scientific reports

Check for updates

OPEN

Bayesian active learning with model selection for spectral experiments

Tomohiro Nabika[1], Kenji Nagata[2], Masaichiro Mizumaki[3], Shun Katakami[1] & Masato Okada[1]✉

Active learning is a common approach to improve the efficiency of spectral experiments. Model selection from the candidates and parameter estimation are often required in the analysis of spectral experiments. Therefore, we proposed an active learning with model selection method using multiple parametric models as learning models. Important points for model selection and its parameter estimation were actively measured using Bayesian posterior distribution. The present study demonstrated the effectiveness of our proposed method for spectral deconvolution and Hamiltonian selection in X-ray photoelectron spectroscopy.

Experimental design to reduce the cost of experiments is a fundamental challenge from science to industry and has been extensively studied[1]. A sequential experimental design, which selects the measurement point sequentially, has been realized by active learning[2].

In spectral experiment, two active learning methods have been primarily evaluated. One method is to use a Gaussian process regression (GPR) model as a learning model[3–8]. As this approach is model-agnostic, it can be applied to an experiment without a formulated physical model. However, its application for the parameter estimation of physical models might be a challenge[9]. Another issue is the approach for measurement noise[2].

The other method is to fix a physical model before the experiment and use it as a learning model[10–13]. This approach is suitable for the parameter estimation of physical models, but cannot be applied to the experiment where the physical model is not fixed.

However, in the analysis of experimental data, a physical model is selected from the candidates and then its parameters are estimated. To improve the efficiency of such experiments, active learning with model selection for parametric models is required. Active learning with model selection has been separately studied in various fields such as linear regression[14], labeling problems[15], and kernel selection for GPR[16]. However, none of these is applicable to spectral experiments.

In this study, we propose an active learning with model selection method using multiple parametric models as learning models to improve the model selection and its parameter estimation for spectral experiments. First, the model and its parameter posterior distribution are calculated; then, they are used to select the next measurement for model selection and its parameter estimation. The posterior probabilities are approximated using the exchange Monte Carlo method[17,18], which allows our methods to be applied to complex physical models.

The results of the present study demonstrated the effectiveness of the proposed method for spectral deconvolution and Hamiltonian selection in X-ray photoelectron spectroscopy (XPS). In the numerical experiment, our method improved the accuracy of model selection and its parameter estimation while reducing the experiment time compared with the experiment without active learning or those with active learning using GPR.

Bayesian model selection and its parameter estimation

We consider the problem of selecting the physical model M from the candidates $\mathcal{M} = \{M_1, \ldots, M_K\}$ and estimating its parameter θ_M. Let $D = \{x_i, y_i\}_{i=1}^N$ be the data, where x_i is the measurement point and y_i is the observed value. If the model M and its parameter θ_M are given, the probability of the data D is given by

[1]Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8561, Japan. [2]Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba, Ibaraki, 305-0047, Japan. [3]Faculty of Science, Course for Physical Sciences, Kumamoto University, Kumamoto, Japan. ✉email: okada@edu.k.u-tokyo.ac.jp

$$p(D|M,\theta_M) = \prod_{i=1}^{N} p(y_i|x_i,M,\theta_M), \tag{1}$$

where the observed value y_i is assumed to be independently generated.

From Bayes' theorem, the posterior probability of model M and its parameter θ_M is given by

$$p(M|D) = \frac{\int p(D|\theta_M,M)p(\theta_M)p(M)d\theta_M}{\sum_{M\in\mathcal{M}}\int p(D|\theta_M,M)p(\theta_M)p(M)d\theta_M}, \tag{2}$$

$$p(\theta_M|D,M) = \frac{p(D|\theta_M,M)p(\theta_M)}{\int p(D|\theta_M,M)p(\theta_M)d\theta_M}, \tag{3}$$

where $p(M)$ and $p(\theta_M)$ are the prior probabilities of model M and its parameter θ_M, respectively. The numerical computation of these posterior distribution can be realized by the exchange Monte Carlo method[17,18].

Bayesian active learning with model selection for parametric models

The objective of the active learning is to maximize the estimation accuracy of model M and its parameter θ_M by sequentially selecting the next measurement point. In this study, we propose an active learning method to select the next measurement point based on two criteria: the expected improvement of the parameter estimation and that of the model selection (Fig. 1). The detailed equation transformations are given in the supplementary materials.

Active learning criterion for parameter estimation

When $\{x, y\}$ is added to the data D, the information gain of the posterior distribution of the parameter θ_M is represented by

$$\mathcal{J}_M(x;y) = H(p(\theta_M|D,M)) - H(p(\theta_M|D\cup\{x,y\},M)). \tag{4}$$

where $H(p)$ is the entropy of $p(\cdot)$. Therefore, the expected gain provided by x is

$$\mathcal{I}_M(x) = \int \mathcal{J}_M(x;y)p(y|x,D,M)dy \tag{5}$$

$$= \int_\Theta KL(p_{x,\theta_M}||p_{x,D})p_D(\theta_M)d\theta_M, \tag{6}$$

where $p_{x,\theta}(y) = p(y|x,\theta_M,M), p_D(\theta_M) = p(\theta_M|D,M), p_{x,D}(y) = p(y|x,D,M) = \int p(y|x,\theta_M,M)p_D(\theta_M)d\theta$, and $KL(p||q)$ is the Kullback–Leibler (KL) divergence between p and q[19]. From convexity of KL divergence, $\mathcal{I}_M(x)$ is bounded as follows:

$$\mathcal{I}_M(x) \le \int\int KL(p_{x,\theta_M}||p_{x,\theta'_M})p_D(\theta_M)p_D(\theta'_M)d\theta'_M d\theta_M \tag{7}$$

$$= \widetilde{\mathcal{I}}_M(x) \tag{8}$$

When the model M is expressed as

$$p(y|x,\theta_M,M) = Poisson(y; f_M(x;\theta_M)) \tag{9}$$

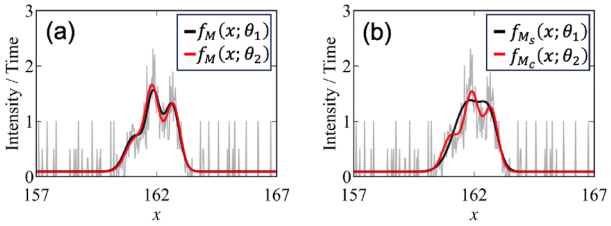

Figure 1. Criteria for active learning. (**a**) Parameter estimation. The gray line represents data D and θ_1, θ_2 follow $p(\theta_M|D,M)$. $\widetilde{\mathcal{I}}_M(x)$ corresponds to the difference between $f_M(x;\theta_1)$ and $f_M(x;\theta_2)$ integrated numerically over $p(\theta_M|D,M)$. (**b**) Model selection. The gray line represents data D and θ_1, θ_2 follow $p(\theta_{M_s}|D,M_s), p(\theta_{M_c}|D,M_c)$, respectively. $\widetilde{\mathcal{I}}_{s,c}(x)$ corresponds to the difference between $f_{M_s}(x;\theta_1)$ and $f_{M_c}(x;\theta_2)$ integrated numerically over $p(\theta_{M_s}|D,M_s)$ and $p(\theta_{M_c}|D,M_c)$.

$$= \frac{f_M(x;\theta_M)^y \exp(-f_M(x;\theta_M))}{y!}, \tag{10}$$

KL divergence of p_{x,θ_M} and p_{x,θ'_M} is calculated as follows:

$$KL(p_{x,\theta_M}||p_{x,\theta'_M}) = f_M(x;\theta'_M) - f_M(x;\theta_M) + f_M(x;\theta'_M)\log\frac{f_M(x;\theta_M)}{f_M(x;\theta'_M)}, \tag{11}$$

where $f_M(x;\theta_M)$ is a physical model, and y follows Poisson distribution. By setting $M = \widehat{M} = argmaxp(M|D)$, $\widetilde{\mathcal{I}}_M(x)$ can be calculated numerically with $p_D(\theta_M)$, which is obtained by the exchange Monte Carlo method.

Therefore, we consider selecting the next measurement point x that maximizes $\widetilde{\mathcal{I}}_M(x)$.

Active learning criterion for model selection

The aforementioned criterion improves the accuracy of parameter estimation when $\widehat{M}$ is a true model. Here, we consider the criterion to make $\widehat{M}$ a true model. When data is small, a higher signal-to-noise ratio can make complex structures in spectral data less discernible, leading to a higher likelihood of selecting simpler models[20]. Therefore, we consider the criterion to select samples that favors the more complex model.

Let the second-best model $M' = argmax_{M\neq\widehat{M}}p(M|D)$, two competitive models $\{M_s, M_c\} = \{\widehat{M}, M'\}$, and M_s have a smaller parameter dimension than M_c. (Specifically, if $\widehat{M}$ is simpler than M', $M_s = \widehat{M}$ and $M_c = M'$; otherwise, $M_s = M'$ and $M_c = \widehat{M}$). We consider the following criterion to make $p(M_c|D\cup\{x,y\})$ bigger than $p(M_s|D\cup\{x,y\})$:

$$\mathcal{I}_{s,c}(x) = \int \log\frac{p(M_c|D\cup\{x,y\})}{p(M_s|D\cup\{x,y\})}p(y|x,D)dy \tag{12}$$

$$= \int \log\frac{p(y|x,D,M_c)}{p(y|x,D,M_s)}p(y|x,D,M_c)dy + C. \tag{13}$$

where C is a constant independent of x.

$\frac{p(y|x,D,M_c)}{p(y|x,D,M_s)}$ is referred to as the Bayes factor, a concept well-explored in Bayesian decision theory[21,22].

From convexity of KL divergence, $\mathcal{I}_{s,c}(x)$ is bounded as follows:

$$\begin{aligned}&\mathcal{I}_{s,c}(x) - C\\ &\leq \int\int KL(p_{x,\theta_{M_c}}||p_{x,\theta_{M_s}})p_D(\theta_{M_c})p_D(\theta_{M_s})d\theta_{M_c}d\theta_{M_s}\end{aligned} \tag{14}$$

$$= \widetilde{\mathcal{I}}_{s,c}(x) \tag{15}$$

$\widetilde{\mathcal{I}}_{s,c}(x)$ can be calculated with $p_D(\theta_{M_c}), p_D(\theta_{M_s})$, which are obtained by the exchange Monte Carlo method.

Therefore, the next measurement point x that maximizes $\widetilde{\mathcal{I}}_{s,c}(x)$ is also selected.

Spectral deconvolution

Our proposed method was applied to the spectral deconvolution in XPS, which poses a challenge in estimating the number of peaks and their parameters[20].

Problem setting

Let M_K be a model with K peaks, the parameter set θ_{M_K} be $\theta_{M_K} = \{\{a_k,\mu_k,\sigma_k\}_{k=1}^K, B\}$, and the physical model $f_{M_K}(x;\theta_K)$ be $f_{M_K}(x;\theta_K) = \sum_{k=1}^K a_k \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right) + B$ (where a_k,μ_k,σ_k, and B correspond to the peak intensity, peak position, peak width, and background intensity, respectively). Since the measurement is performed by photon counting in XPS, the probability distribution of the number of observed photons $p(y|f_M(x;\theta_M))$ is $Poisson(y; f_M(x;\theta_M)\times T)$ with measurement time T.

Detailed algorithm for spectral deconvolution

To apply our method, a set of candidate models must be given in advance; however, in the Bayesian spectral deconvolution, the number of peaks K can take any integer. Therefore, we consider changing the candidate model set sequentially.

We define the initial model set as $\mathcal{M} = \{M_1, M_2, M_3\}$. At each step, let $\widehat{K}$ be the number of peaks of the best predicted model $\widehat{M}$ ($\widehat{M} = M_{\widehat{K}}$). The following model set was used in the next estimation:

$$\mathcal{M} = \begin{cases}\{M_1, M_2, M_3\} & (\widehat{K} = 1)\\ \{M_{\widehat{K}-1}, M_{\widehat{K}}, M_{\widehat{K}+1}\}. & (otherwise)\end{cases} \tag{16}$$

In addition, in the spectral measurement, a short time measurement is performed first, followed by a long time measurement. The specific algorithm that takes these considerations into account is shown in Algorithm 1.

Require: Number of measurement points per one experiment n, Number of experiments k, Measurement points set $\mathcal{X} = \{x_i\}_{i=1}^{N}$
Ensure: Data $D = \{(x_i, y_i)\}_{i=1}^{N+n\times k}$
Measure $y_1, ..., y_N$ with $x_1, ..., x_N$.
Data $D = \{(x_i, y_i)\}_{i\in\{1,...,N\}}$
Candidate model set $\mathcal{M} = \{M_1, M_2, M_3\}$
for $i \in \{1, ..., k\}$ **do**
  $\widehat{M} = M_{\widehat{K}} = \mathrm{argmax}_{M_K\in\mathcal{M}} p(M_K|D)$
  M_s, M_c is the best and second best model in $\mathcal{M}$. (M_c is more complex than M_s)
  Calculate criteria $\{\widetilde{\mathcal{I}}_{\widehat{M}}(x_i)\}_{x_i\in\mathcal{X}}, \{\widetilde{\mathcal{I}}_{s,c}(x_i)\}_{x_i\in\mathcal{X}}$.
  Select $\frac{n}{2}$ points $\{x'_1, ..., x'_{\frac{n}{2}}\}$ in descending order of $\{\widetilde{\mathcal{I}}_{\widehat{M}}(x_i)\}_{x_i\in\mathcal{X}}$.
  Select $\frac{n}{2}$ points $\{x'_{\frac{n}{2}+1}, ..., x'_n\}$ in descending order of $\{\widetilde{\mathcal{I}}_{s,c}(x_i)\}_{x_i\in\mathcal{X}}$.
  Measure $\{y'_1, ..., y'_n\}$ in $\{x'_1, ..., x'_n\}$.
  $D = D \cup \{(x'_1, y'_1), ..., (x'_n, y'_n)\}$
  if $\widehat{K} = 1$ **then**
    $\mathcal{M} = \{M_1, M_2, M_3\}$
  else
    $\mathcal{M} = \{M_{\widehat{K}-1}, M_{\widehat{K}}, M_{\widehat{K}+1}\}$
  end if
end for

Algorithm 1. Sequential experiment for Bayesian spectral deconvolution.

Conventional methods

We compare our method with the following two conventional methods.

Passive learning

Passive learning measures the same measurement time at all measurement points. This method is the most common method in spectral experiments.

Active learning with GPR

In active learning with GPR, a GPR model is used as a learning model, and the next measurement point that maximizes the expected improvement of the measured value estimation is selected. A detailed algorithm is given in the supplementary materials.

Result

Let the true model be the model M_3 with $K = 3$ peaks, and the true values of the parameters $\theta^*_{M_3} = \{\{a^*_k, \mu^*_k, \sigma^*_k\}_{k=1}^3, B^*\}$ be as follows:

$$\begin{pmatrix} a^*_1 \\ a^*_2 \\ a^*_3 \end{pmatrix} = \begin{pmatrix} 0.587 \\ 1.522 \\ 1.183 \end{pmatrix}, \quad \begin{pmatrix} \mu^*_1 \\ \mu^*_2 \\ \mu^*_3 \end{pmatrix} = \begin{pmatrix} 161.032 \\ 161.852 \\ 162.677 \end{pmatrix}, \tag{17}$$

$$\begin{pmatrix} \sigma^*_1 \\ \sigma^*_2 \\ \sigma^*_3 \end{pmatrix} = \begin{pmatrix} 0.341 \\ 0.275 \\ 0.260 \end{pmatrix}, \ B = 0.1. \tag{18}$$

The modeling function $f_{M_3}(x; \theta^*_{M_3})$ is shown in Fig. 2.

Let the measurement time for one measurement in active learning T be $T = 1$, number of measurement points per one experiment $n = 10$, and the candidate set of measurement points $\mathcal{X}$ be $\mathcal{X} = \{157 + 0.025(i-1)\ (eV)\}_{i=1}^{400}$. The prior distributions are shown in the supplementary materials.

The flow of the measurement is shown in Fig. 3. The signal-to-noise ratio is poor at all points at first. However, as the experiment progresses, the signal-to-noise ratio near the peaks improves due to focused measurements. The data and the fitting by the MAP estimator $(\hat{\theta}_{M_3} = \{\{\hat{a}_k, \hat{\mu}_k, \hat{\sigma}_k\}_{k=1}^K, \hat{B}\} = \mathrm{argmax}_\theta\, p(\theta|D, M_3))$ when the total measurement time is 2400 are shown in Fig. 4 (the parameter indices are set so that $\mu_1 < \mu_2 < \mu_3$). This figure shows that the proposed method focuses on the measurement points near the peaks that are considered to be important in the spectral deconvolution.

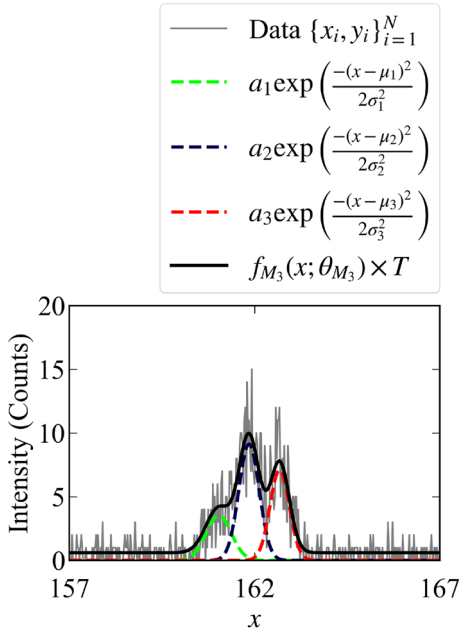

Figure 2. Value of the modeling function $f_{M_3}(x;\theta_{M_3})$ and the example of the observed data $\{x_i, y_i\}_{i=1}^{400}$ when $T = 6$.

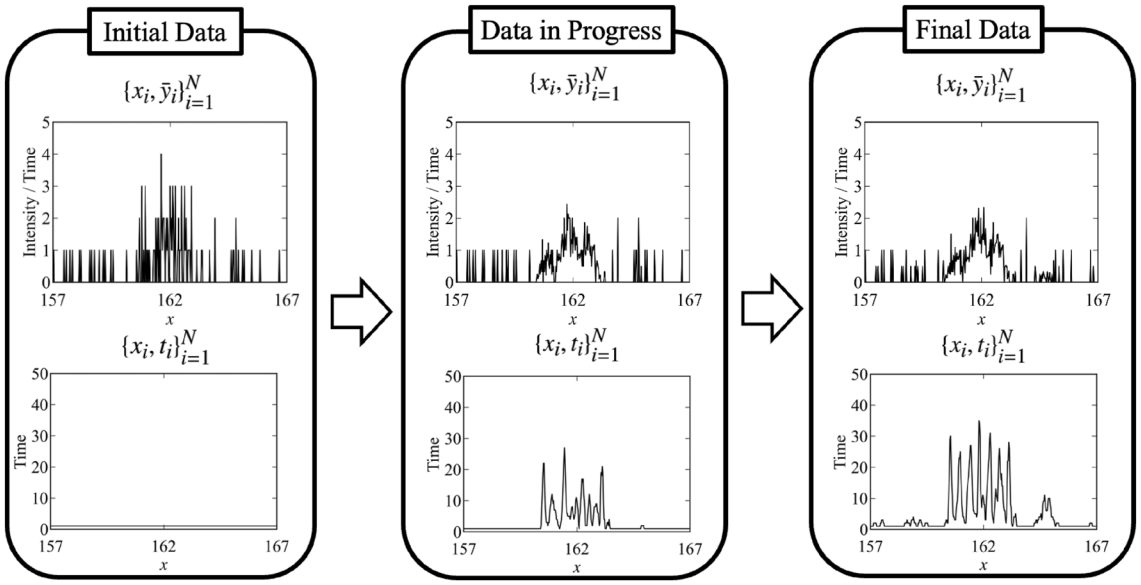

Figure 3. Flow of the proposed method for spectral deconvolution. The upper figure shows the observed values per measurement time $\bar{y}_i = \frac{\sum_{x_j = x_i} y_j}{t_i}$, and the lower figure shows the total measurement time per measurement point $t_i = \#\{j | x_j = x_i\} \times T$. Although the signal-to-noise ratio of the initial data is poor at all measurement points, the signal-to-noise ratio of the data near the peak is improved by repeating the experiments.

In addition, we calculated $p(K = 3|D)$ and $p(\theta_{M_3}|D, M_3)$ when the total measurement time is $\{400 + 100i\}_{i=0}^{36}$. Figure 5A shows the result of the model selection. Active learning with GPR does not improve the model selection because of the high intensity measurement noise. However, our method improves the model selection compared to passive learning. Figure 5B shows the 99% credible interval of the parameter estimation of peak positions μ_1, μ_2, μ_3. Our method narrowed the interval width and improved the parameter estimation.

Moreover, the results of 10 independent measurements were compared: (a) passive learning with total measurement time 2400, (b) active learning with total measurement time 2400, and (c) passive learning with total measurement time 7200. Figure 6 shows the result of the model selection. Figure 7 shows the parameter estimation accuracy. Here, we defined the parameter estimation accuracy $W_{\mu_1}, W_{\mu_2}, W_{\mu_3}$ for μ_1, μ_2, μ_3 as follows:

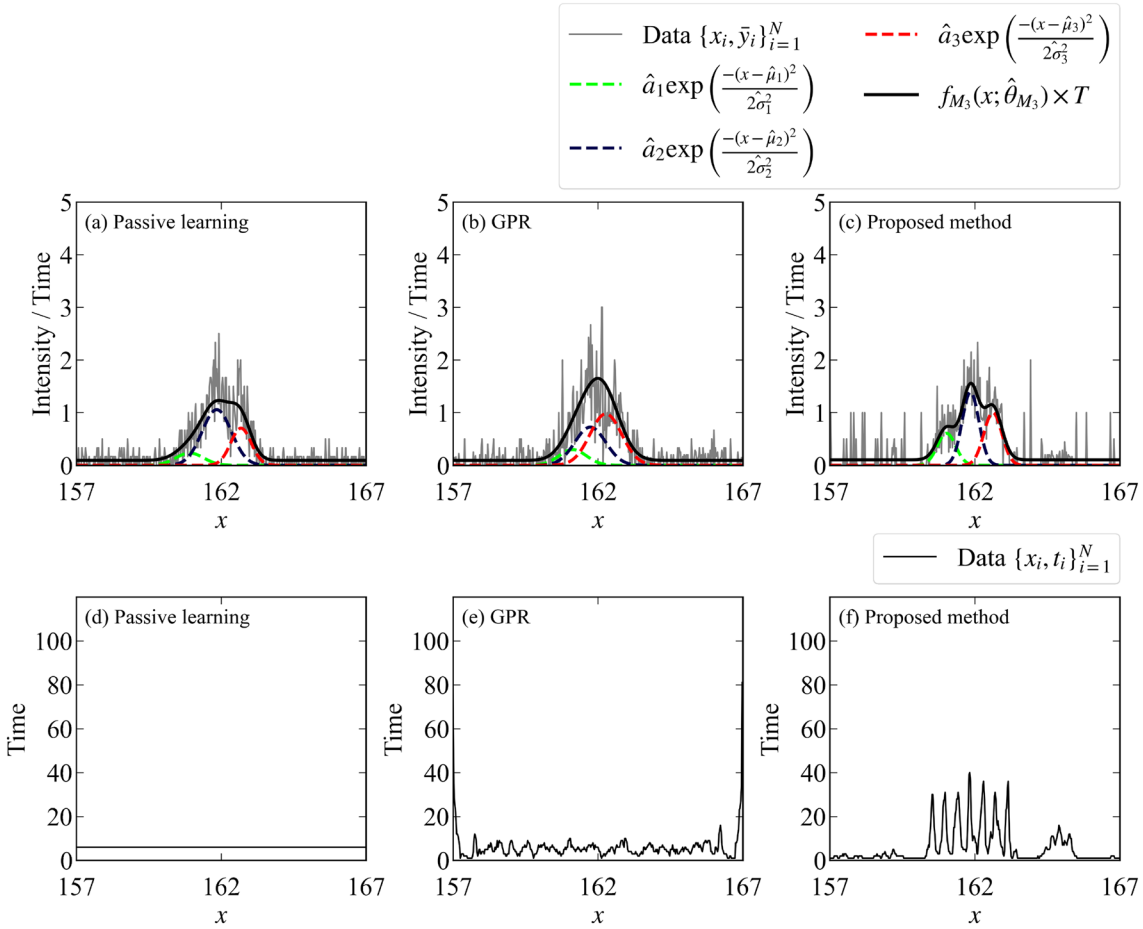

Figure 4. Data and fitting obtained by experiments on the spectral deconvolution. The upper figure shows that the number of photons observed per measurement time $\bar{y}_i$ and the fitting by the MAP estimator. The lower figure shows the total measurement time per measurement point. (**a**,**d**) Passive learning. (**b**,**e**) Active learning with GPR. (**c**,**f**) Proposed method.

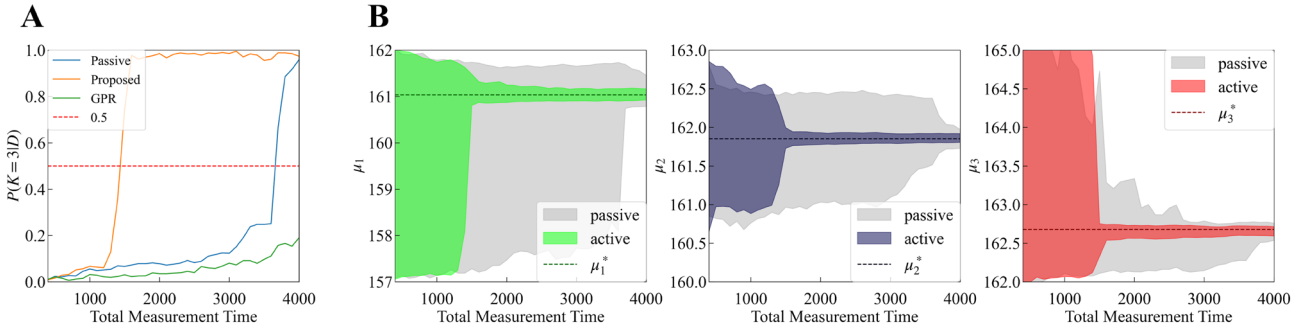

Figure 5. (**A**) Model selection results. The horizontal axis is the total measurement time; the vertical axis, the probability of the true model; the blue line, the result of passive learning; the green line, the result of active learning with GPR; and the orange line, the result of our method. (**B**) 99% credible interval of the parameter estimation of peak positions μ_1, μ_2, μ_3. The horizontal axis is the total measurement time. The gray area represents the result of passive learning; the colored area, the result of our method; and the dotted lines, the true value of μ_1, μ_2, μ_3.

$$W_{\mu_i} = \max_{\alpha \in [0.005, 0.995]} |\mu_i^* - \mu_{i,\alpha}| \tag{19}$$

where

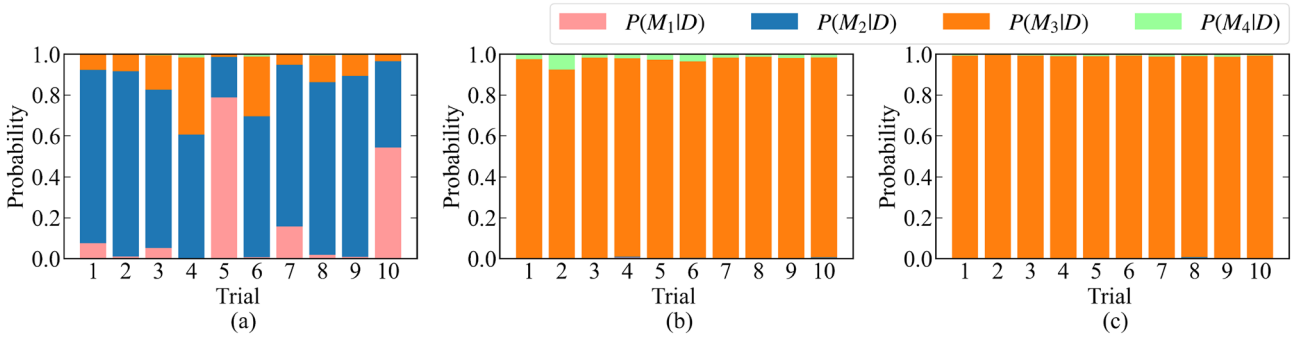

Figure 6. Bar graphs of $p(M_1|D), p(M_2|D), p(M_3|D), p(M_4|D)$ for the 10 independent trials. (**a**) Passive learning with a total measurement time of 2400. (**b**) Proposed method with a total measurement time of 2400. (**c**) Passive learning with a total measurement time of 7200.

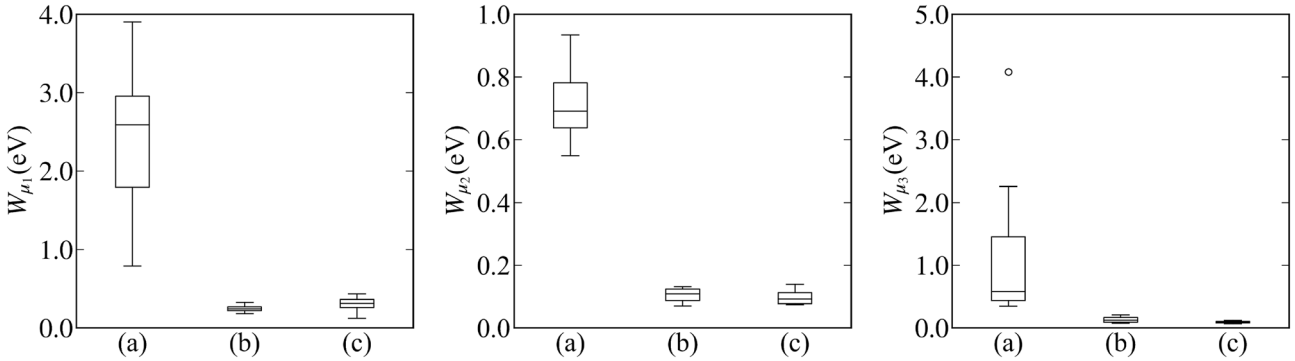

Figure 7. Boxplots represent the accuracy of parameter estimation of the peak positions. The left panel, middle panel, and right panels show the boxplots of W_{μ_1}, W_{μ_2}, and W_{μ_3} respectively. (**a**) Passive learning with a total measurement time of 2400. (**b**) Proposed method with a total measurement time of 2400. (**c**) Passive learning with a total measurement time of 7200.

$$\mu_{i,\alpha} = \min_{\mu} \left\{ \left(\int_{\mu_i < \mu} p(\mu_i|D,K)d\mu_i \right) > \alpha \right\}. \tag{20}$$

Both results show that our method improved the estimation accuracy and shortened the measurement time.

Hamiltonian selection

The Hamiltonian selection in XPS[23] was also considered in this study.

Problem setting

Let M_2 be a model using a two-state Hamiltonian, H_2, and M_3 be a model using a three-state Hamiltonian H_3, and let $\mathcal{M} = \{M_2, M_3\}$ be a set of candidate models. Let $\theta_{M_2} = \{\Delta, V, \Gamma, U_{fc}, b\}$ and $\theta_{M_3} = \{\Delta, V, \Gamma, U_{fc}, U_{ff}, b\}$. The physical model $f_{M_2}(x;\theta_{M_2})$ and $f_{M_3}(x;\theta_{M_3})$ are shown in the supplementary materials. As the measurement is performed by photon counting in XPS, the probability distribution of the number of observed photons $p(y|f_M(x;\theta_M))$ is considered to be $Poisson(y; f_M(x;\theta_M) \times T)$ with measurement time T.

Detailed algorithm for Hamiltonian selection

Unlike in the case of spectral deconvolution, the model set $\mathcal{M} = \{M_2, M_3\}$ is fixed. The specific algorithm is shown in Algorithm 2.

Require: Number of measurement points per one experiment n, Number of experiments k, Measurement points set $\mathcal{X} = \{x_i\}_{i=1}^{N}$
Ensure: Data $D = \{(x_i, y_i)\}_{i=1}^{N+n\times k}$
Measure $y_1, ..., y_N$ with $x_1, ..., x_N$.
Data $D = \{(x_i, y_i)\}_{i\in\{1,...,N\}}$
$\mathcal{M} = \{M_2, M_3\}$
for $i \in \{1, ..., n\}$ **do**
  $\widehat{M} = \mathrm{argmax}_M P(M|D)$, $M_s = M_2, M_c = M_3$.
  Calculate criteria $\{\widetilde{\mathcal{I}}_{\widehat{M}}(x_i)\}_{x_i\in\mathcal{X}}, \{\widetilde{\mathcal{I}}_{s,c}(x_i)\}_{x_i\in\mathcal{X}}$.
  Select $\frac{n}{2}$ points $\{x'_1, ..., x'_{\frac{n}{2}}\}$ in descending order of $\{\widetilde{\mathcal{I}}_{\widehat{M}}(x_i)\}_{x_i\in\mathcal{X}}$.
  Select $\frac{n}{2}$ points $\{x'_{\frac{n}{2}+1}, ..., x'_n\}$ in descending order of $\{\widetilde{\mathcal{I}}_{s,c}(x_i)\}_{x_i\in\mathcal{X}}$.
  Measure $\{y'_1, ..., y'_n\}$ in $\{x'_1, ..., x'_n\}$.
  $D = D \cup \{(x'_1, y'_1), ..., (x'_n, y'_n)\}$
end for

Algorithm 2. Sequential experiment for Bayesian Hamiltonian Selection.

Conventional methods

We compare our method with passive learning and active learning with GPR as in the case of spectral deconvolution.

Result

Let the true model be the model M_3 with H_3 and the true values of its parameters be as follows:

$$\Delta^* = 7.66, V^* = 0.76, U_{ff} = 10.5, \tag{21}$$

$$U_{fc} = 12.7, \ \Gamma = 0.7, \quad b = 0. \tag{22}$$

This true parameter is derived from[23]. The physical function $f_{M_3}(x; \theta_K LdivergenceoM_3)$ with the true parameter $\theta^*_{M_3} = \{\Delta^*, V^*, \Gamma^*, U^*_{fc}, U^*_{ff}, b^*\}$ is shown in Fig. 8. The peak around $x = 5$ is small, indicating that the model selection from model M_2 that generates two peaks and model M_3 that generates three peaks is difficult. Let the measurement time for one measurement in active learning T be $T = 1$, number of measurement points per one experiment $n = 10$, and the candidate set of measurement points $\mathcal{X}$ be $\mathcal{X} = \{-30 + 0.125(i-1)\}_{i=1}^{400}$. The prior distribution is shown in the supplementary materials.

The flow of the measurement is shown in Fig. 9. The signal-to-noise ratio is poor at all points at first. However, as the experiment progresses, the signal-to-noise ratio near the peaks improves due to focused measurements. The data and the fitting by the MAP estimator ($\hat{\theta}_{M_3} = \underset{\theta}{\mathrm{argmax}}\, p(\theta|D, M_3)$) when the total measurement time is

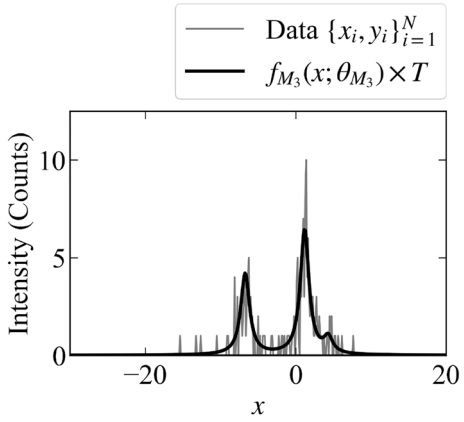

Figure 8. Plot of the modeling function $f_{M_3}(x; \theta^*_{M_3})$ and the example of the observed data $\{x_i, y_i\}_{i=1}^{400}$ when $T = 25$. The peak around $x = 5$ is small, indicating the difficulty of the model selection.

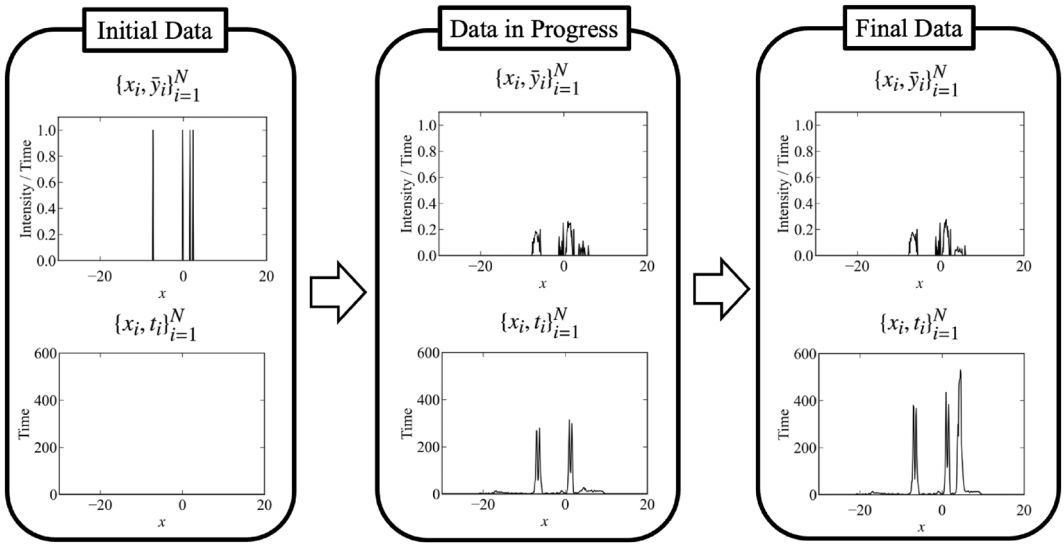

Figure 9. Flow of the proposed method for Hamiltonian selection. The upper figure shows the observed values per measurement time $\bar{y}_i = \frac{\sum_{x_j = x_i} y_j}{t_i}$, and the lower figure shows the total measurement time per measurement point $t_i = \#\{j | x_j = x_i\} \times T$. It can be observed that the area near the peaks, particularly near the small peak around $x = 5$, is measured intensively.

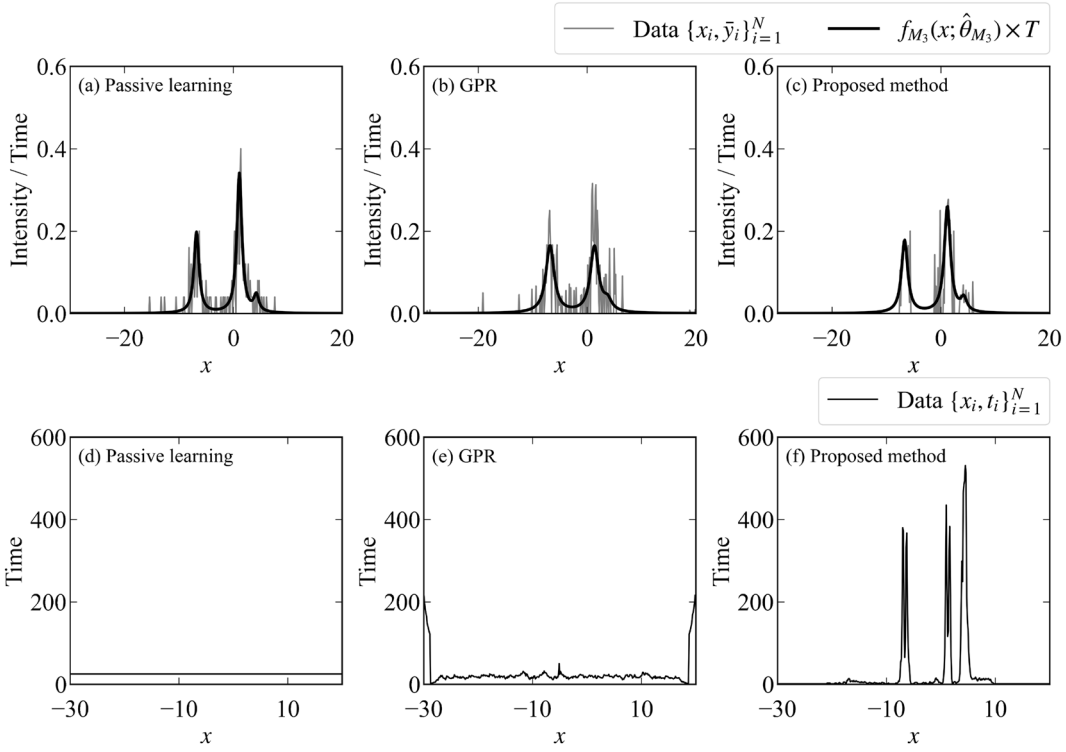

Figure 10. Data and fitting obtained by experiments on the Hamiltonian selection. The upper figure shows that the number of photons observed per measurement time $\bar{y}_i$ and the fitting by the MAP estimator. The lower figure shows the total measurement time per measurement point. (**a**,**d**) Passive learning. (**b**,**e**) Active Learning with GPR. (**c**,**f**) Proposed method.

10,000 are shown in Fig. 10. This figure shows that the proposed method focuses on the area near the peaks, particularly near the small peak around $x = 5$.

In addition, we calculated $p(M_3|D)$ and $p(\theta_{M_3}|D, M_3)$ when the total measurement time is $\{400 + 300i\}_{i=0}^{32}$. Figure 11A shows the result of the model selection. As in the previous section, active learning with GPR did not improve the model selection because of the high intensity measurement noise. However, our method improved the model selection compared to passive learning. Figure 11B shows the 99% credible interval of the parameter estimation of Δ, Γ, U_{fc}. Our method narrowed the interval width and improved the parameter estimation.

Moreover, 10 independent measurements were performed, and the following results were compared: (a) passive learning with total measurement time of 10,000, (b) active learning with total measurement time of 10,000, and (c) passive learning with total measurement time of 40,000. Figure 12 shows the result of the model selection. Figure 13 shows the accuracy of parameter estimation. Here, we defined the accuracy of parameter estimation $W_\Delta, W_\Gamma, W_{U_{fc}}$ for Δ, Γ, U_{fc} as follows:

$$W_\Delta = \max_{\alpha \in [0.005, 0.995]} |\Delta^* - \Delta_\alpha|, \tag{23}$$

$$W_\Gamma = \max_{\alpha \in [0.005, 0.995]} |\Gamma^* - \Gamma_\alpha|, \tag{24}$$

$$W_{U_{fc}} = \max_{\alpha \in [0.005, 0.995]} |U_{fc}^* - U_{fc,\alpha}|, \tag{25}$$

where

$$\Delta_\alpha = \min_{\Delta} \left\{ \left(\int_{\Delta < \Delta^*} p(\Delta|D, M_3) d\Delta \right) > \alpha \right\}, \tag{26}$$

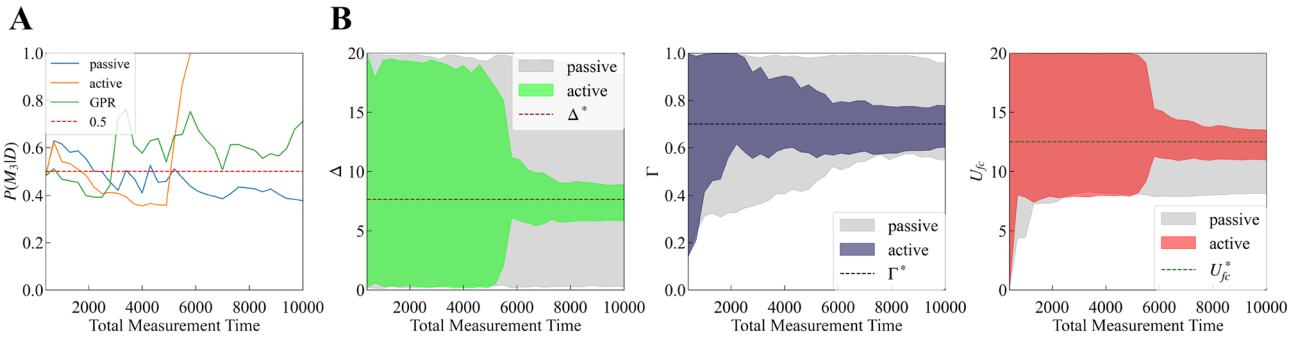

Figure 11. (**A**) Model selection results. The horizontal axis is the total measurement time; the vertical axis is the probability of the true model; the blue line is the result of passive learning; the green line is the result of active learning with GPR; and the orange line is the result of our method. (**B**) 99% credible interval of the parameter estimation of Δ, Γ, U_{fc}. The horizontal axis is the total measurement time. The gray area indicates the result of passive learning, while the colored area indicates the result of our method; the dotted lines represent the true value Δ, Γ, U_{fc}.

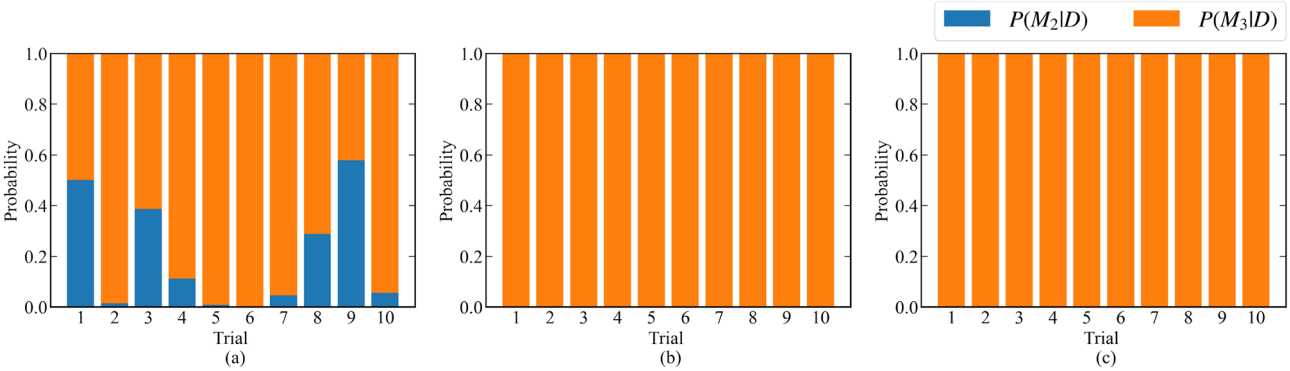

Figure 12. Bar graphs of $p(M_2|D), p(M_3|D)$ for the 10 independent trials. (**a**) Passive learning with a total measurement time of 10,000. (**b**) Proposed method with a total measurement time of 10,000. (**c**) Passive learning with a total measurement time of 40,000.

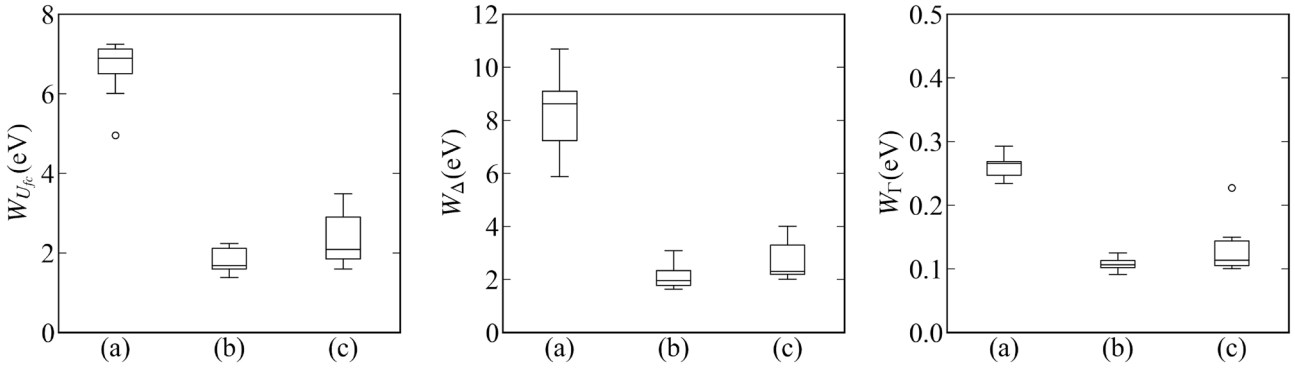

Figure 13. Boxplots represent the accuracy of parameter estimation of Hamiltonian parameters. The left panel, middle panel, and right panels show the boxplots of W_Δ, W_Γ, and $W_{U_{fc}}$, respectively. (**a**) Passive learning with a total measurement time of 10,000. (**b**) Proposed method with a total measurement time of 10,000. (**c**) Passive learning with a total measurement time of 40,000.

$$\Gamma_\alpha = \min_{\Gamma} \left\{ \left(\int_{\Gamma<\Gamma^*} p(\Gamma|D, M_3) d\Gamma \right) > \alpha \right\}, \tag{27}$$

$$U_{fc,\alpha} = \min_{U_{fc}} \left\{ \left(\int_{U_{fc}<U_{fc}^*} p(U_{fc}|D, M_3) dU_{fc} \right) > \alpha \right\}. \tag{28}$$

Both results show that our method improved the estimation accuracy and shortened the measurement time.

Conclusion and future work

We developed an active learning method using multiple parametric models as learning models to improve the accuracy of model selection and parameter estimation in spectral experiments. In our method, the next measurement points that are important for model selection and its parameter estimation were selected using the model and its parameter posterior distribution. We applied our method to two spectral experiments, namely spectral deconvolution and Hamiltonian selection. In both experiments, the proposed method improved the model selection and accuracy of parameter estimation compared with passive learning and active learning with GPR.

To apply our method to a broader range of actual spectral experiments, we need to consider the following points. Firstly, there is a concern about the calculation cost of the proposed method. To reduce the actual experimental time using the proposed method, the computational time of the Monte Carlo method should be sufficiently small compared to the experiment time. Nevertheless, in cases with a large number of parameters, the exploration range of the Monte Carlo method expands, leading to longer convergence times. Additionally, the computation time per iteration often scales proportionally with the number of measurement points. Therefore, in scenarios with a high number of measurement points, such as in the case of high-dimensional spectral data, the measurement time can significantly increase. To mitigate computational time, one approach is to employ the concept of sequential Monte Carlo methods[24], utilizing samples obtained from previous simulations to perform sampling from the new posterior distribution. Moreover, the convergence can be improved by appropriately setting the prior distribution using prior knowledge about the experiment[25].

Another challenge is adjusting the Monte Carlo parameters automatically. The Monte Carlo method has many parameters and setting them up for each experiment is difficult. Thus, an algorithm like NUTS[26] to adjust the Monte Carlo parameters will be required.

Finally, our method is only applicable when candidate models are known in advance. However, background functions often have limited prior knowledge, making modeling challenging in many cases. Such challenges can be solved by using semi-parametric models[27].

Data availability

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

Received: 15 December 2023; Accepted: 11 February 2024
Published online: 14 February 2024

References

1. Rainforth, T., Foster, A., Ivanova, D. R. & Smith, F. B. Modern Bayesian experimental design. arXiv:2302.14545 (arXiv preprint) (2023).
2. Hino, H. Active learning: Problem settings and recent developments. arXiv:2012.04225 (arXiv preprint) (2020).
3. Ueno, T. *et al.* Adaptive design of an X-ray magnetic circular dichroism spectroscopy experiment with Gaussian process modelling. *NPJ Comput. Mater.* **4**, 4 (2018).
4. Ueno, T., Ishibashi, H., Hino, H. & Ono, K. Automated stopping criterion for spectral measurements with active learning. *NPJ Comput. Mater.* **7**, 139 (2021).

5. Noack, M. M. *et al.* A kriging-based approach to autonomous experimentation with applications to X-ray scattering. *Sci. Rep.* **9**, 11809 (2019).
6. Noack, M. M., Doerk, G. S., Li, R., Fukuto, M. & Yager, K. G. Advances in kriging-based autonomous X-ray scattering experiments. *Sci. Rep.* **10**, 1325 (2020).
7. Noack, M. M. *et al.* Gaussian processes for autonomous data acquisition at large-scale synchrotron and neutron facilities. *Nat. Rev. Phys.* **3**, 685–697 (2021).
8. Holman, E. A. *et al.* Autonomous adaptive data acquisition for scanning hyperspectral imaging. *Commun. Biol.* **3**, 684 (2020).
9. Teixeira Parente, M. *et al.* Active learning-assisted neutron spectroscopy with log-Gaussian processes. *Nat. Commun.* **14**, 2246 (2023).
10. Dushenko, S., Ambal, K. & McMichael, R. D. Sequential Bayesian experiment design for optically detected magnetic resonance of nitrogen-vacancy centers. *Phys. Rev. Appl.* **14**, 054036 (2020).
11. McMichael, R. D., Dushenko, S. & Blakley, S. M. Sequential Bayesian experiment design for adaptive Ramsey sequence measurements. *J. Appl. Phys.* **130**, 25 (2021).
12. McMichael, R. D. & Blakley, S. M. Simplified algorithms for adaptive experiment design in parameter estimation. *Phys. Rev. Appl.* **18**, 054001 (2022).
13. Caouette-Mansour, M. *et al.* Robust spin relaxometry with fast adaptive Bayesian estimation. *Phys. Rev. Appl.* **17**, 064031 (2022).
14. Sugiyama, M. & Rubens, N. Active learning with model selection in linear regression. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, 518–529 (2008).
15. Ali, A., Caruana, R. & Kapoor, A. Active learning with model selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28 (2014).
16. Gardner, J. *et al.* Bayesian active model selection with an application to automated audiometry. *Adv. Neural Inf. Process. Syst.* **28**, 25 (2015).
17. Hukushima, K. & Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **65**, 1604–1608 (1996).
18. Nagata, K., Sugita, S. & Okada, M. Bayesian spectral deconvolution with the exchange Monte Carlo method. *Neural Netw.* **28**, 82–89 (2012).
19. Pronzato, L. & Pázman, A. *Design of Experiments in Nonlinear Models: Asymptotic Normality, Optimality Criteria and Small-Sample Properties* (Springer, 2013).
20. Nagata, K., Muraoka, R., Mototake, Y.-I., Sasaki, T. & Okada, M. Bayesian spectral deconvolution based on Poisson distribution: Bayesian measurement and virtual measurement analytics (VMA). *J. Phys. Soc. Jpn.* **88**, 044003 (2019).
21. Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. *Bayesian Data Analysis* (Chapman and Hall, 1995).
22. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **20**, 773–795 (1995).
23. Mototake, Y.-I., Mizumaki, M., Akai, I. & Okada, M. Bayesian Hamiltonian selection in X-ray photoelectron spectroscopy. *J. Phys. Soc. Jpn.* **88**, 034004 (2019).
24. Cappé, O., Godsill, S. J. & Moulines, E. An overview of existing methods and recent advances in sequential Monte Carlo. *Proc. IEEE* **95**, 899–924 (2007).
25. Kashiwamura, S. *et al.* Bayesian spectral deconvolution of X-ray absorption near edge structure discriminating between high-and low-energy domains. *J. Phys. Soc. Jpn.* **91**, 074009 (2022).
26. Hoffman, M. D. *et al.* The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
27. Tokuda, S. *et al.* Unveiling quasiparticle dynamics of topological insulators through Bayesian modelling. *Commun. Phys.* **4**, 170 (2021).

Acknowledgements

This work was supported by JST, CREST (Grant Numbers JPMJCR1761 and JPMJCR1861), Japan, and JSPS KAKENHI Grant-in-Aid for Scientific Research(A) (No. 23H00486).

Author contributions

M.O. decided the research plan, T.N., K.N. and M.M. developed the proposed method, T.N. and S.K. did the numerical experiment, and T.N. wrote the main manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-54329-w.

Correspondence and requests for materials should be addressed to M.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.