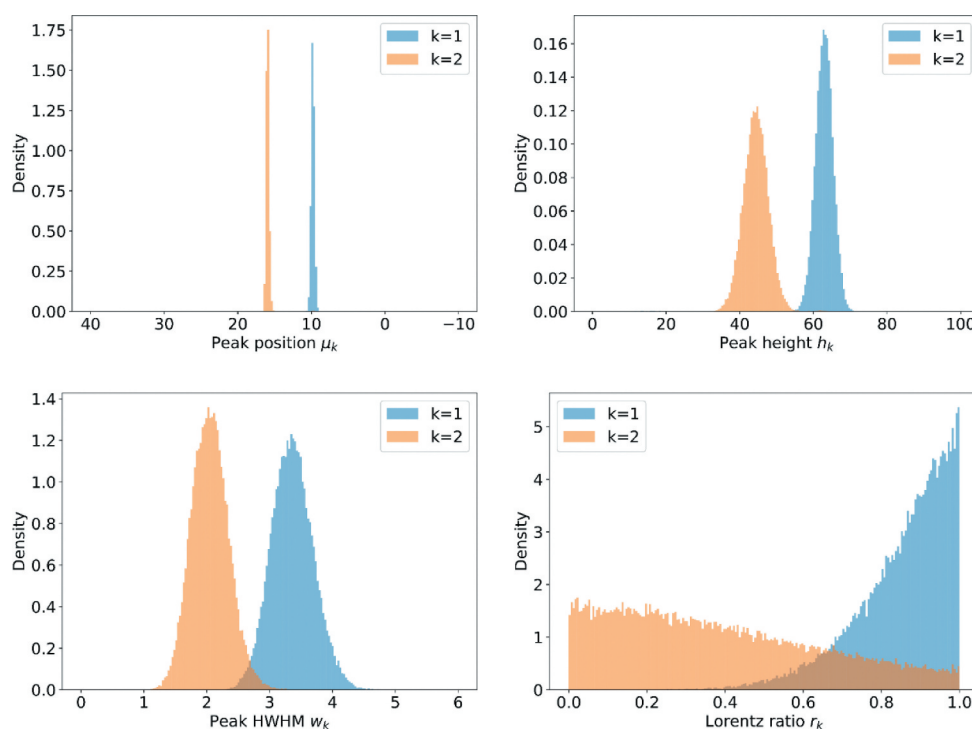

Figure 12. Posterior probability densities of peak parameters $\{\mu_k, h_k, w_k, r_k\}(k = 1, 2)$ for two peaks located at about $E_B = 10$ and 16 eV for the valence spectrum of SiO_2.

Table 3. Calculated confidence intervals of the posterior distribution for all peak parameters and those estimated using Eq. (19).

Parameter	Confidence interval for peak $k = 1$	Confidence interval for peak $k = 2$	Confidence interval estimated using Eq. (19) and the BIC-fitting model
μ	0.22	0.21	0.34
h	2.35	3.69	5.48
w	0.33	0.32	0.40
r	0.12	0.27	—

posterior probability distributions of these peaks are given in Table 3.

The confidence interval of the parameters of the two peaks is obtained, using the peak parameters obtained by the approximated formula (19) and BIC-fitting. It is only necessary to know the estimated peak parameters without using the Bayesian EMC method. The approximated formula assumes that the heights and widths of the two peaks are identical, but this is not so for real spectra. In applying the approximated formula to the real spectrum, we decide to use the average of the two peak heights for h_0 and the average of peak widths for w_0. The STD of noise σ is estimated as the root mean square of the residual between the original spectrum and the fitted spectrum. We further use the peak-to-peak distance Δ obtained from the fitted spectrum.

From the peak parameters of the two peaks obtained by BIC-fitting, we have $h_0 = 53.8, w_0 = 2.7, \Delta = 6.17, \sigma = 10.3$. The estimated S/N ratio is therefore 5.2.

The estimated confidence interval is given in the right-hand column of Table 3. The approximated formula reproduces well the actual confidence interval obtained from the posterior distribution. The S/N ratio is less than 10, and the Lorentz–Gauss mixing ratio r is thus outside the applicable range of the approximated formula (19), and the confidence interval of r cannot be calculated. The results confirm that we can estimate a confidence interval comparable to that estimated by the Bayesian EMC method using the gradient method and the approximated formula (19) when the fitting is good. The approximated formula (19) is also applicable to the case where the heights of the two peaks are more different. Details are shown in Appendix E.

Approximated formula (19) was derived assuming that the spectrum consists of two pseudo-Voigt functions. These simplified two functions mean the nearest neighbor two peaks of the spectrum consists of more than two peaks. Whereas, we must be careful when a target peak is sandwiched between two peaks that are almost equally spaced because the tails of the target peak are overlapping with those of the other two peaks. The Lorentz–Gauss mixing ratio contributes to the shape of the tail of the peak. Therefore, it is difficult to estimate the Lorentz–Gauss mixing ratio of an inner target peak

whose both tails are not clearly distinguished. By using the Bayesian EMC method, in principle, we may be able to obtain the confidence intervals of arbitrary parameters even when a target peak is sandwiched between two peaks and we have to consider what parameter should be used to make a model formula of STD. This will be a future work.

5. Conclusions

We developed a BIC-fitting method with confidence-interval estimation in spectral decomposition. By adopting the Bayesian EMC method, we may be able to not only estimate the number of peaks but also optimize peak parameters, such as the Lorentz–Gauss mixing ratio, in addition to the peak intensity, peak position, and peak width. Using this method, we may also be able to obtain the confidence interval through the STD of the Bayesian posterior distribution. We set various peak-to-peak distances and S/N ratios to generate data on a computer, and then applied Bayesian estimation to obtain the behavior of model selection and the STD of Bayesian posterior distributions for each peak parameter through computer simulation. As a result, an approximated formula expressing the relationship between the obtained STD and the peak-to-peak distance or S/N ratio was derived. In terms of practical use, we confirmed the usefulness of the approximated formula for a real valence spectrum of SiO_2. The confidence interval of each parameter was estimated using the peak parameter obtained by BIC-fitting, and it was confirmed that the value agreed well with the confidence interval obtained directly from the posterior probability obtained using the Bayesian EMC method. In short, even with low-cost optimization methods such as BIC-fitting, we can now estimate confidence intervals of fitting parameters that are comparable to those estimated by high-cost Bayesian EMC methods. Using the approximated formula derived in this study, we may be able to estimate the S/N ratio required to obtain the desired parameters with the desired confidence interval, which will be useful in experimental planning such as the setting of measurement time and energy resolution for individual measurements.

In this study, we treated the peak shapes of the XPS spectra as pseudo-Voigt functions. In practice, the suitable basis function is a Voigt function defined by the convolution of a Lorentzian function derived from the natural widths and a Gaussian function derived from a device. We will consider peak fitting based on convoluted Voigt functions in our future work.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Hiroshi Shinotsuka http://orcid.org/0000-0001-5147-1396
Kenji Nagata http://orcid.org/0000-0001-9894-4461
Hideki Yoshikawa http://orcid.org/0000-0002-7389-8865
Yoh-Ichi Mototake http://orcid.org/0000-0002-6044-1620
Hayaru Shouno http://orcid.org/0000-0002-2412-0184
Masato Okada http://orcid.org/0000-0002-9040-8784

References

[1] Kobayashi K, Yabashi M. High resolution-high energy x-ray photoelectron spectroscopy using third-generation synchrotron radiation source, and its application to Si-high k insulator systems. Appl Phys Lett. 2003;83:1005–1007.

[2] Takata Y, Yabashi M, Tamasaku K, et al. Development of hard X-ray photoelectron spectroscopy at BL29XU in SPring-8. Nucl Instrum Methods Phys Res, Sec A. 2005;547:50–55.

[3] Hook AL, Anderson DG, Langer R, et al. High through-put methods applied in biomaterial development and discovery. Biomaterials. 2010;31:187–198.

[4] Matsumura T, Nagamura N, Akaho S, et al. Spectrum adapted expectation-maximization algorithm for high-throughput peak shift analysis. Sci Technol Adv Mater. 2019;20:733–745.

[5] Shinotsuka H, Yoshikawa H, Murakami R, et al. Automated information compression of XPS spectrum using information criteria. J Electron Spectrosc Relat Phenom. 2020;239:146903.

[6] Nagata K, Sugita S, Okada M. Bayesian spectral deconvolution with the exchange Monte Carlo method. Neural Netw. 2012;28:82–89.

[7] Hesse R, Streubel P, Szargan R. 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. Surf Interface Anal. 2007;39:381–391.

[8] Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc Jpn. 1996;65:1604–1608.

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

[10] Marquardt DW. An algorithm for least squares estimation of nonlinear parameters. J Soc Ind Appl Math. 1963;11:431–441.

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

[12] Watanabe S. Algebraic analysis for nonidentifiable learning machines. Neural Comput. 2001;13:899–933.

[13] Watanabe S. Algebraic geometrical methods for hierarchical learning machines. Neural Netw. 2001; 14:1049–1060.

[14] Nagata K, Watanabe S. Asymptotic behavior of exchange ratio in exchange Monte Carlo method. Neural Netw. 2008;21:980–988.

[15] Yoshihara K. The introduction of common data processing system version 12. J Surf Anal. 2017;23: 138–148.

[16] Ogata Y. A Monte Carlo method for an objective Bayesian procedure. Ann Inst Stat Math. 1990; 42:403–433.

[17] Branch MA, Coleman TF, Li Y. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. SIAM J Sci Comput. 1999;21:1–23.

Appendix A. Bayesian spectral deconvolution

In this section, we describe Bayesian spectral decomposition used to fit the spectral data and estimate the number of peaks.

A.1 Bayesian estimation

Bayesian estimation is a framework in which the process of generating data in a probabilistic model is formulated and an estimation is made by tracing back the causal relationship using the Bayesian theorem. We first define a probabilistic model. It is assumed that the spectral data y are generated by adding noise ε to the spectrum function $f(x; \theta)$:

$$y = f(x; \theta) + \varepsilon. \tag{A-1}$$

We assume that ε follows a Gaussian distribution with zero mean and variance σ^2, and the conditional probability $p(y|x, \theta)$ of the spectral data y is then formulated as

$$p(y|x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \{y_i - f(x_i; \theta)\}^2\right). \tag{A-2}$$

Assuming that data $\{y_i\}_{i=1}^n$ are independent and identically distributed, using the error function $E(\theta)$, we obtain the probability of the data set $D = \{(x_i, y_i)\}_{i=1}^n$ as

$$p(D|\theta) = \prod_{i=1}^n p(y_i|x_i, \theta) \propto \exp\left(-\frac{n}{\sigma^2} E(\theta)\right). \tag{A-3}$$

For the Gaussian probability model, the maximum likelihood estimation method that maximizes $p(D|\theta)$ is equivalent to the well-known least-squares method.

In the Bayesian estimation, we regard not only the data D but also the parameter set θ and the number of peaks K as random variables; i.e., we construct a model of the simultaneous probability distribution $p(D, \theta, K)$:

$$\begin{aligned} p(D, \theta, K) &= p(D|\theta, K)p(\theta|K)p(K) \\ &\propto \exp\left(-\frac{n}{\sigma^2} E(\theta)\right)p(\theta|K)p(K). \end{aligned} \tag{A-4}$$

Here, $p(K)$ and $p(\theta|K)$ are prior distributions, and they should be set in advance.

We estimate the parameter θ using the posterior distribution $p(\theta|D, K)$:

$$\begin{aligned} p(\theta|D, K) &= \frac{p(D, \theta, K)}{p(D, K)} = \frac{p(D, \theta, K)}{\int p(D, \theta, K)d\theta} \\ &= \frac{\exp\left(-\frac{n}{\sigma^2}E(\theta)\right)p(\theta|K)}{\int \exp\left(-\frac{n}{\sigma^2}E(\theta)\right)p(\theta|K)d\theta} \\ &= \frac{1}{Z(K)} \exp\left(-\frac{n}{\sigma^2} E(\theta)\right)p(\theta|K), \end{aligned} \tag{A-5}$$

where

$$Z(K) = \int \exp\left(-\frac{n}{\sigma^2} E(\theta)\right)p(\theta|K)d\theta. \tag{A-6}$$

This technique for obtaining the parameter θ that maximizes the posterior distribution $p(\theta|D, K)$ is a maximum a posteriori (MAP) estimation, and the parameter θ at this time is called the MAP estimator. Bayesian estimation has the advantage that both the optimum value of the parameter and the confidence interval can be evaluated by obtaining the width of the posterior distribution $p(\theta|D, K)$.

We use the posterior probability $p(K|D)$ to estimate the number of peaks K. This can be done by adopting the procedure of probability marginalization:

$$\begin{aligned} p(K|D) &= \frac{p(D, K)}{p(D)} = \frac{\int p(D, \theta, K)d\theta}{\sum_k \int p(D, \theta, K)d\theta} \\ &= \frac{\int \exp\left(-\frac{n}{\sigma^2}E(\theta)\right)p(\theta|K)p(K)d\theta}{\sum_k \int \exp\left(-\frac{n}{\sigma^2}E(\theta)\right)p(\theta|k)p(k)d\theta} = \frac{p(K)Z(K)}{\sum_k p(k)Z(k)}, \end{aligned} \tag{A-7}$$

where $\sum_k$ represents the sum of the numbers of peaks k. The denominator in Eq. (A-7) does not depend on K because it is the sum of all peaks k. We thus obtain

$$p(K|D) \propto p(K)Z(K). \tag{A-8}$$

Therefore, to obtain the posterior probability $p(K|D)$ with respect to the number of peaks, we need to calculate $Z(K)$. The Bayesian free energy is defined as

$$\begin{aligned} F(K) &= -\log Z(K) \\ &= -\log \int \exp\left(-\frac{n}{\sigma^2} E(\theta)\right)p(\theta|K)d\theta. \end{aligned} \tag{A-9}$$

If $p(K)$ is uniformly distributed, the maximization of the posterior probability $p(K|D)$ is equivalent to the minimization of the Bayesian free energy $F(K)$. The number of peaks K can therefore be estimated by minimizing the Bayesian free energy.

A.2 Exchange Monte Carlo method

How to evaluate a posterior probability $p(\theta|D, K)$ is a problem when calculating the spectral decomposition through Bayesian estimation. In the case of spectral decomposition, the posterior distribution for the parameter θ is difficult to handle because the distribution does not become a known distribution, such as a Gaussian distribution. As an example, it is difficult to determine the MAP estimators because of the presence of local solutions. Furthermore, when we attempt to determine the confidence interval of the parameter θ, it is necessary to determine the shape of the distribution itself, which is much more difficult.

The exchange Monte Carlo (EMC) method [8] solves the above problems. In this study, the method described below is called the Bayesian EMC method, which is a Markov-chain Monte Carlo method. We prepare a plurality of distributions,

$$p_\beta(\theta|D) = \frac{1}{Z_\beta(K)} \exp\left(-\frac{n\beta}{\sigma^2} E(\theta)\right) p(\theta|K), \qquad (A-10)$$

$$Z_\beta(K) = \exp\left(-\frac{n\beta}{\sigma^2} E(\theta)\right) p(\theta|K) d\theta, \qquad (A-11)$$

in which the inverse temperature β is introduced into the posterior distribution, and simultaneously perform sampling in parallel. That is, the target distribution becomes the simultaneous distribution

$$p(\theta_1, \ldots, \theta_M) = \prod_{m=1}^M p_{\beta_m}(\theta_m|D). \qquad (A-12)$$

Here, the inverse temperature $\{\beta_m\}_{m=1}^M$ is $0 = \beta_1 < \beta_2 < \cdots < \beta_M = 1$ and M is the number of prepared temperatures. θ_m is the fitting parameters for inverse temperature β_m. For $\beta_M = 1$, $p_{\beta_M}(\theta_M|D)$ matches the posterior distribution $p(\theta|D, K)$.

The algorithm of the EMC method comprises two updates.

Metropolis sampling for each temperature. For each inverse temperature β_m, we sample θ_m from the distribution of $p_{\beta_m}(\theta|D)$ using the Metropolis algorithm.

State exchange between adjacent temperatures. We exchange states between adjacent temperatures and set $\{\theta_m, \theta_{m+1}\} \rightarrow \{\theta_{m+1}, \theta_m\}$. We determine whether to exchange on the basis of the probability

$$p(\theta_m \leftrightarrow \theta_{m+1}) = \min(1, \nu), \qquad (A-13)$$

$$\begin{aligned} \nu &= \frac{p_{\beta_m}(\theta_{m+1}|D) p_{\beta_{m+1}}(\theta_m|D)}{p_{\beta_m}(\theta_m|D) p_{\beta_{m+1}}(\theta_{m+1}|D)} \\ &= \exp\{(\beta_{m+1} - \beta_m)(E(\theta_{m+1}) - E(\theta_m))\} \end{aligned} \qquad (A-14)$$

As a result, the parameter θ_M obtained from the M-th inverse temperature β_M can be regarded as sampling from the posterior distribution $p(\theta|D, K)$. Thus, by repeating the above procedure sufficiently many times and recording θ_M, we obtain a sample sequence from the posterior distribution $p(\theta|D, K)$ and we can determine the MAP estimators and confidence interval.

There are two advantages of the EMC method. The first is that heating and annealing effects can be introduced by stochastically exchanging samples between adjacent temperatures during the sampling of each distribution. We thereby escape local solutions and efficiently search for global solutions. This is an important advantage for the accurate estimation of MAP estimators and confidence intervals. The second is that the free energy $F(K)$ can also be calculated using the results of the EMC method. We introduce the inverse temperature β into the free energy:

$$f(\beta) = -\log \int \exp\left(-\frac{n\beta}{\sigma^2} E(\theta)\right) p(\theta|K) d\theta. \qquad (A-15)$$

At the high-temperature limit $\beta = 0$, it holds that $f(\beta) = 0$. The desired free energy $F(K) = f(\beta = 1)$ is expressed as

$$F(K) = f(\beta = 1) = \int_0^1 d\beta \frac{\partial f}{\partial \beta}, \qquad (A-16)$$

$$\begin{aligned} \frac{\partial f}{\partial \beta} &= \frac{\frac{n}{\sigma^2} E(\theta) \exp\left(-\frac{n\beta}{\sigma^2} E(\theta)\right) p(\theta|K) d\theta}{\exp\left(-\frac{n\beta}{\sigma^2} E(\theta)\right) p(\theta|K) d\theta} \\ &= \frac{n}{\sigma^2} E(\theta) p_\beta(\theta|D) d\theta \equiv \frac{n}{\sigma^2} E(\theta)_{p_\beta(\theta|D)}. \end{aligned} \qquad (A-17)$$

$F(K)$ is therefore an integral of the expected value $\frac{n}{\sigma^2} E(\theta)_{p_\beta(\theta|D)}$ under the probability distribution $p_\beta(\theta|D)$ with respect to the inverse temperature β in the range from 0 to 1. In addition, because the results of the EMC method provide samples at each inverse temperature β_m, the expected value $\frac{n}{\sigma^2} E(\theta)_{p_{\beta_m}(\theta|D)}$ can be calculated. Therefore, by using the piecewise quadrature method, we can calculate the free energy and perform model selection [6,16].

Appendix B. Effects of changes in the data set

We here discuss how the data set affects the posterior distribution of parameter θ shown in Figure 4. Specifically, we generate five data sets in which only the noise (initial random seed) is changed under the same condition as that in Figure 1(a), and we then perform Bayesian estimation for each data set. The results of model selection and the posterior distributions of the parameters w_1, w_2, r_1, and r_2 are shown in Fig. A-1. The results for the zero seed are the same as those in Figure 2(a) and 4. It is seen that the results of model selection are stable regardless of the data set. In Fig. A-1, deviation from the true values for w_1 and r_1 in Figure 4 is strongly affected by the noise added to the data. To reduce this deviation, spectral data must be measured so that the S/N ratio improves. It also seems that the
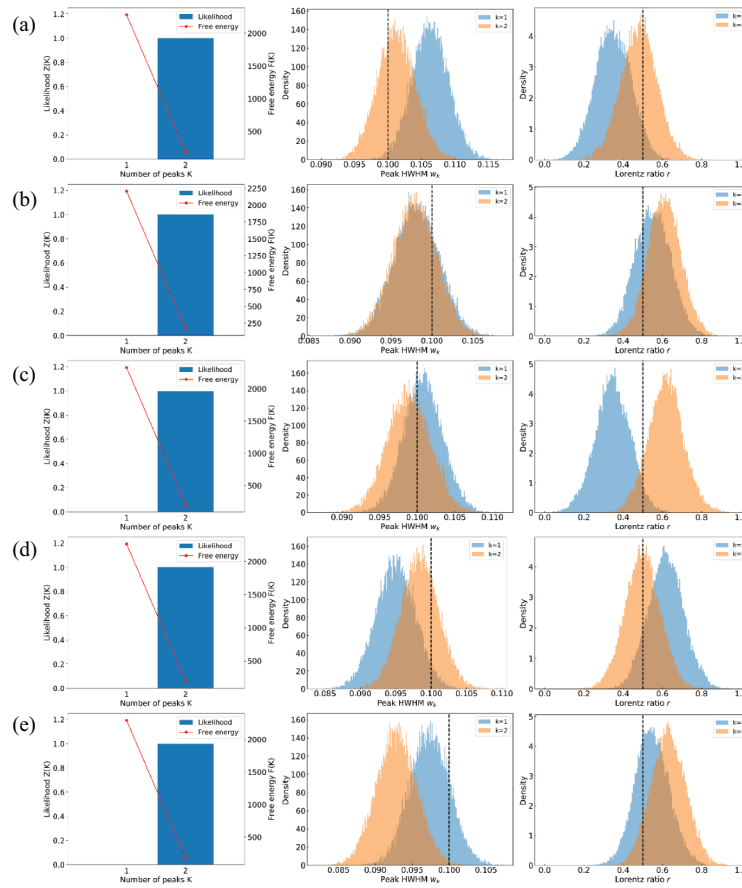
Figure A-1. Results of model selection and posterior distributions $p(w_k|D, K)$ and $p(r_k|D, K)$. (a)–(e) only differ in the initial random seeds used in generating the spectral data sets under the same conditions as those in Figure 1(a).

deviations from the true values are correlated between parameters. For example, there is a negative correlation between w_1 and r_1; when w_1 shifts to a larger value, the corresponding r_1 shifts to a smaller value. We have discussed this result in detail in Appendix C.

Appendix C. Correlation structure in the posterior distribution of the parameters $p(\theta|D)$

The two-dimensional distribution of parameters θ is shown in Fig. A-2. The diagonal components of the figure show the histogram of the posterior distribution for each parameter. The lower off-diagonal components show two-dimensional distributions, whereas the upper off-diagonal components represent correlation coefficients defined by

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{s_X s_Y}, \qquad (A-18)$$

where s_X and s_Y are respectively the STDs of X and Y data, whereas $\text{cov}(X, Y)$ is the covariance between X and Y data. μ is almost uncorrelated with other parameters, whereas the parameters h, w, and r are strongly correlated with each other. This correlation is considered to be an intrinsic property of the model that comes from the constraint that the change in the spectral shape is small. Considering the various correlations, it may be possible to estimate a property derived from a spectral parameter, such as a peak area intensity.

Appendix D. Features of the regression function and the optimization method of the parameters

The functional form of Eq. (19) with fitting parameters $B, C, D,$ and E is

$$s\left(\Delta, \frac{\sigma}{h_0}\right) = \frac{\sigma}{h_0} B\left\{ \left(\frac{\Delta - E}{D}\right)^C + 1 \right\} \qquad (A-19)$$

σ/h_0 is the reciprocal of the S/N ratio. B is associated with the asymptote of the value of the STD at which the scaled peak-to-peak distance Δ is sufficiently large. C is a negative value related to the curvature of the function. The larger the $|C|$ value, the more rapidly the STD increases as Δ decreases. D is a parameter for adjusting the scale in the Δ-axis direction. E is the position of the asymptote on the Δ axis, and the normal deviation $\tilde{s}$ diverges to positive infinity as Δ approaches E; i.e., the domain of this function is $\Delta > E$.
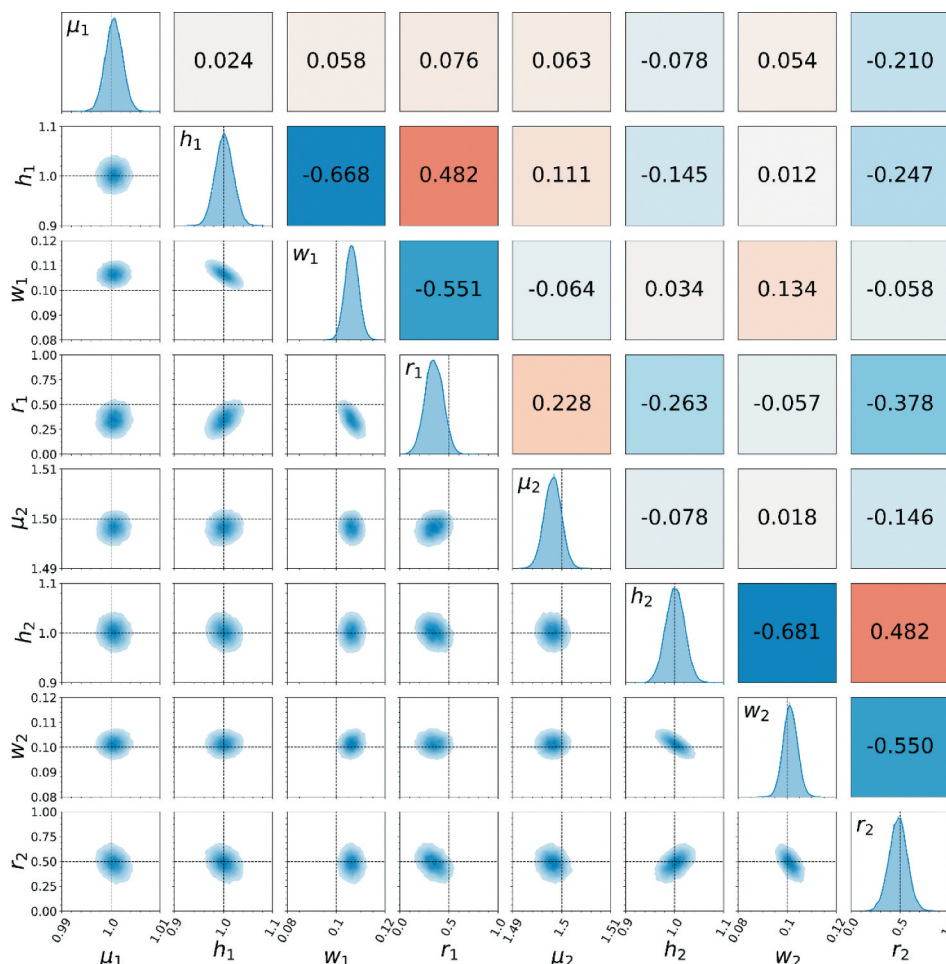
Figure A-2. Two-dimensional distribution of parameters θ. The diagonal components of the figure show the histogram of the posterior distribution of each parameter. The lower off-diagonal components are two-dimensional distributions, whereas the upper off-diagonal components show the coefficients of correlation between two parameters. The dotted lines indicate the true parameter values used to generate the spectral data.

The fitting parameters are optimized using the trust region reflective algorithm [17]. The parameters are optimized so as to minimize the square error on a logarithmic scale with a constraint condition. We fix $D = 2.5$, as described in the text, and impose the constraints $0 < B < \infty$, $-\infty < C < 0$ and $-5 < E < 0$. We also assume that the range of peak-to-peak distances to be optimized is $\Delta \geq 0.4$. The ranges of the S/N ratio to be optimized are $S/N \geq 10.0$ for r, $S/N \geq 1.0$ for μ, $S/N \geq 2.0$ for h, and $S/N \geq 5.0$ for w.

We perform the optimization using the curve_fit function in the SciPy library of Python3.6.

Appendix E. Application of approximated formula (19) to a combination of a large peak and a weak peak

In a real spectrum, it is common for a satellite peak to overlap a main peak. Therefore, assuming an application to a real spectrum, we simulated the case where the heights of the two peaks are different. First, artificial spectra were generated under two conditions, and the posterior distribution obtained by Bayesian EMC was investigated in detail. Next, we systematically changed the peak height and investigated whether the approximation formula (19) for the confidence interval of the parameter can be applied.

First, we show two calculation examples. Two artificial spectra were generated under two conditions as shown in Figs. E-1 (a) and (b). The conditions for the artificial spectra are as follows. The peak heights are $\left(h_1^*, h_2^*\right) = (1.0, 0.5)$ for Fig. E-1(a) and $\left(h_1^*, h_2^*\right) = (1.0, 0.1)$ for Fig. E-1(b). Other peak parameters are common, the peak positions are $\left(\mu_1^*, \mu_2^*\right) = (1.0, 1.2)$, the HWHMs are $w_1^* = w_2^* = 0.1$, and the Lorentz–Gauss mixing ratios are $r_1^* = r_2^* = 0.5$. The background is not taken into consideration. The STD of Gaussian noise is $\sigma = 0.02$. We applied the Bayesian EMC method to each spectrum. As a result, a model with two peaks was selected for both spectra. The fitted results of Bayesian estimation for each spectrum are shown in Figs. E-1 (a) and (b).

Figures E-2 and E-3 show the posterior distributions of the fitting parameters corresponding to the spectra in Figs. E-1 (a) and E-1 (b), respectively. Figure E-2 shows that each parameter is estimated with good accuracy in that the distribution roughly includes the true value and appears similarly to a Gaussian distribution. The confidence interval of the parameters of the second peak is broader than that of the first peak, except for the confidence interval of peak height. The small second peak has a relatively
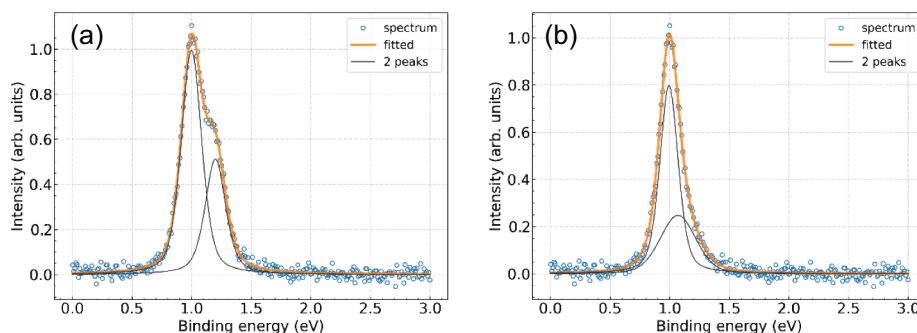
Figure E-1. Two examples of artificial spectra that have two peaks with different heights: (a) $h_1^* = 1.0$, $h_2^* = 0.5$, (b) $h_1^* = 1.0$, $h_2^* = 0.1$. The other parameters are the same. Open circles are the artificial spectrum, the orange line is the fitted spectrum, and the black lines are the peak components.
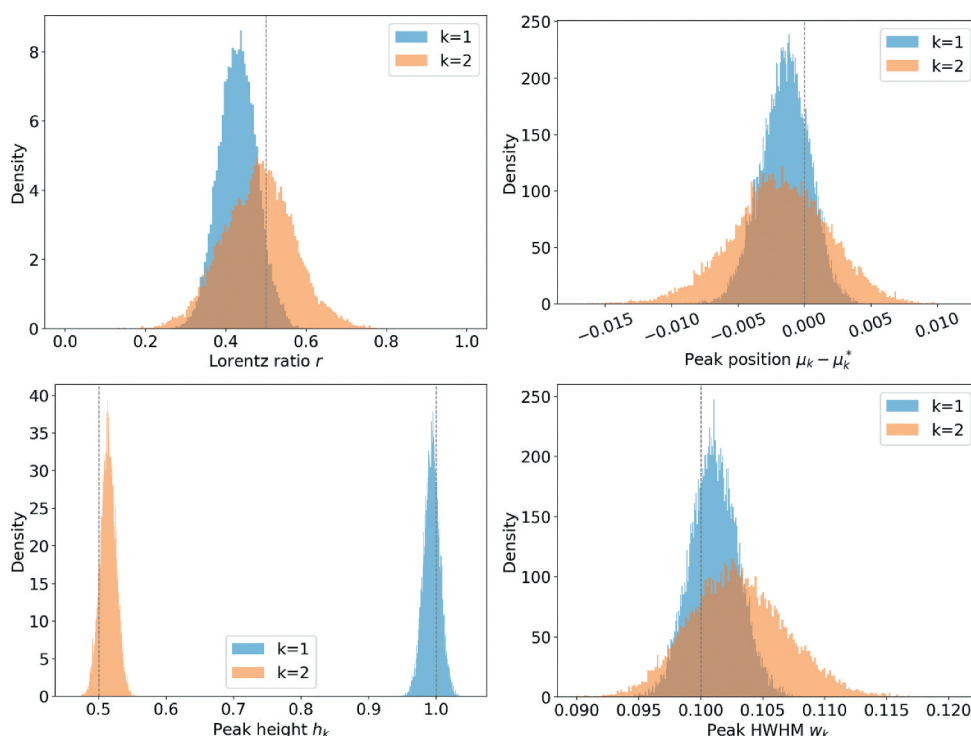

Figure E-2. Posterior distribution of each parameter when Bayesian estimation is performed on the spectral data in Fig. E-1(a). The dashed lines indicate the true parameter values used to generate the spectral data.

low S/N ratio because the Gaussian noise is uniformly applied to the entire spectral range. As a result, the parameters of the second peak became more difficult to estimate, and their confidence intervals became wide. On the contrary, the posterior distributions of two peak heights have similar confidence intervals because the Gaussian noise is uniformly applied. Figure E-3 shows that the posterior distributions of the second peak parameters are unclear. It is, thus, difficult to estimate the parameters of a small peak. This is probably because the height of the second peak is small and it is very strongly affected by noise.

Next, to investigate the applicability of approximated formula (19), we prepared artificially measured spectra in which the height of the second peak was systematically changed. The conditions of the spectra were almost the same as those in Fig. E-1 (a), but the heights of the second peak were changed as $h_2^* = 0.1, 0.2, 0.3, \ldots, 1.0$. We applied the Bayesian EMC method to each spectrum and calculated the confidence intervals for each fitting parameter after filtered by PCA as in Section 4. Figure E-4 shows the confidence intervals as a function of the height of the second peak h_2^*. We also plotted the values of confidence intervals calculated using approximated formula (19) using the true peak parameters. From this figure, it is confirmed that even if the heights of the two peaks are different, the approximated formula (19) can reproduce the actual confidence intervals obtained by the Bayesian EMC method. However, when $h_2^* = 0.1$, the confidence interval of the parameter of the second peak does not agree well with the approximated formula (19). Note that it is difficult to apply the approximated formula (19) when the parameters are difficult to estimate owing to the strong effect of noise.
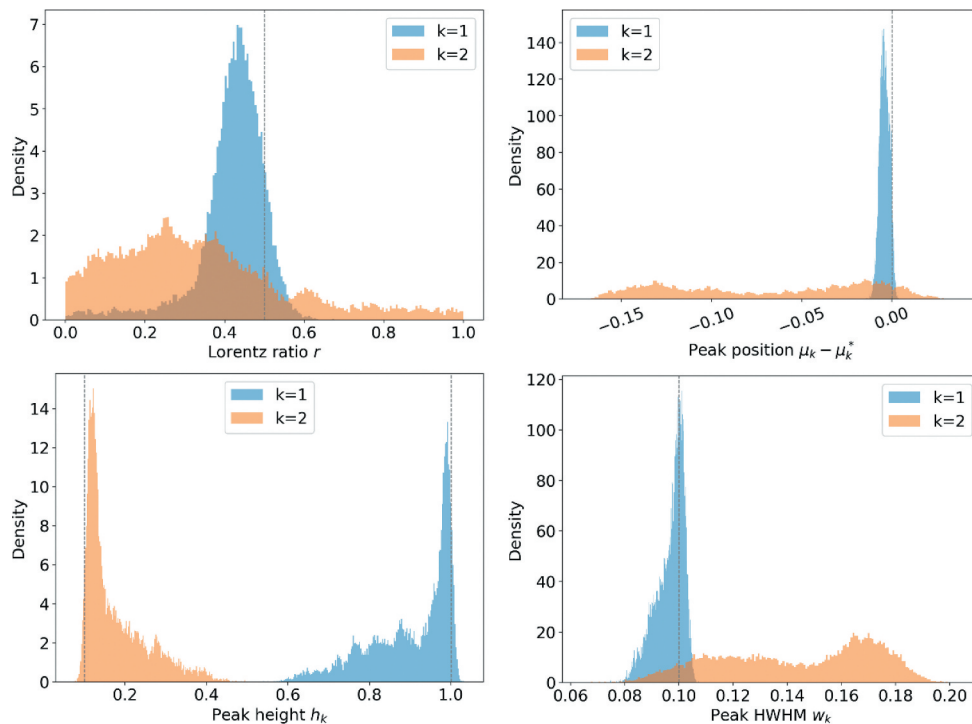
Figure E-3. Posterior distribution of each parameter when Bayesian estimation is performed on the spectral data in Fig. E-1(b). The dashed lines indicate the true parameter values used to generate the spectral data.
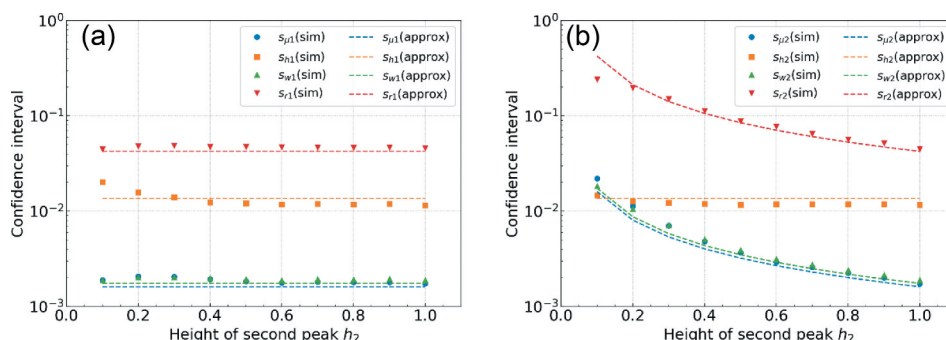

Figure E-4. Confidence intervals of peak parameters as a function of the height of the second peak h_2 for peaks 1 (a) and 2 (b). (sim) indicates values simulated by the Bayesian EMC method, (approx) indicates values calculated with the approximated formula (19).