Journal of Electron Spectroscopy and Related Phenomena 273 (2024) 147449

Contents lists available at ScienceDirect

Journal of Electron Spectroscopy and Related Phenomena

journal homepage: www.elsevier.com/locate/elspec

Bayesian inference method utilizing SESSA in quantitative layer structure estimation from XPS data

Atsushi Machida [a], Kenji Nagata [b], Ryo Murakami [c], Hiroshi Shinotsuka [c], Hayaru Shouno [d], Hideki Yoshikawa [b,c], Masato Okada [c,e,*]

[a] Graduate School of Science, The University of Tokyo, Bunkyo, Japan
[b] Center for Basic Research on Materials, National Institute for Materials Science, Tsukuba, Japan
[c] Materials Data Platform, Research Network and Facility Services Division, National Institute for Materials Science, Tsukuba, Japan
[d] Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu, Japan
[e] Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan

ARTICLE INFO

Keywords:
X-ray photoelectron spectroscopy
Bayesian estimation
Exchange Monte Carlo method
SESSA

ABSTRACT

X-ray photoelectron spectroscopy (XPS) is a surface analysis technique for the nondestructive identification of elemental species and chemical states of solid samples, and the measured spectra are affected by not only sample-specific information but also factors dependent on the measurement environment. This feature makes it difficult to analyze the data for the chemical state identification of mixed samples when referring to the data measured with different models or in different environments. In a previous study, Bayesian inference was successfully applied to the analysis of XPS narrow-scan spectra, but the challenge was to apply Bayesian inference to XPS spectra of samples that are nonuniform in the depth direction. We propose a method to infer the layer structure of a sample from XPS spectra by incorporating Bayesian inference into the simulation of electron spectra for surface analysis (SESSA). SESSA can simulate XPS spectra of samples with specified composition and microstructure, and is already in use as a simulator with highly reproducible results. By utilizing the proposed method, one can estimate the layer structure of a sample from XPS data on the basis of the posterior probability distribution. In a typical XPS measurement, wide-scan data are acquired to qualitatively identify elemental species, and narrow-scan data are acquired to the estimate detailed composition and chemical state information of a sample. In this study, we have shown that given wide-scan or narrow-scan data without angle resolution, Bayesian inference can be applied to quantitatively analyze the layer structure information.

1. Introduction

X-ray photoelectron spectroscopy (XPS) is a surface analysis technique to nondestructively identify elemental species and their chemical states in solid samples through of the measurement of the energy spectrum of photoelectrons emitted by a sample irradiated with X-rays. XPS is characterized by the fact that not only the intrinsic sample-specific information but also extrinsic factors such as charging and energy resolution, which depend on the measurement environment, affect the actual measurement spectrum. This feature of XPS makes it difficult to analyze the data when identifying the chemical states contained in a mixed sample by referring to the individual chemical state XPS spectra measured in different environments in the literature. To solve this problem, Machida et al. studied the application of Bayesian inference to spectral data analysis in XPS, and they succeeded in automating data analysis for chemical state identification and in quantitatively determining its accuracy [1]. In this study, in the XPS data generation process, the sample is assumed to be spatially uniform and the background due to inelastic photoelectron scattering is assumed to be generated by the approximate formula of the Shirley method, and the usefulness of this model is demonstrated by analyzing artificial data. On the other hand, this generative model could not analyze the microstructural structure of a sample that is nonuniform in the depth direction, such as a sample with having a layered structure. Angle-resolved XPS is commonly used for the nondestructive analysis of the microstructural structure of samples that are nonuniform in the depth direction [2–4]. On the other hand, as is well known to XPS experts, it is possible to estimate the depth-directed microstructural information of a sample from only its wide-scan spectrum without angle resolution with a wide energy range [5]. Wide-scan spectra are often limited to qualitative analysis owing to the many and complex factors that generate them, but they are rich in depth information. In a previous study [1], we applied Bayesian inference methods to

* Corresponding author.
E-mail address: okada@edu.k.u-tokyo.ac.jp (M. Okada).

https://doi.org/10.1016/j.elspec.2024.147449
Received 9 February 2024; Received in revised form 30 April 2024; Accepted 22 May 2024
Available online 29 May 2024
0368-2048/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the analysis of XPS narrow-scan spectra of the inner-shell levels of each element in the mixed samples that are uniform in the depth direction and confirmed their validity for the component identification of the samples. In this study, we applied Bayesian inference methods to XPS wide-scan spectra of mixed samples of nonuniform systems consisting of even more complex factors, enabling the data analysis of wide-scan spectra, which is often limited to qualitative analysis in the past, leading to quantitative analysis. The background of spectral data is correlated with the layer structure of the sample as it covers a wide energy range from the peak to a high binding energy [6], and therefore in this study, the background of a wide-scan spectrum was included in the analysis target of Bayesian inference. However, the Shirley method used in our previous study [1] is an approximate for the background of a narrow-scan spectral model and cannot reproduce the background over a wide-scan, which means that it is difficult to estimate the layer structure of a sample. To overcome the above-mentioned issues, we turned to the simulation of electron spectra for surface analysis (SESSA), developed by Smekal and coworkers [7,8]. SESSA contains the physical data necessary to quantitatively interpret XPS spectra of samples with a specified composition and microstructure, and it can simulate the XPS spectra for user-defined sample structures and measurement conditions, and the simulated XPS spectra can be compared with the measured spectra. It also simulates not only peaks of photoelectrons and Auger electrons but also the background due to the inelastic scattering of signal electrons; thus, it has been used as a simulator with highly reproducible results of analyses that use XPS data.

In this study, we propose a Bayesian inference method that employs SESSA as a generative model. Bayesian inference is a framework for probabilistically inferring variables, assuming that model parameters behave probabilistically. Also, by probabilistically modeling the data generation process, we can calculate the likelihood probability of obtaining data based on certain parameters, and by applying Bayes' theorem, we can get the posterior probability of the parameters when certain data is obtained [9,10]. By combining this framework with SESSA, it is possible to estimate sample information with reliability evaluation from a variety of sample candidates. To incorporate SESSA into Bayesian inference, the graphical user interface-based software was replaced by a virtual window system that can be run from the command line. Exchange Monte Carlo methods, which are widely used in statistical mechanics and machine learning to efficiently sample probability distributions in high-dimensional spaces by running multiple Markov chains simultaneously and exchanging states, were used to derive the posterior probability distribution of the parameters related to the layered structure of a sample from the given data. By using the method proposed in this study, one can estimate the layer structure of a sample from XPS data via the posterior probability distribution with confidence intervals. In a typical XPS measurement, wide-scan spectra are acquired for the qualitative identification of elemental species, including checking for the presence of impurities, and narrow-scan spectra are acquired to estimate detailed information on the composition and chemical state of the sample. We have confirmed the usefulness of the proposed method by conducting simulations in which Bayesian inference is applied to situations in which wide-scan and narrow-scan data without angle-resolved measurements are given separately, and by examining the extent to which the quantitative analysis of layer structure information is possible from both approaches.

2. Methods

2.1. Data generative model

In this study, artificial data from SESSA are used for the analysis. In this section, we assume a layered sample and explain the process of data generation from the assumed sample. For a layered sample, three assumptions are made. First, the interface of each layer is steep. Second, the substate is pure Si and the number of layers on the Si substrate is known in advance, but not the thicknesses of layers. Third, the elemental species composing each layer are also known in advance, but not the composition. From these assumptions, the parameter set for the sample to be estimated is $\theta_{sample} = \{T_j, \{x_{j,m}\}_{m=1}^{M_j}\}_{j=1}^{L}$, where L is the number of layers excluding the Si substrate and M_j is the number of constituent elements in each layer j. T_j is the thickness of layer j and $x_{j,m}$ is the ratio of element m in layer j. The elemental ratio is constrained to be $\sum_m^{M_j} x_{j,m} = 1$ for all j. Two points should be noted here. First, the output directly obtained from SESSA has an arbitrary unit on the vertical axis and does not reflect the actual intensity of the measurement. Therefore, in this study, an intensity correction parameter h was introduced and multiplied directly to the output from SESSA to adjust it. Second, the background of wide-scan spectra was calculated by combining peak backgrounds and matrix background because SESSA has an upper limit of the total number of energy points in a simulated spectrum, and wide-scan and narrow-scan spectra with background were calculated separately. Unlike the previous study [1], SESSA also simulates the backgrounds due to inelastic scattering from peak components within a specified energy range. However, when simulating over a specified energy range, the effect of the matrix background from peaks at energies lower than the specified range is not taken into account. Therefore, another simulation was performed and this was added as a correction. A constant background of a few counts, c, was also added as a correction based on the assumption that noise is generated by the detector, assuming the experimental case. In summary, the spectral intensity $f(x;\boldsymbol{\Theta})$ at energy x under the given parameter set $\boldsymbol{\Theta}$ can be formulated as

$$f(x;\boldsymbol{\Theta}) = \left\{S(x;\theta_{sample}) + B(x;\theta_{sample}, c)\right\} \times h, \tag{1}$$

where h is an intensity correction parameter for adjusting SESSA output to the measurement data and varies depending on the irradiated X-ray intensity and measurement time. The parameter set in the data generative model is $\boldsymbol{\Theta} = \{\theta_{sample}, h, c\}$. However, in the case of narrow-scan data, the noise due to the detector is assumed to be sufficiently small and $c = 0$, and only background correction by SESSA is applied to generate the data. In the case of wide-scan data, all peaks are included in the range, i.e., the entire matrix background is taken into account, and no constant background correction is performed by SESSA. The spectral intensity defined in Eq. (1) corresponds to the expected value in the measurement and does not take into account the statistical noise from the observations. Since the XPS data are usually count data under the experimental condition of pulse counting detection, the intensity y at the energy x under the given parameter set $\boldsymbol{\Theta}$ follows the Poisson distribution shown below:

$$p(y|\boldsymbol{\Theta}) = \frac{f(x;\boldsymbol{\Theta})^y \exp\left(-f(x;\boldsymbol{\Theta})\right)}{y!}. \tag{2}$$

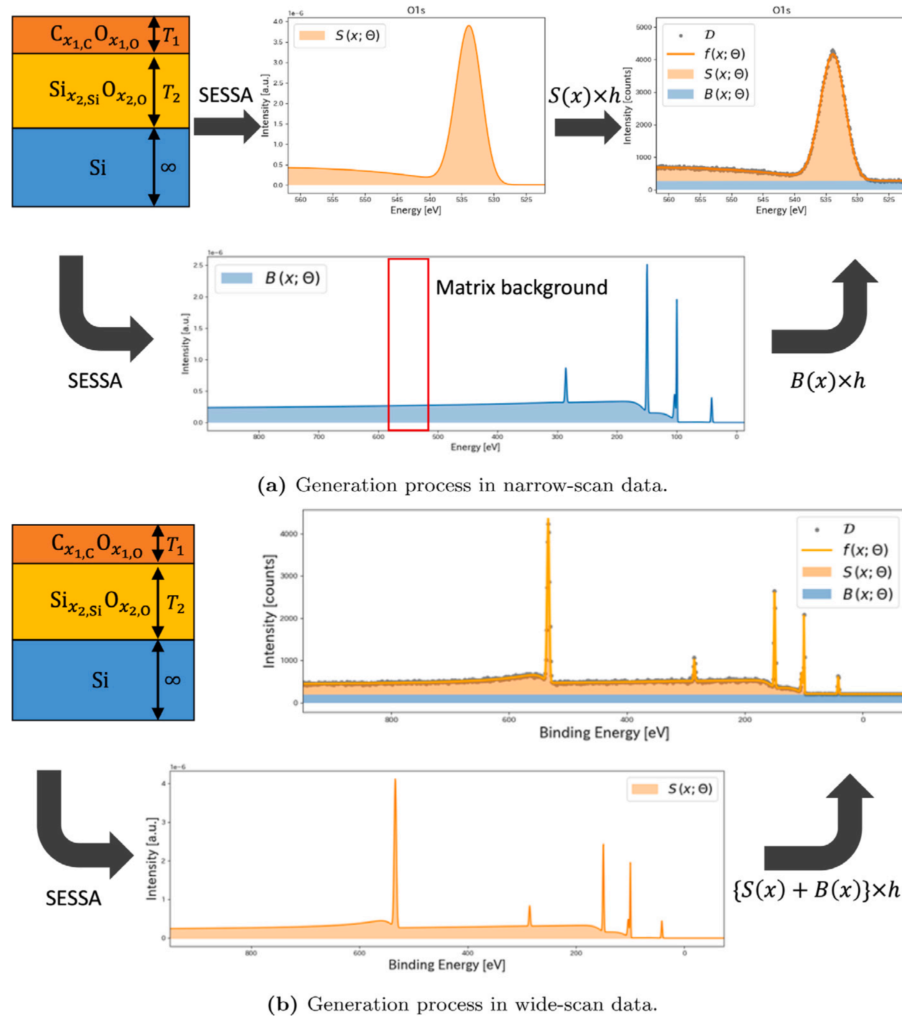
Fig. 1 shows the actual process used to generate the dataset $\mathcal{D}$. The dataset $\mathcal{D}$ is generated from a Poisson distribution with $f(x;\boldsymbol{\Theta})$ as the expected value, where $f(x;\boldsymbol{\Theta})$ consists of two components, $S(x;\boldsymbol{\Theta})$ and $B(x;\boldsymbol{\Theta})$.

2.2. Estimation model

In the wide-scan case, the data generation process assumed in the estimation employs the model described in Section 2.1. In the narrow-scan case, however, the generation of the spectrum requires two SESSA simulations, which is computationally time-consuming. Therefore, in the estimation, we assumed the following matrix background:

$$B_{narrow}(x;\boldsymbol{\Theta}) = l + \frac{x - x_{min}}{x_{max} - x_{min}}(r - l), \tag{3}$$

where x_{min} and x_{max} are respectively the minimum and maximum values of energies in the range and $\{l, r\}$ is the background intensity at the endpoints. Note that when estimating using this model, the parameter set is $\boldsymbol{\Theta} = \{\theta_{sample}, h, l, r\}$, with no true value for $\{l, r\}$. Given



(a) Generation process in narrow-scan data.

(b) Generation process in wide-scan data.

Fig. 1. Generation process for the data used in this study. For clarity, the background used in Fig. 1(b) is shown at a higher intensity than that used in the estimation experiment, also such a large detector-dependent background is not realistic in actual XPS measurements. Also, h is the intensity parameter corresponding to the measurement time.

a dataset $\mathcal{D}$ with N data, the likelihood of the data given the parameter set $\boldsymbol{\Theta}$ is as follows:

$$p(\mathcal{D}|\boldsymbol{\Theta}) = \prod_{i=1}^{N} p(y_i|\boldsymbol{\Theta}) = \prod_{i=1}^{N} \frac{f(x_i;\boldsymbol{\Theta})^{y_i} \exp\left(-f(x_i;\boldsymbol{\Theta})\right)}{y_i!}. \tag{4}$$

Bayes' theorem yields the following posterior distribution for the parameters when using this likelihood:

$$p(\boldsymbol{\Theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{p(\mathcal{D})}. \tag{5}$$

However, $p(\boldsymbol{\Theta})$ is a prior distribution with respect to the parameters, and Bayesian estimation can incorporate a priori information with respect to the parameters in the form of such a probability distribution in the estimation. $p(\mathcal{D}) = \int p(\mathcal{D}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}$ is a normalization constant independent of $\boldsymbol{\Theta}$. In the Bayesian estimation employed in this study, parameters are estimated probabilistically by obtaining this posterior distribution.

2.3. Exchange Monte Carlo method

If the parameters are multidimensional, it is difficult to obtain Eq. (5) analytically. Therefore, in this study, we prepared replicas given by the following equations and performed exchange Monte Carlo sampling [11]:

$$q(\boldsymbol{\Theta};\beta) = p(\boldsymbol{\Theta}) \exp\left(-\beta E(\boldsymbol{\Theta})\right), \tag{6}$$

$$E(\boldsymbol{\Theta}) = -\log\left(p(\mathcal{D}|\boldsymbol{\Theta})\right). \tag{7}$$

In Eq. (6), β is an auxiliary variable called the inverse temperature, which is used to relate the energy of the system to the probability distribution of states. In exchange Monte Carlo, multiple Markov chains are run at different inverse temperatures and the states are exchanged between the chains, resulting in efficient sampling. E is an error function, and $q(\boldsymbol{\Theta};\beta = 1) \propto p(\boldsymbol{\Theta}|\mathcal{D})$. The exchange Monte Carlo method enables global sampling without falling into local solutions by preparing the inverse temperature sequence $0 = \beta_1 < \beta_2 < \cdots < \beta_\tau = 1$ and performing Monte Carlo sampling for each replica, as well as exchanging between replicas. Exchange Monte Carlo methods are used to estimate the posterior distribution by sampling parameters that follow the posterior distribution.

2.4. Set up

In this study, the parameters related to peak width and peak position not explained above have the same as those referenced by default in SESSA (see Table 1).

We used the computer that has Intel(R) Xeon(R) CPU E5–2697A v4 at 2.60 GHz with 2 cores and 251 GB memory. Monte Carlo calculations were performed for 6000 steps for each replica, 3000 steps of which were rejected as burn-in. Thirty-two replicas were prepared, and the inverse temperatures were set as follows:

$$\beta_i = \gamma^{i-32} \quad (i = 2, 3, \ldots, 31). \tag{8}$$

Table 1
Parameters related to peak position and peak width referenced by default in SESSA.

Peak	Type	Position [eV]	Width [eV]
Si2s (Substrate)	Gauss	149.7	1.00
$Si2p_{1/2}$ (Substrate)	Gauss	100.3	0.58
$Si2p_{3/2}$ (Substrate)	Gauss	99.7	0.59
Si2s (2nd layer)	Gauss	149.7	1.00
$Si2p_{1/2}$ (2nd layer)	Gauss	103.7	1.39
$Si2p_{3/2}$ (2nd layer)	Gauss	103.7	1.39
O1s (2nd layer)	Gauss	533.0	1.70
O2s (2nd layer)	Gauss	41.6	1.00
C1s (1st layer)	Gauss	285.7	1.24
O1s (1st layer)	Gauss	534.7	1.80
O2s (1st layer)	Gauss	41.6	1.00

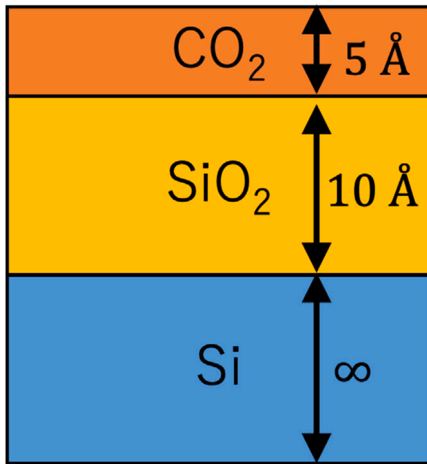


Fig. 2. Assumed sample: SiO_2 layer on a Si substrate with a thickness of 10 Å and CO_2 adsorbed on the surface with a thickness of 5 Å.

3. Experiments and discussion

In this study, we generated artificial data from a hypothetical sample as shown in Fig. 2, and conducted two experiments to estimate sample information using the proposed method. One is an estimation experiment from one-set wide-scan data, and the other is an estimation experiment from narrow-scan data focusing on three atomic orbitals, O1s, C1s, and Si2p. The parameter θ_{sample} and its true value θ^*_{sample} are as follows:

$$\theta_{sample} = \{T_1, T_2, x_{1,\mathrm{C}}, x_{1,\mathrm{O}}, x_{2,\mathrm{Si}}, x_{2,\mathrm{O}}\}, \tag{9}$$

$$\theta^*_{sample} = \{T^*_1, T^*_2, x^*_{1,\mathrm{C}}, x^*_{1,\mathrm{O}}, x^*_{2,\mathrm{Si}}, x^*_{2,\mathrm{O}}\}, \tag{10}$$

$$T^*_1 = 10\,\text{Å}, \quad T^*_2 = 5\,\text{Å}, \tag{11}$$

$$x^*_{1,\mathrm{C}} = 0.333333, \quad x^*_{1,\mathrm{O}} = 0.666666, \tag{12}$$

$$x^*_{2,\mathrm{Si}} = 0.333333, \quad x^*_{2,\mathrm{O}} = 0.666666. \tag{13}$$

3.1. Estimation from one-set wide-scan data

3.1.1. Artificial data and parameter set

First, estimation was performed from the wide-scan data shown in Fig. 3. The energy step in this data set is 1.0 eV. However, the true value of the intensity correction parameter was set to $h^* = 1.0 \times 10^9$ and the true value of the background correction parameter was set to $c^* = 2.0$. Since there is only one set of data, the dataset and parameter set for this experiment are as follows:

$$\mathcal{D} = \{x_i, y_i\}_{i=1}^{N}, \tag{14}$$

$$\boldsymbol{\Theta} = \{\theta_{sample}, h, c\}, \tag{15}$$

where N is the number of data points for the wide-scan data, and h and c are the intensity and background correction parameters, respectively. The prior distributions for each parameter were set as follows:

$$p(\boldsymbol{\Theta}) = p(T_1)p(T_2)p(x_{1,\mathrm{C}}, x_{1,\mathrm{O}})p(x_{2,\mathrm{Si}}, x_{2,\mathrm{O}})p(h)p(c), \tag{16}$$

$$p(T_1) = Gamma(T_1; n_T, \mu_T) \equiv \frac{1}{\Gamma(n_T)\mu_T^{n_T}} T_1^{(n_T-1)} \exp\left(-\frac{T_1}{\mu_T}\right), \tag{17}$$

$$p(T_2) = Gamma(T_2; n_T, \mu_T), \tag{18}$$

$$p(x_{1,\mathrm{C}}, x_{1,\mathrm{O}}) = Dirichlet(x_{1,\mathrm{C}}, x_{1,\mathrm{O}}; \alpha) \equiv \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2} x_{1,\mathrm{C}}^{\alpha-1} x_{1,\mathrm{O}}^{\alpha-1}, \tag{19}$$

$$p(x_{2,\mathrm{Si}}, x_{2,\mathrm{O}}) = Dirichlet(x_{2,\mathrm{Si}}, x_{2,\mathrm{O}}; \alpha), \tag{20}$$

$$p(h) = Gamma(h; n_h, \mu_h), \tag{21}$$

$$p(c) = Gamma(c; n_c, \mu_c), \tag{22}$$

$$n_T = 6, \quad \mu_T = 1.5, \quad \alpha = 1, \tag{23}$$

$$n_h = n_c = 11, \tag{24}$$

$$\mu_h = 1.0 \times 10^8, \tag{25}$$

$$\mu_c = 2.0 \times 10^{-1}. \tag{26}$$

Since all parameters are non-negative, a gamma distribution was used. However, the Dirichlet distribution was used for the elemantal ratios because of the limitation that the sum of the elemantal ratios must be 1 in each layer.

3.1.2. Results

Fig. 4 shows the posterior distributions and maximum a posteriori (MAP) estimates of each sample parameter obtained by Bayesian estimation. MAP estimation is a method for estimating the value that maximizes the posterior probability.

Fig. 4 also shows that sampling for both elemental ratio and thickness is concentrated around the true value. The MAP estimates do not always agree with the true values, but this can be attributed to the effect of noise in the data. On the other hand, the distribution of the elemental ratios of the CO_2 layer is more extensive than that of other parameters, indicating that the accuracy is low. The peak at around 290 eV in Fig. 3 corresponds to that derived from the C1s orbital, indicating that the data intensity is low. It is considered that the insufficient intensity here makes it difficult to accurately estimate the elemental ratio of the C elements. Below are the results of fitting the data using the MAP estimates (see Fig. 5). It can be seen that the fitting can be performed correctly on the data.

3.2. Estimation from narrow-scan data

3.2.1. Artificial data and parameter set

Next, an estimation experiment was conducted using the three sets of narrow-scan data shown in Fig. 6. The energy step in these data sets is 0.10 eV. Since there are intensity parameter h and background correction parameters l and r for each set of narrow-scan data, the parameter set θ_{sample} in this experiment is as shown in Eq. (27).

$$\boldsymbol{\Theta} = \{\theta_{sample}, \{h_k, l_k, r_k\}_{k=\mathrm{O1s,C1s,Si2p}}\} \tag{27}$$

Note that there are no true values for the background correction parameters l and r, and the true values of the intensity correction parameters are $h^*_{\mathrm{O1s}} = 1.0{\times}10^9$, $h^*_{\mathrm{C1s}} = 5.0{\times}10^9$ and $h^*_{\mathrm{Si2p}} = 2.5{\times}10^9$. Since the estimation is performed simultaneously from three sets of data, the likelihood is as in Eq. (28). The dataset is $\mathcal{D} = \{\{x_i, y_i\}_{i=1}^{N_k}\}_{k=1}^{3}$, where k is the label corresponding to the three atomic orbitals O1s, C1s, and Si2p, and N_k is the number of data points in the narrow-scan data for atomic orbital k.

$$p(\mathcal{D}|\boldsymbol{\Theta}) = \prod_{k=1}^{3}\prod_{i=1}^{N_k} p(y_i|\boldsymbol{\Theta}) \tag{28}$$

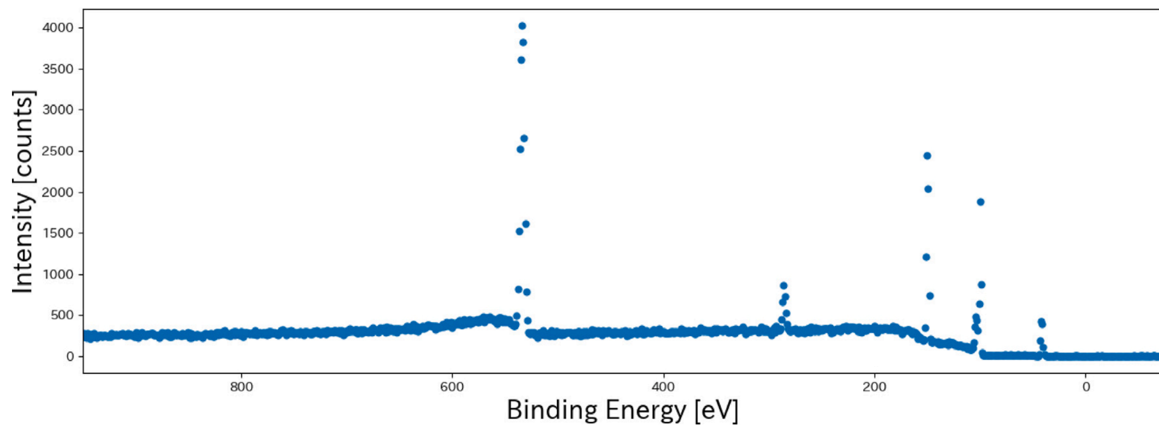
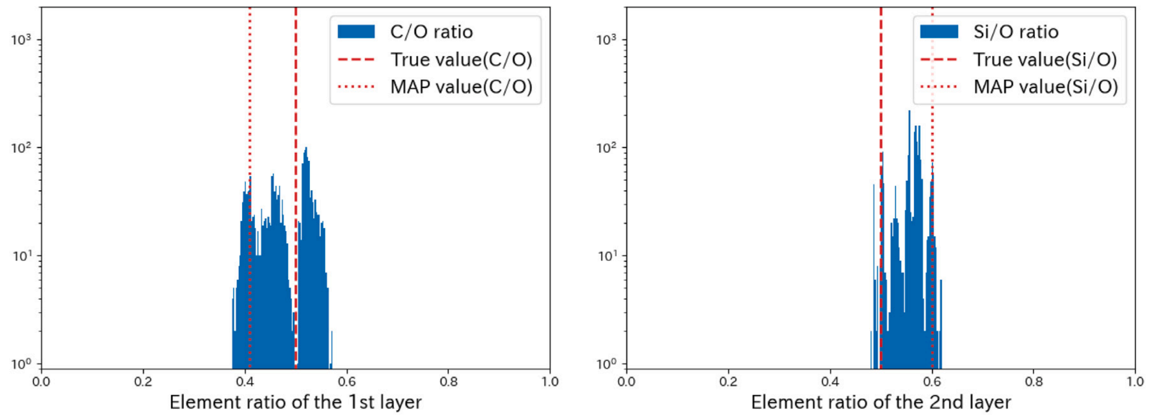
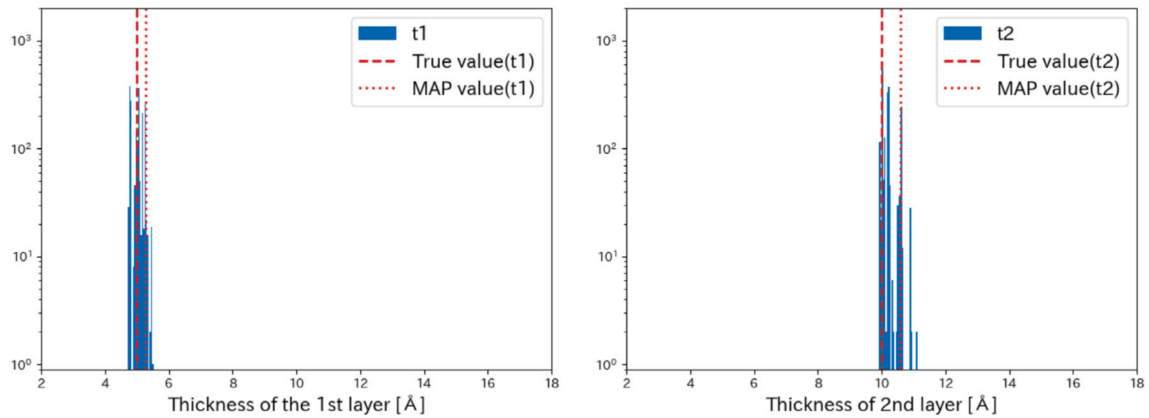


Fig. 3. Wide-scan data generated from the assumed sample.



(a) Posterior distribution of elemantal ratios in the first layer. **(b)** Posterior distribution of elemantal ratios in the second layer.



(c) Posterior distribution of thickness in the first layer. **(d)** Posterior distribution of thickness in the second layer.

Fig. 4. Results of posterior distributions of sample parameters. The vertical axis shows the sampling frequency, and the dotted lines represent the true values and MAP estimates. The vertical axis is on a logarithmic scale.

The prior distributions for each parameter were set as follows:

$$p(\boldsymbol{\Theta}) = p(T_1)p(T_2)p(x_{1,\mathrm{C}}, x_{1,\mathrm{O}})p(x_{2,\mathrm{Si}}, x_{2,\mathrm{O}})\prod_{k=1}^{3} p(h_k)p(l_k)p(r_k), \quad (29)$$

$$p(T_1) = Gamma(T_1; n_T, \mu_T), \quad (30)$$

$$p(T_2) = Gamma(T_2; n_T, \mu_T), \quad (31)$$

$$p(x_{1,\mathrm{C}}, x_{1,\mathrm{O}}) = Dirichlet(x_{1,\mathrm{C}}, x_{1,\mathrm{O}}; \alpha), \quad (32)$$

$$p(x_{2,\mathrm{Si}}, x_{2,\mathrm{O}}) = Dirichlet(x_{2,\mathrm{Si}}, x_{2,\mathrm{O}}; \alpha), \quad (33)$$

$$p(h_k) = Gamma(h_k; n_k, \mu_{h_k}) \quad \text{for } k = \text{O1s,C1s,Si2p}, \quad (34)$$

$$p(l_k) = Gamma(l_k; n_k, \mu_{l_k}) \quad \text{for } k = \text{O1s,C1s,Si2p}, \quad (35)$$

$$p(r_k) = Gamma(r_k; n_k, \mu_{r_k}) \quad \text{for } k = \text{O1s,C1s,Si2p}, \quad (36)$$

$$n_{\mathrm{O1s}} = n_{\mathrm{C1s}} = n_{\mathrm{Si2p}} = 11, \quad (37)$$

$$\mu_{h_{\mathrm{O1s}}} = 1.0 \times 10^8, \quad (38)$$

$$\mu_{h_{\mathrm{C1s}}} = 5.0 \times 10^8, \quad (39)$$

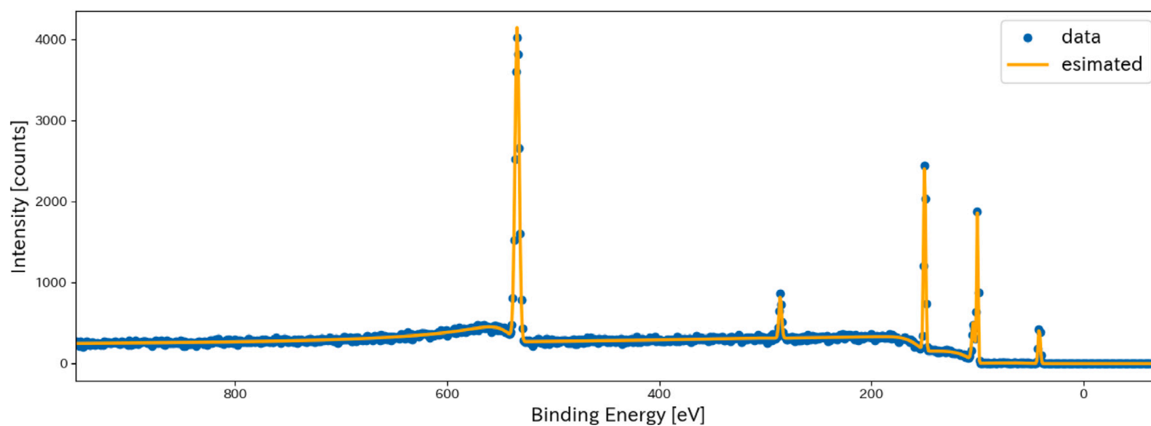
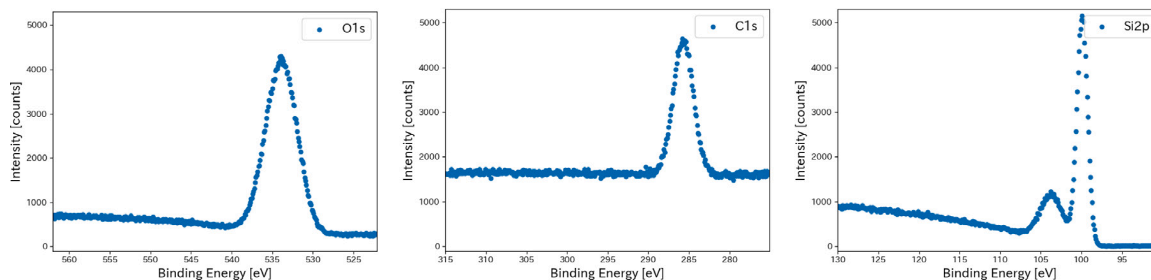


Fig. 5. Fitting results obtained using MAP estimates.



(a) Artificial data derived from O1s. **(b)** Artificial data derived from C1s. **(c)** Artificial data derived from Si2p.

Fig. 6. Three sets of narrow-scan data generated from the assumed sample.

$$\mu_{h_{\mathrm{Si2p}}} = 2.5 \times 10^8, \tag{40}$$

$$\mu_{l_{\mathrm{O1s}}} = \mu_{r_{\mathrm{O1s}}} = 27, \tag{41}$$

$$\mu_{l_{\mathrm{C1s}}} = \mu_{r_{\mathrm{C1s}}} = 1.6 \times 10^2, \tag{42}$$

$$\mu_{l_{\mathrm{Si2p}}} = \mu_{r_{\mathrm{Si2p}}} = 1.0. \tag{43}$$

3.2.2. Results

Fig. 7 shows the posterior distributions and estimated values of each sample parameter obtained by Bayesian estimation.

Fig. 7 also shows that the sampling concentrated around the true values for both elemantal ratio and thickness. The MAP estimates do not always agree with the true values, but this can be attributed to the effect of noise in the data. However, as in Section 3.1, the elemental ratios of the CO_2 layer have wider distributions than the other parameters, and the accuracy is low. We will discuss the reasons for this.

From the setup in this experiment, the main component contributing to the narrow-scan data for the C1s orbital is only the C element in the CO_2 layer. However, the data intensity in Fig. 6(b) is dominated by the matrix background, and the intensity from the C1s orbital is low. This makes it difficult to estimate the elemental ratio of the C elements. Below are the results of fitting the data using the MAP estimates (see Fig. 8). It can be seen that the fitting can be performed correctly on the data.

3.3. Discussion

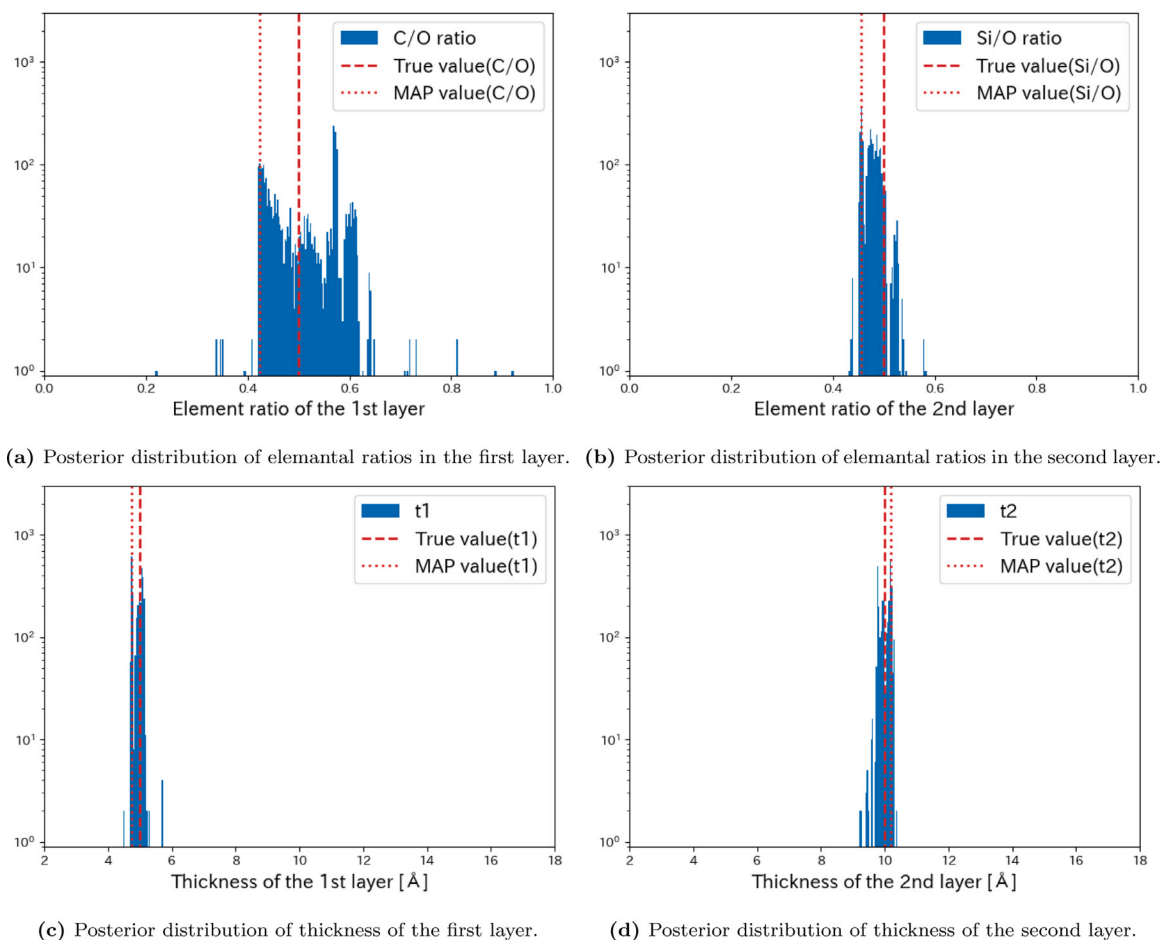
To discuss the results in Sections 3.1.2 and 3.2.2 quantitatively, Bayesian credible interval (BCI) [12,13] results for each experiment are presented in Tables 2 and 3. BCI is the interval of the posterior in which a parameter is included under a certain probability and is used in Bayesian statistics to indicate parameter uncertainty. In this study, the 95% BCI was calculated from the sampling obtained.

Table 2
True values, MAP estimates and BCI in estimation from wide scan data.

parameter	True value	MAP value	95% BCI
C/O ratio (1st layer)	0.500	0.410	[0.394 0.554]
Si/O ratio (2nd layer)	0.500	0.601	[0.489 0.604]
Thickness (1st layer) [Å]	5.0	5.3	[4.8 5.3]
Thickness (2nd layer) [Å]	10.0	10.6	[10.0 10.6]

From the results in Sections 3.1.2 and 3.2.2, it appears that the sampling is distributed around the true values and the fitting with the MAP estimates is successful. Similarly, it can be seen from Tables 2 and 3 that the 95% BCI for all parameters in both results also contain true values. From the results shown in Section 3.2.2, it was found that for the present three-layer model, information on composition and film thickness in the depth direction can be obtained sufficiently by analyzing narrow-scan spectra up to 30 eV away from the peak without angle-resolved measurement. Also these results indicates that wide-scan data can be sufficiently used for the quantitative analysis by this method. Moreover, quantitative analysis of wide-scan data can contribute to the efficiency of the comprehensive analysis of XPS data, because such analysis can lead to the immediate detection of unexpected situations in actual samples (e.g., the presence of impurity elements).

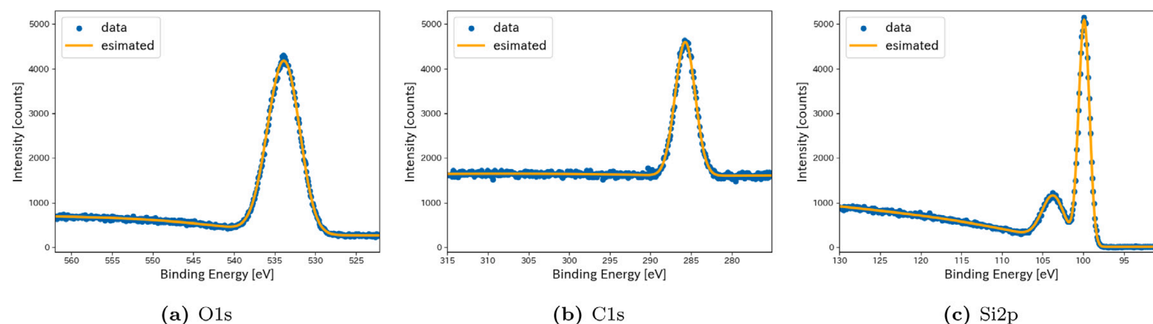
However, we believe that there is room for improvement with regard to the accuracy of the estimation, which can be attributed to the small number of Monte Carlo steps. In this study, the number of steps was set to 6000 for the sake of computation time, but this may not have been sufficient to converge to the posterior distribution. The estimation accuracy can also be improved by increasing the intensity of the data used.



(a) Posterior distribution of elemantal ratios in the first layer. **(b)** Posterior distribution of elemantal ratios in the second layer.

(c) Posterior distribution of thickness of the first layer. **(d)** Posterior distribution of thickness of the second layer.

Fig. 7. Results of posterior distributions of sample parameters. The vertical axis shows the sampling frequency, and the dotted lines represent the true values and MAP estimates. The vertical axis is on a logarithmic scale.



(a) O1s **(b)** C1s **(c)** Si2p

Fig. 8. Fitting results obtained using MAP estimates.

Table 3
True values, MAP estimates and BCI in estimation from narrow scan data.

parameter	True value	MAP value	95% BCI
C/O ratio (1st layer)	0.500	0.424	[0.422 0.613]
Si/O ratio (2nd layer)	0.500	0.456	[0.454 0.522]
Thickness (1st layer) [Å]	5.0	4.7	[4.7 5.1]
Thickness (2nd layer) [Å]	10.0	10.2	[9.8 10.3]

4. Conclusion

In this paper, we proposed a Bayesian estimation method utilizing the simulator SESSA for XPS data analysis. In conventional XPS data analysis, it has been considered difficult to obtain information on the layer structure using only narrow-scan data, and analysis in combination with wide-scan data analysis has been necessary. It was also believed that wide-scan data alone could only provide qualitative findings and not quantitative layer structure information. We used Bayesian inference to estimate the layer structure from artificial data for these two situations. The analysis of two sets of synthetic data in this study show that the proposed method can be used to estimate the layer structure information independently from both wide-scan and narrow-scan data. However, we believe that there is room for improvement in terms of estimation accuracy, which could be improved by increasing the intensity of the data or by using a sufficiently large number of Monte Carlo steps. This achievement is significant for the automated analysis of XPS data, and future prospects include the extension of

this method and its application to real data. However, for computational convenience, in this paper, we made some assumptions about the sample in the estimation; thus, the estimation can be considered successful. It is necessary to confirm the validity of various sample conditions by changing the assumptions to different sample conditions and by conducting experiments. In addition, it is necessary to solve the heavy computation time of SESSA in order to conduct more general experiments. These issues will be addressed in the future.

CRediT authorship contribution statement

Atsushi Machida: Writing – original draft. **Kenji Nagata:** Project administration. **Ryo Murakami:** Writing – review & editing. **Hiroshi Shinotsuka:** Writing – review & editing. **Hayaru Shouno:** Project administration. **Hideki Yoshikawa:** Project administration. **Masato Okada:** Project administration.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP23KJ0471 and JP23H00486, and CREST Grant Number JPMJCR1761 from the Japan Science and Technology Agency (JST).

References

[1] A. Machida, K. Nagata, R. Murakami, H. Shinotsuka, H. Shouno, H. Yoshikawa, M. Okada, Bayesian estimation for XPS spectral analysis at multiple core levels, Sci. Technol. Adv. Mater.: Methods 1 (1) (2021) 123–133, http://dx.doi.org/10.1080/27660400.2021.1943172.

[2] P.J. Cumpson, Angle-resolved XPS and AES: Depth-resolution limits and a general comparison of properties of depth-profile reconstruction methods, J. Electron Spectrosc. Relat. Phenom. 73 (1) (1995) 25–52, http://dx.doi.org/10.1016/0368-2048(94)02270-4.

[3] P.J. Cumpson, Angle-resolved XPS depth-profiling strategies, Appl. Surf. Sci. 144 (1999) 16–20, http://dx.doi.org/10.1016/S0169-4332(98)00752-1.

[4] V. Lockett, R. Sedev, C. Bassell, J. Ralston, Angle-resolved X-ray photoelectron spectroscopy of the surface of imidazolium ionic liquids, Phys. Chem. Chem. Phys. 10 (9) (2008) 1330–1335, http://dx.doi.org/10.1039/B713584J.

[5] S. Tougaard, Practical guide to the use of backgrounds in quantitative XPS, J. Vacuum Sci. Technol. A 39 (1) (2021) http://dx.doi.org/10.1116/6.0000661.

[6] M. Chudzicki, W.S.M. Werner, A.G. Shard, Y.C. Wang, D.G. Castner, C.J. Powell, Evaluating the internal structure of core–shell nanoparticles using X-ray photoelectron intensities and simulated spectra, J. Phys. Chem. C 119 (31) (2015) 17687–17696, http://dx.doi.org/10.1021/acs.jpcc.5b04517.

[7] W. Smekal, W.S. Werner, C.J. Powell, Simulation of electron spectra for surface analysis (SESSA): A novel software tool for quantitative auger-electron spectroscopy and X-ray photoelectron spectroscopy, Surf. Interface Anal.: Int. J. Devoted Develop. Appl. Tech. Anal. Surf. Interfaces Thin Films 37 (11) (2005) 1059–1067, http://dx.doi.org/10.1002/sia.2097.

[8] W. Werner, W. Smekal, C.J. Powell, J. Gorham, Simulation of Electron Spectra for Surface Analysis (SESSA) Version 2.2 User's Guide, 2021, http://dx.doi.org/10.6028/NIST.NSRDS.100-2021.

[9] K. Nagata, S. Sugita, M. Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, Neural Netw. 28 (2012) 82–89, http://dx.doi.org/10.1016/j.neunet.2011.12.001.

[10] R. Murakami, H. Tanaka, H. Shinotsuka, K. Nagata, H. Shouno, H. Yoshikawa, Development of multiple core-level XPS spectra decomposition method based on the Bayesian information criterion, J. Electron Spectrosc. Relat. Phenom. 245 (2020) 147003, http://dx.doi.org/10.1016/j.elspec.2020.147003.

[11] K. Hukushima, K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, J. Phys. Soc. Japan 65 (6) (1996) 1604–1608, http://dx.doi.org/10.1143/JPSJ.65.1604.

[12] L. Hespanhol, C.S. Vallio, L.M. Costa, B.T. Saragiotto, Understanding and interpreting confidence and credible intervals around effect estimates, Braz. J. Phys. Therapy 23 (4) (2019) 290–301, http://dx.doi.org/10.1016/j.bjpt.2018.12.006.

[13] J.A. Bittl, Y. He, Bayesian analysis: A practical approach to interpret clinical trials and create clinical practice guidelines, Circulation: Cardiovasc. Qual. Outcomes 10 (8) (2017) e003563, http://dx.doi.org/10.1161/CIRCOUTCOMES.117.003563.