Bayesian estimation for XPS spectral analysis at multiple core levels

Atsushi Machida\textsuperscript{a}, Kenji Nagata\textsuperscript{b}, Ryo Murakami\textsuperscript{c}, Hiroshi Shinotsuka\textsuperscript{b}, Hayaru Shouno\textsuperscript{c}, Hideki Yoshikawa\textsuperscript{b} and Masato Okada \textsuperscript{a}\textsuperscript{b}\textsuperscript{d}

\textsuperscript{a}Graduate School of Science, The University of Tokyo, Bunkyo, Japan; \textsuperscript{b}Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba, Japan; \textsuperscript{c}Department of Informatics, The University of Electro-Communications, Chofu, Japan; \textsuperscript{d}Graduate School of Frontier Science, The University of Tokyo, Kashiwa, Japan

**ABSTRACT**
X-ray photoelectron spectroscopy (XPS) is a widely used measurement technique in material surface analysis, but its analysis is subject to operator arbitrariness in the results. In a previous paper, a method based on genetic algorithms was proposed to estimate the composition ratios of compounds from XPS data using reference spectra and it was shown that it is possible to analyze them automatically from the reference spectra data. In this paper, we newly proposed a Bayesian spectral decomposition method based on the exchange Monte Carlo method and tested it on artificial data. This method provides a posterior distribution of the model parameters. This not only allows the estimation of compositional ratios for samples, but also allows statistical reliability assessment. In addition, we simulated an artificial data analysis to clarify the effect on the identification of compounds and the estimation of their compositional ratios by varying the signal-to-noise ratio of the data.

1. Introduction
X-ray photoelectron spectroscopy (XPS) is a technique to measure the kinetic energy distribution of photoelectrons emitted by X-ray irradiation and to obtain information on the type, amount and chemical bonding state of elements present on the surface of a sample, which is used in various surface analyses of materials. In the field of spectroscopy, reference spectrum databases are being rapidly constructed, and it is necessary to develop automated spectrum analysis methods using reference spectrum data. Previous studies developed a method to estimate parameters such as composition ratios, peak position adjustments and peak width adjustments of samples using genetic algorithms [1]. Additionally, this method makes it possible to identify the combinations of compounds in samples using Bayesian information criterion. The genetic algorithms, however, do not allow us to make statistical reliability assessments about the results of the estimation.

This paper proposes a Bayesian inference using the exchange Monte Carlo (MC) method for identifying kinds of compounds and estimating their composition ratios from XPS data of samples. In the Bayesian framework, it is possible to obtain the probability distribution of model parameters given the data, which allows us to estimate the reliability of the results. Nagata et al. estimated the position, width and intensity of the spectrum of each compound in a sample using the exchange MC method with a Gaussian distribution as the basis function [2]. On the other hand, considering the distribution represented by a pseudo-Voigt function to fit XPS data...
[3], we propose a framework to estimate the model parameters including the composition ratio by the exchange MC method using the reference spectrum of each compound given in advance. The compound identification is statistically equivalent to the framework for estimating the number of peaks and can be solved by the model selection framework [2]. However, it is not realistic from a computational viewpoint to evaluate the Bayesian free energy for all combinations of compounds. Therefore, we introduce an indicator parameter that expresses the presence or absence of each compound and propose a framework for compound identification by considering the indicator parameter as a model parameter and obtaining its posterior distribution.

To verify the effectiveness of the proposed method, we conducted Bayesian estimation using artificial data. In particular, taking advantage of the fact that the measurement time in XPS is strongly related to the signal-to-noise ratio of the data, we generated data with various signal-to-noise ratios by introducing the measurement time in the data generation. For these data, the statistical reliabilities of compound identification and composition ratio estimation are evaluated in the framework of Bayesian inference, and the relationship between measurement time and estimation accuracy is clarified.

2. Models

2.1. Composition ratio estimation

In this section, we will describe the model for estimating the composition ratio and other parameters of observed samples from XPS data. In XPS analysis, the orbitals have a one-to-one relationship with the energy range. The spectrum $f_i(x)$ of the $i$th orbital is expressed as

$$f_i(x) = f(x; \Theta_i)$$

$$= S(x; \Theta_S) + B(x; \Theta_S; \Theta_B),$$

where $\Theta_i = \{\Theta_S, \Theta_B\}$ is the set of parameters on orbital $i$, which is composed of the signal parameter set $\Theta_S$ and the background parameter set $\Theta_B$. The signal spectrum function $S(x; \Theta_S)$ and the background spectrum function $B(x; \Theta_S; \Theta_B)$ are provided as

$$S(x; \Theta_S) = h_i \sum_{k=1}^{K} c_k R_{ik}(x; \mu_k, \tilde{w}_k),$$

$$R_{ik}(x; \mu_k, \tilde{w}_k) = A_{ik} \sum_{j=1}^{M_k} m_{kj} V(x; \mu_k + \mu_{kj}, \tilde{w}_{kj}, r_{kj}),$$

$$V(x; \mu, w, r) = \frac{(1-r)^2 \ln 2}{\sqrt{\pi w}} \exp\left(-\frac{2}{w}(x - \mu)^2\right) + \frac{r}{\pi \left((x - \mu)^2 + w^2\right)},$$

$$B(x; \Theta_S, \Theta_B) = (a_i - b_i) \int_{0}^{\infty} d\nu S(u; \Theta_S) + b_i, \quad \int_{0}^{\infty} d\nu S(u; \Theta_S)$$

where $h_i$ is the intensity parameter and $K$ is the number of compounds. The parameter $c_k$ is the composition ratio of the $k$th compound in the sample and satisfies $\sum_k c_k = 1$. The function $R_{ik}(x; \mu_k, \tilde{w}_k)$ is the reference spectrum of the $k$th compound in the $i$th orbital, which plays the role of a basis function. $\mu_k$ is the adjustment parameter for the peak position and $\tilde{w}_k$ is the adjustment parameter for the peak width. Note that $\mu_k$ does not depend on the $i$th orbital. This is modeled by considering the effect of charge-up, which is the main cause of the misalignment of peak positions in XPS and assumes that the misalignment does not depend on the orbital and is uniform in the whole energy region. The pseudo-Voigt function $V(x; \mu, w, r)$ is used for a single peak [3]. This function is defined as the sum of a Gaussian function and a Lorentz function. The parameter $A_{ik}$ is the probability of the atoms of the $k$th compound in the $i$th orbital and satisfies $\sum_i A_{ik} = 1$. It can be interpreted as the probability of obtaining the reference spectrum of the target atoms from the $k$th compound. The parameter $M_{ik}$ is the number of peaks obtained from the $k$th compound in the $i$th orbital. The parameter $m_{kj}$ is the area ratio of each peak and the parameter $\mu_{kj}$ is its position. The parameter $\tilde{w}_{kj}$ is the peak width and the parameter $r_{kj}$ is the value that determines the ratio of the Gaussian and Lorentz functions. The background $B$ is modeled on the basis of the Shirley method [4–6]. The parameters $a_i$ and $b_i$ determine the edge intensity of the background, where $a_i > b_i$. The set of parameters representing the reference spectrum

$$\left\{\left\{A_{ik}, m_{kj}, \mu_{kj}, \tilde{w}_{kj}, r_{kj}\right\}_{k=1}^{M_k}\right\}_{i=1,k=1}^{L,K} \quad \left(7\right)$$

is assumed to be given beforehand. Therefore, the set of parameters $\Theta_i$ of the $i$th orbital necessary to estimate the spectrum $f_i(x) = f(x; \Theta_i)$ is defined as,

$$\Theta_i = \{\Theta_S, \Theta_B\},$$

$$\Theta_S = \{h_i, \{c_k, \mu_k, \tilde{w}_k\}_{k=1}^{K}\},$$

$$\Theta_B = \{a_i, b_i\}. \quad \left(10\right)$$

Since the XPS observation data are count data, the conditional probability of intensity $y$ given energy $x$
in the \(l\)th orbital and the parameter set \(\Theta_l\) follows a Poisson distribution:

\[
p(y|\Theta_l) = \frac{f(x; \Theta_l)^y \exp(-f(x; \Theta_l))}{y!}.
\]

(11)

We assume that each data point is i.i.d., then the conditional probability given a dataset of observations \(\{X, Y\} = \{\{x_{il}, y_{il}\}_{l=1}^L\}_{l=1}^L\) for \(L\) orbitals is

\[
p(Y|\Theta) = \prod_{l=1}^L \prod_{i=1}^{N_l} p(y_{il}|\Theta_l)
\]

(12)

for the set of parameters \(\Theta = \{\Theta_l\}_{l=1}^L\). Also the cost function \(E(\Theta)\) is

\[
E(\Theta) = -\sum_{l=1}^L \sum_{i=1}^{N_l} \{y_{il} \log(f(x_{il}; \Theta_l))
\]

\[
- f(x_{il}; \Theta_l) - \log(y_{il})\}.
\]

(13)

From Bayes’ theorem, the posterior distribution of the parameter set \(\Theta\) is provided as

\[
p(\Theta|Y) = \frac{p(Y|\Theta)p(\Theta)}{p(Y)} = \frac{\varphi(\Theta) \prod_{l=1}^L \prod_{i=1}^{N_l} p(y_{il}|\Theta_l)}{p(Y)}
\]

(14)

where \(\varphi(\Theta)\) is the prior distribution corresponding to \(p(\Theta). \ p(Y)\) is a normalized constant determined by the dataset and does not depend on \(\Theta\) because Equation (14) is a probability distribution.

\[
p(Y) = \int d\Theta p(Y|\Theta)\varphi(\Theta)
\]

(15)

From the posterior distribution obtained above, the parameters are estimated.

\[
p(\Theta|Y) \propto \varphi(\Theta) \prod_{l=1}^L \prod_{i=1}^{N_l} \{y_{il} \log(f(x_{il}; \Theta_l))
\]

\[
- f(x_{il}; \Theta_l) - \log(y_{il})\}
\]

(16)

\[
= \varphi(\Theta) \exp\left(\sum_{l=1}^L \sum_{i=1}^{N_l} \{y_{il} \log(f(x_{il}; \Theta_l))
\]

\[
- f(x_{il}; \Theta_l) - \log(y_{il})\} \right)
\]

(17)

\[
= \varphi(\Theta) \exp(-E(\Theta))
\]

(18)

By using the posterior probability distribution of Equation (18), it is possible to estimate not only each parameter but also its confidence interval. In Experiment 1, we will clarify the estimation accuracy and confidence intervals of various parameters by calculating the posterior probability distribution for artificial data.

Note that one of the features of this model is that it is not sufficient to perform the estimation for each orbital but it is necessary to consider all the orbitals. If each parameter set \(\Theta_l\) of the \(l\)th orbital is independent of the other, the above cost function \(E(\Theta)\) can be expressed as the sum of the cost functions \(E_l(\Theta_l)\) for the \(l\)th orbital. In this case, the posterior probability distribution should also be discussed for each orbital. However, in the model in this paper, all the orbitals must be treated as a simultaneous probability distribution because \(c_k\) and \(\hat{\mu}_k\) exist in the parameter set \(\Theta_l\) as parameters independent of the \(l\)th orbital.

\[\text{2.2. Compound identification}\]

Depending on the combination of compounds used for estimation, the estimation accuracy of the model in the previous section may deteriorate due to the contribution of compounds that are not included in the sample. To avoid this problem, it is necessary to find the true combination from various combinations of compounds. Therefore, we propose an extended model for model selection using the framework of Bayesian estimation in the previous section. We introduce a new parameter \(g \in \{0, 1\}\) into Equation (3) and consider the following model.

\[
S(x; \theta_n) = h_j \sum_{k=1}^K g_k c_k R_k(x; \hat{\mu}_k, \hat{\omega}_k)
\]

(19)

\(g_k = 1\) denotes that the compound \(k\) is included in the spectrum data and \(g_k = 0\) denotes that it is not included in the data. It is possible not only to estimate the composition ratio of each compound but also to identify whether it is present in the sample by simultaneously estimating the indicator parameters \(g = (g_1, \ldots, g_K)\). When identifying combinations of compounds, it is difficult to predetermine the required set of composition ratios in advance. Therefore, in this model, instead of \(c_k\), we introduce \(c_k\), which eliminates the constraint \(\sum_k c_k = 1\). Since the estimated \(c_k\) indicates the relative amounts of the compounds, the composition ratios for the identified combination of compounds can be calculated by correcting \(c_k\) so that \(\sum_k g_k c_k = 1\), based on the obtained \(g_k\) and \(c_k\).

Applying Bayes’ theorem, as in the previous section, we can derive the following equation as the joint posterior probability \(p(g, \Theta|Y)\) of indicator \(g\) and parameter set \(\Theta\):

\[
p(g, \Theta|Y) \propto \varphi(g)\varphi(\Theta) \exp(-E(g, \Theta)),
\]

(20)

\[
E(g, \Theta) = -\sum_{l=1}^L \sum_{i=1}^{N_l} \{y_{il} \log(f(x_{il}; g, \Theta_l))
\]

\[
- f(x_{il}; g, \Theta_l) - \log(y_{il})\}.
\]

(21)

\[
f(x_{il}; g, \Theta_l) = S(x; g, \theta_{n_l}) + B(x; \theta_{n_l}, \theta_{B_l}).
\]

(22)

Using the above joint posterior distribution, we first estimate the indicator \(g\) from the following marginal posterior distribution:
However, where information.

\[ F(g) = -\log d\Theta \varphi(\Theta) \exp(-E(g, \Theta)). \]  

Using the integration on the right-hand side, we define the Bayesian free energy as [7,8]

\[ p(g | Y) = \int d\Theta p(g, \Theta | Y) \]

\[ = \varphi(g) \int d\Theta \varphi(\Theta) \exp(-E(g, \Theta)). \]

If the prior \( \varphi(g) \) is uniform, maximization of the marginal posterior distribution is equivalent to minimization of the Bayesian free energy [9]. This framework makes it possible to identify combinations of compounds. After the indicator is estimated, other parameters are estimated by using the posterior \( p(\Theta | g, Y) \) over the parameter set \( \Theta \).

\[ p(\Theta | g, Y) = \frac{p(g, \Theta | Y)}{p(g | Y)} \]

\[ \propto \varphi(\Theta) \exp(-E(g, \Theta)) \]

This estimation is equivalent to that in the previous section.

In Experiment 2, we perform simulations on this model and discuss the estimation accuracy for compound identification and the estimation accuracy of the parameter set \( \Theta \).

3. Sampling algorithm

In Experiment 1, we sample parameters from Equation (18) and perform maximum a posteriori (MAP) estimation. We use the MC method as a sampling method. However, since the MC method can lead to local minima depending on the initial values, we prepare the following replicas [10]:

\[ q(\Theta; \beta) = \varphi(\Theta) \exp(-\beta E(\Theta)) \]

where \( \beta \) is an auxiliary variable called the inverse temperature and we can see that \( q(\Theta; \beta = 1) = p(\Theta | X, Y) \).

Here, we prepare the inverse temperature sequence \( 0 = \beta_1 < \beta_2 < \cdots < \beta_s = 1 \) and sample the parameter set \( \Theta \) by the exchange MC method using these replicas.

In this framework, we can calculate the free energy and then select an optimal model, but it is difficult to calculate the free energy for all the candidate model combinations because of their large number. In Experiment 2, we sample the indicator parameter \( g \) as well as the parameter set \( \Theta \) from Equation (20). Also, the exchange MC method is used in Experiment 2 for sampling. The probability distribution of each replica is

\[ q(g, \Theta; \beta) = \varphi(g) \varphi(\Theta) \exp(-\beta E(g, \Theta)) \]

\[ \propto \varphi(\Theta) \exp(-\beta E(g, \Theta) \beta \varphi(\Theta) \exp(-E(g, \Theta)). \]

After simultaneous sampling of the indicator parameter \( g \) and the parameter set \( \Theta \) is obtained, the compound is identified using the most sampled indicator parameter as \( g^* \) and MAP estimation for \( \Theta \) is performed according to the conditional posterior probability \( p(\Theta | g, Y) \).

For each replica, the parameters are updated by the Metropolis algorithm for a certain number of steps. In particular, for the parameter \( c \) with the constraint of \( \sum_k c_k = 1 \), when updating \( c_k \), we simultaneously updated \( c_{k'} \) where \( k \neq k' \). After that, we exchange replicas between adjacent inverse temperatures. By doing this for a sufficient number of steps, we can sample the parameters without falling into a local solution [10].

To update the parameters, adaptive step-size tuning is used [11,12]. By this method, the step size of continuous variables is tuned so that the acceptance rate becomes the target value set.

We used the computer which has Intel(R) Xeon(R) CPU E5-2697A v4 at 2.60 GHz CPU with 2 cores and 251GB memory.

4. Experiment 1: composition ratio estimation

Following the model described in Section 2.1, we generate artificial data corresponding to three orbitals, Sr3d, Ti2p and O1s, assuming a sample containing four compounds, SrCO3, SrO, SrTiO3 and TiO2. For the parameters of the reference spectrum described in Equation (7), the peak position \( \mu \) and the area ratio \( m \) are taken from the literature [13–19], the peak width \( w \) and Lorentzian width \( r \) are set to 0.6 and 0.1, respectively, and \( A \) is determined according to the compound. The composition ratios are set to 0.1, 0.3, 0.4 and 0.2 for the above four compounds, respectively. The true values of the other parameters are set as

\[ a^* = \begin{pmatrix} 2.00 \times 10^2 \\ 2.00 \times 10^2 \\ 2.00 \times 10^2 \end{pmatrix}, \quad b^* = \begin{pmatrix} 1.00 \times 10^2 \\ 1.00 \times 10^2 \end{pmatrix}, \]

\[ h^* = \begin{pmatrix} 1.00 \times 10^4 \\ 1.00 \times 10^4 \end{pmatrix}, \quad \hat{\mu}^* = \begin{pmatrix} 0.353 \\ 0.080 \\ 0.196 \\ 0.448 \end{pmatrix}, \]

\[ w^* = \begin{pmatrix} 1.09 & 0.96 & 1.00 & 1.02 \\ 1.04 & 1.01 & 1.02 & 1.02 \\ 1.07 & 0.99 & 0.89 & 1.03 \end{pmatrix}. \]

Note that the true parameter set is \( \Theta^* = \{ \{ a_j^*, b_j^* \}_j \}_i, \{ \hat{c}_k \}, \{ \hat{\mu}_{k_l} \}_l, \{ \hat{w}_{k_l} \}_l \}_i, \) with \( i = 1, 2, 3 \) as Sr3d, Ti2p, O1s and \( k = 1, 2, 3, 4 \) as SrCO3, SrO, SrTiO3, TiO2, respectively. Figure 1 shows the actual
Table 1. Comparison between true value and MAP estimation of parameters. \( l = 1, 2, 3 \) as Sr3d, Ti2p and O1s and \( k = 1, 2, 3, 4 \) as SrCO\(_3\), SrO, SrTiO\(_3\) and TiO\(_2\), respectively.

| Parameter | Index | True Value | MAP Estimation | Parameter | Index | True Value | MAP Estimation |
|-----------|-------|------------|----------------|-----------|-------|------------|----------------|
| \( c_k \) | \( k = 1 \) | 0.100 | 0.090 | \( \mu_k \) | \( k = 1 \) | 0.100 | 0.090 |
| \( k = 2 \) | 0.300 | 0.328 | \( k = 2 \) | 0.300 | 0.328 |
| \( k = 3 \) | 0.400 | 0.386 | \( k = 3 \) | 0.400 | 0.386 |
| \( k = 4 \) | 0.200 | 0.196 | \( k = 4 \) | 0.200 | 0.196 |
| \( \sigma_k \) | \( l = 1 \) | 2.00 \times 10^2 | 1.99 \times 10^2 | \( \omega_{\mu_k} \) | \( l = 1 \) | 1.09 | 0.93 |
| \( l = 2 \) | 2.00 \times 10^2 | 1.95 \times 10^2 | \( l = 1 \) | 1.02 | 0.94 |
| \( l = 3 \) | 1.00 \times 10^2 | 1.01 \times 10^2 | \( l = 2 \) | 1.02 | 0.99 |
| \( l = 4 \) | 1.00 \times 10^2 | 1.00 \times 10^2 | \( l = 3 \) | 1.02 | 0.99 |
| \( h_k \) | \( l = 1 \) | 1.00 \times 10^4 | 0.97 \times 10^4 | \( l = 3 \) | 0.99 | 1.00 |
| \( l = 2 \) | 1.00 \times 10^4 | 1.05 \times 10^4 | \( l = 2 \) | 0.89 | 1.04 |
| \( l = 3 \) | 1.00 \times 10^4 | 1.00 \times 10^4 | \( l = 3 \) | 1.03 | 0.98 |

The number of temperatures is set to 64. For the exchange MC simulation, we performed 100,000 steps of calculations and rejected 50,000 of them as burn-in for Experiment 1.

Table 1 shows the MAP estimations and Figure 2 shows the fitting results using the estimation. From Figure 2, we can see that the fitting is accurate. In addition, Table 1 shows that the MAP estimations are also reasonable. In fact, according to Figure 3(a), the composition ratio \( c \) can be estimated with high accuracy. Other parameters such as \( \mu, a, b \) and \( h \) are also estimated with high accuracy (Figure 3(b-e)). However, that of the peak width adjustment parameter \( \omega \) is not high (Figure 3(f-h)). This trend is confirmed throughout the entire experiment and it can be concluded that the peak width adjustment parameter is difficult to estimate for this model. This is an advantage of the Bayesian framework, i.e. we can discuss not only the estimation value but also its estimation accuracy by comparing the prior and posterior distributions. Next, we investigate the relationship between the signal-to-noise ratio and the estimation accuracy. The observation time is an important parameter in XPS measurement for adjusting the signal-to-noise ratio. Thus, we newly introduce an observation time.
parameter $T$ into the model described in Section 2.1 as follows:

$$a_l = a'_l \times T, \quad b_l = b'_l \times T, \quad h_l = h'_l \times T.$$ \hspace{1cm} (39)

$T$ is a latent parameter, which is consistent with the fact that the intensity of the XPS data is proportional to the observation time. If the observation time is $T = 1000$ for the data in Figure 1, the true values of $a'_l, b'_l$ and $h'_l$ are respectively 0.2, 0.1 and 10.0 for all $l$. Here, we generate new data for $T = 10, 100$ and 10,000 for the true values of $a'_l, b'_l$ and $h'_l$. For these data, we simulate the Bayesian estimation for the parameter $\Theta^* = \{a'_l, b'_l, h'_l\}_{l=1}^{5}$, $\{c^*_k, \hat{a}_k, \{\hat{w}_{lk}\}_{l=1}^{3}\}_{k=1}^{4}$ using the reference spectra of the four compounds, SrCO$_3$, SrO, SrTiO$_3$ and TiO$_2$.

Figure 4 and Table 2 respectively show the fitting results and MAP estimates of the parameters for these data. Furthermore, Figure 5 shows the posterior distributions of the composition ratios for each observation time. As can be seen from the overall results, the fitting results and the accuracy of the estimation of the composition ratios improve as the observation time $T$ increases. Table 3 shows the mean and standard deviation of the posterior distributions of each parameter when fitted to a normal distribution as an index for discussing the estimation accuracy quantitatively. In

Figure 2. Results of fitting by MAP estimation for Sr3d orbital (a), Ti2p orbital (b) and O1s orbital (c).

Figure 3. Posterior distributions for each parameter. The vertical axis represents the number of times sampled and the dotted line represents the true value of each parameter. (a) Composition ratio parameter, (b) peak position adjustment parameter, (c) and (d) background endpoint intensities, (e) intensity parameter and (f), (g) and (h) peak width adjustment parameters.
Figure 4. Fitting results for each observation time. The first column from the left is $T = 10$, the second column is $T = 100$, the third column is $T = 1000$ and the fourth column is $T = 10000$.

Table 2. Comparison between true value and MAP estimation of composition ratios for each compound.

| compound | true value | MAP estimation |
|----------|------------|----------------|
| SrCO$_3$ | 0.100 | 0.100 | 0.094 | 0.090 | 0.099 |
| SrO    | 0.300 | 0.291 | 0.216 | 0.328 | 0.297 |
| SrTiO$_3$ | 0.400 | 0.343 | 0.403 | 0.386 | 0.402 |
| TiO$_2$ | 0.200 | 0.266 | 0.287 | 0.196 | 0.202 |

particular, the results in Figure 5 suggest that $T = 10000$ is necessary to obtain sufficient estimation accuracy for composition ratio estimation. From Table 2, it can be seen that the estimation of SrCO$_3$ is adequate for all observation times $T$. This may be due to the fact that SrCO$_3$ is necessary to represent the leftmost peak component in the O1s data.

5. Experiment 2: compound identification

For the estimation in Experiment 1, we use only the reference spectra of the compounds contained in the sample. In Experiment 2, we use the reference spectra of candidate compounds according to the model described in Section 2.2 to identify the compounds contained in the observed sample and to estimate their composition ratios. In addition to the four compounds used in Experiment 1, SrCO$_3$, SrO, SrTiO$_3$ and TiO$_2$, six compounds, namely SrSO$_4$, TiO, Ti$_2$O$_3$, Sr, TiC and Ti, are used as candidates for compound identification from the data shown in Figure 1. For the parameters of the reference spectrum of the newly added compound, the peak position $\mu$ and the area ratio $m$ are taken from the literature [19–26], the peak width $w$ and Lorentz ratio $r$ are set to 0.6 and 0.1, respectively, and $A$ is determined according to the compound. This means that the total number of candidate compounds is 10. The true values of the indicator parameters defined in Equation (19) are $g_k^c = 1$ for $k = 1, 2, 3, 4$ and $g_k^c = 0$ for the others. The prior distribution of $c_k$ corresponding to the composition ratio is as

Figure 5. Posterior distributions of the composition ratios $c$ for each observation time.
Table 3. Mean and standard deviation of the posterior distribution of the composition ratio of each compound at each observation time, fitted with a normal distribution. From this result, we can quantitatively say that the estimation accuracy improves as the observation time increases.

| compound    | observation time | mean  | standard deviation |
|-------------|------------------|-------|--------------------|
| SrCO$_3$    | $T = 10$         | 0.122 | 0.034              |
|             | $T = 100$        | 0.094 | 0.011              |
|             | $T = 1000$       | 0.094 | 0.005              |
|             | $T = 10000$      | 0.097 | 0.002              |
| SrO         | $T = 10$         | 0.368 | 0.085              |
|             | $T = 100$        | 0.279 | 0.027              |
|             | $T = 1000$       | 0.309 | 0.010              |
|             | $T = 10000$      | 0.299 | 0.004              |
| SrTiO$_3$   | $T = 10$         | 0.327 | 0.040              |
|             | $T = 100$        | 0.390 | 0.014              |
|             | $T = 1000$       | 0.392 | 0.006              |
|             | $T = 10000$      | 0.403 | 0.002              |
| TiO$_2$     | $T = 10$         | 0.183 | 0.070              |
|             | $T = 100$        | 0.237 | 0.023              |
|             | $T = 1000$       | 0.205 | 0.008              |
|             | $T = 10000$      | 0.201 | 0.003              |

Figure 6. Posterior distribution of the combination of compounds. The vertical axis shows the combinations of compounds for which the indicator parameter is 1 and the horizontal axis shows the posterior probability. The probability of the true combination of compounds is 93.0%. Note that the labels on the vertical axis are sorted in the order of the most sampled number of combinations.

$$
\varphi(c_k) = \text{Gamma}(c_k k_\nu, \lambda_\nu),
$$

(40)

$$
k_\nu = 2.0, \quad \lambda_\nu = 0.1.
$$

(41)

The number of temperatures is set to 64. For the exchange MC simulation, we performed 50,000 steps of calculations and rejected 20,000 of them as burn-in for Experiment 2.

Figure 6 shows the posterior distribution of the combination of compounds. The posterior probability is calculated by using the sampling result of the indicator parameter from the exchange MC method. From this result, we can see that the true compound combination, SrCO$_3$, SrO, SrTiO$_3$ and TiO$_2$, is estimated with more than 90% probability. Table 4 shows a comparison between the true value and MAP estimation for the composition ratio. Figure 7 shows the fitting result obtained by MAP estimation for parameter $\Theta$, which is well fitted with the given data. Figure 8 shows the posterior distribution of the composition ratios $c_k$, which is the similar result as in Figure 3(a) and indicates that the estimation accuracy of the composition ratio is satisfactory.

Furthermore, we investigate how the accuracy of compound identification changes with the measurement time. As in Experiment 1, the observation time $T$ is introduced as in Equation (39), and compound identification is performed for data with $T = 10$, 100, 1000, 10,000. Figure 9 shows the posterior distribution of compound identification at various observation times $T$. In addition, Figure 10 shows the conditional posterior distribution of the composition ratios for the true compound combination. As can be seen from Figures 9 and 10, as the observation time decreases, the estimation accuracy of both compound identification and composition ratio estimation deteriorates. The result in Figure 10 shows that $T = 1000$ is necessary to obtain sufficient estimation accuracy of the composition ratios. On the other hand, in Figure 9, even at $T = 10$, the true compound combination shows the highest posterior

| compound    | true value | MAP estimation |
|-------------|------------|----------------|
| SrCO$_3$    | 0.100      | 0.102          |
| SrO         | 0.300      | 0.307          |
| SrTiO$_3$   | 0.400      | 0.398          |
| TiO$_2$     | 0.200      | 0.194          |

Table 4. Comparison between true value and MAP estimation of composition ratios for each compound.

Figure 7. Results of fitting by the MAP estimation for Sr3d orbital (a), Ti2p orbital (b) and O1s orbital (c).
probability, 43.6%, 88.8%, 83.0%, 99.5% for $T = 10, 100, 1000, 10,000$, respectively. This indicates that compound identification is easier than the estimation of composition ratio with sufficient accuracy. This fact was obtained by investigating the estimation accuracy from Bayesian estimation.

Figure 8. Conditional posterior distributions of the composition ratios for the true combination of compounds. In Experiment 2, the composition ratios are calculated by normalizing so that $\sum_k c'_k = 1$.

Figure 9. Posterior distributions of the combination of compounds for each observation time. The true combination of compounds is estimated for all observation times with probabilities of 43.6%, 88.8%, 93.0%, 99.5% in (a)-(d) respectively. Note that the labels on the vertical axis are sorted in the order of the most sampled number of combinations.

6. Conclusion

In this paper, we proposed a framework of Bayesian inference for identifying compounds and estimating their composition ratios by exploiting the reference spectra of compounds in XPS spectral analysis. Also, we introduced indicator parameters in the compound identification and realized an efficient model selection algorithm by simultaneously sampling the indicators in the MC method. In a case like the present study, we have shown that not only the parameter estimates but also their confidence intervals can be obtained from the given data by using Bayesian estimation and comparing the prior and posterior distributions. As a result, that it is possible to estimate the observation time required for composition ratio estimation by looking at the shape of the posterior probability distribution of the parameters. Furthermore, it was found that the signal-to-noise ratio required for compound identification is smaller than that required for accurate composition ratio estimation under conditions such as those in this study, where the compound has peaks that appear in multiple energy ranges. This result is consistent with that of Nagata [27]. The investigation of other cases is a future issue.
This result was based on artificial data but it is also expected to be applicable to real data. Such verification with real data and developing a system that provides feedback to the instrument are a future task. To achieve more realistic modeling of XPS, it is important not only to develop estimation methods using convolution Voigt functions instead of pseudo-Voigt functions but also to explore the design of the prior distribution and hyperparameters.

Disclosure of potential conflicts of interest
No potential conflict of interest was reported by the authors.

Funding
This work was supported by CREST (JP MJCR1761) from the Japan Science and Technology Agency (JST).

ORCID
Masato Okada https://orcid.org/0000-0002-9040-8784

References
[1] Murakami R, Tanaka H, Shinotsuka H, et al. Development of multiple core-level XPS spectra decomposition method based on the Bayesian information criterion. J Electron Spectrosc Relat Phenom. 2020;245:147003.
[2] Nagata K, Sugita S, Okada M. Bayesian spectral deconvolution with the exchange Monte Carlo method. Neural Netw. 2012;28:82–89.
[3] 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.
[4] Herrera-Gomez A, Bravo-Sanchez M, Ceballos-Sanchez O, et al. Practical methods for background subtraction in photoemission spectra. Surf Interface Anal. 2016;48:897–905.
[5] Matsumoto R, Nishizawa Y, Kataoka N, et al. Reproducibility of XPS analysis for film thickness of SiO2/Si by active Shirley method. J Electron Spectrosc Relat Phenom. 2016;207:55–59.
[6] Murakami R, Kageyama H, Nakamura K, et al. Background estimation in X-ray photoelectron spectroscopy data using an active Shirley method with automated selection of the analytical range. e-J Surf Nanotechnol. 2019;17:61–68.
[7] Akaike H. Likelihood and the Bayes procedure. Trabajos de Estadistica y de Investigacion Operativa. 1980;31:143–166.
[8] Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6(2):461–464.
[9] Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 1995;82(4):711–732.
[10] Hukushima K, Nemoto K. Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc Jpn. 1996;65(6):1604–1608.
[11] Garthwaite PH, Fan Y, Sisson A. Adaptive optimal scaling of Metropolis–Hastings algorithms using the Robbins–Monro process. Commun Stat. 2016;45(17):5098–5111.
[12] Okajima K, Nagata K, Okada M. Fast Bayesian deconvolution using dimple reversible jump moves. J Phys Soc Jpn. 2021;90(3):034001.
[13] Pilleux ME, Grabmann CR, Fuenzalida VM. Hydrothermal strontium titanate films on titanium: an XPS and AES depth-profiling study. J Am Ceram Soc. 1994;77(6):1601–1604.
[14] Sosulnikov MI, Teterin YA. X-ray photoelectron study of calcium, strontium, barium and their oxides. Dokl Akad Nauk SSSR. 1991;317:418–421.
[15] Aas N, Pringle TJ, Bowker M. Adsorption and decomposition of methanol on TiO2, SrTiO3, and SrO. J Chem Soc Faraday Trans. 1994;90(7):1015–1022.
[16] Chaschechnikova IT, Vorontyshev VM, Borovik VV, et al. Strong metal-carrier interaction in cobalt- and nickel-titanium dioxide co-hydrogenation catalysts. Theor Exp Chem. 1993;28:176–178.
[17] Vasquez RP. X-ray photoelectron spectroscopy study of Sr and Ba compounds. J Electron Spectrosc Relat Phenom. 1991;56(3):217–240.
[18] Lange F, Schmelz H, An KH. X-ray photoelectron spectroscopy study of oxides of arsenic supported on TiO2. J Electron Spectrosc Relat Phenom. 1991;57(3):307–315.
[19] Briggs D, Grant JT. Surface analysis by Auger and X-ray photoelectron spectroscopy. UK: IM Publications and SurfacedataScientificlimited; 2003. p. 37.
[20] Galuska AA, Uht JC, Marquez N. Reactive and non-reactive ion mixing of Ti films on carbon substrates. J Vac Sci Technol A. 1988;6:110–122.
[21] Simon D, Perrin C, Bardolle J. ESCA study of Nb and Ti oxides. Applications to the determination of the nature of the superficial films formed during the oxidation of Nb–Ti and Nb–Ti alloys. J de Microscopie et de Spectroscopie Electron. 1976;12(2):175–186.
[22] Gonbeuf D, Guimon C, Pifster-Guillouzo G, et al XPS study of thin films of titanium oxysulfides. Surf Sci. 1991;254(1–3):81–89.
[23] Badrinarayanan S, Sinha S, Mandale AB. XPS studies of nitrogen ion implanted zirconium and titanium. J Electron Spectrosc Relat Phenom. 1989;49(3):303–309.

[24] Kuznetsov MV, Zhuravlev JF, Gubanov VA. XPS analysis of adsorption of oxygen molecules on the surface of Ti and TiN\textsubscript{x} films in vacuum. J Electron Spectrosc Relat Phenom. 1992;58(3):169–176.

[25] Werfel F, Brummer O. Corundum structure oxides studied by XPS. Phys Scr. 1983;28(1):92–96.

[26] Christie AB, Lee J, Sutherland I, et al. An XPS study of ion-induced compositional changes with group II and group IV compounds. Appl Surf Sci. 1983;15(1):224–237.

[27] Nagata K, Muraoka R, Mototake Y, et al. Bayesian spectral deconvolution based on poisson distribution: bayesian measurement and virtual measurement analytics (VMA). J Phys Soc Jpn. 2019;88:44003.