doi:10.1017/S1431927618013065
Microsc. Microanal. 24 (Suppl 2), 2018

Optimal Design of Experiment for X-Ray Spectromicroscopy by Machine Learning

Tetsuro Ueno[1,*], Hideitsu Hino[2], and Kanta Ono[3]

[1.] Quantum Beam Science Research Directorate, National Institutes for Quantum and Radiological Science and Technology, Sayo, Japan
[2.] The Institute of Statistical Mathematics, Tachikawa, Japan
[3.] Institute of Materials Structure Science, High Energy Accelerator Research Organization, Tsukuba, Japan
* Corresponding author, ueno.tetsuro@qst.go.jp

The total measurement time of an X-ray spectromicroscopy experiment using a scanning transmission X-ray microscope (STXM) is determined by a multiplication of a number of energy points n_e, sample scanning points n_s, and measurement time per each point t_m plus overhead. Overhead consists of time for data acquisition, moving of sample scanners, beamline optics and undulator properties (gap and phase of magnet arrays). An X-ray spectromicroscopy experiment with an STXM is performed as an image acquisition by sample scanning in an energy-by-energy regime. Moreover, moving of beamline optics such as a grating and mirrors takes longer time than that of piezoelectric actuators for sample scanning. Therefore, it is a good strategy to reduce n_e to reduce total measurement time. Another strategy to reduce total measurement time is an optimization of t_m. One can reduce t_m at the expense of a signal-to-noise (S/N) ratio of spectra, which is proportional to $t_m^{1/2}$. It is important to reduce total measurement time without degrading the quality of spectra to extract physical or chemical parameters by analysis. Machine learning techniques are expected to resolve this issue. Ueno *et al*. proposed the adaptive design of an X-ray magnetic circular dichroism (XMCD) spectroscopy experiment by Gaussian process (GP) modeling, a machine learning technique, and successfully reduced the total number of energy points to measure to evaluate magnetic moments with required accuracy [1].

In the present study, we performed GP modeling of X-ray absorption spectra (XAS) with variable number of energy points and S/N ratios to assess optimal experimental conditions. Gaussian process modeling can predict a spectrum from a limited number of data points, and it is useful to find the optimal condition for X-ray absorption spectroscopy experiment. We used simulated XAS for GP modeling. X-ray absorption spectrum for Sm^{3+} was calculated using CTM4XAS software [2]. Gaussian noise was added to the noiseless spectrum to produce spectra with various S/N ratio which is defined as a ratio between the intensity of the M_5 main peak of Sm^{3+} and standard deviation σ of a Gaussian distribution. Gaussian process modeling was performed by *R* software with the *DiceKriging* package [3].

Figure 1 shows the concept of the present study. First, a dataset of simulated XAS spectra with various S/N ratios and number of data points is prepared. Next, GP modeling is applied to each XAS data and predicted XAS is calculated. Finally, a similarity (Pearson correlation coefficient) between a predicted spectrum and the reference spectrum (XAS with no noise and full data points) is evaluated.

Figure 2 shows the results for GP modeling, and the similarity between GP-predicted spectra and the reference spectrum. As an overall trend, similarity values for high data points density and high S/N ratio are close to 1. Thus, the accuracy of spectrum prediction increases in such conditions. Similarity values do not exceed 0.8 for very low S/N ratio (S/N = 5). Therefore, XAS spectra with poor S/N ratio fail to retrieve the predicted spectrum with high similarity in spite of increasing data points. On the other hand,

high similarity values above 0.9 are observed for S/N ≥ 50 with at least 30 energy points. Similarity values reach almost 1 for S/N ≥ 100 with high data points density ($n_e \geq 100$). That means S/N = 100 is enough in this case. For example, we can expect four-times-faster measurement for same n_e to obtain same Pearson correlation coefficient, if we set S/N = 100 rather than S/N = 200.

In conclusion, we examined GP modeling of XAS with various S/N and n_e to find optimal experimental condition. Necessary and sufficient experimental conditions can be obtained by setting energy point density and S/N ratio depending on expected spectrum structures. Moreover, real-time GP modeling during measurement can automatically find optimal experimental conditions [4].

References:

[1] T Ueno *et al*., npj Comput. Mater. **4** (2018) 4.
[2] E Stavitski and F M F de Groot, Micron **41** (2010) 687.
[3] O Roustant, D Ginsbourger, and Y Deville, J. Stat. Softw. **51** (2012) 1.
[4] HH acknowledges the support from CREST (No. JPMJCR1761) from Japan Science and Technology Agency (JST).

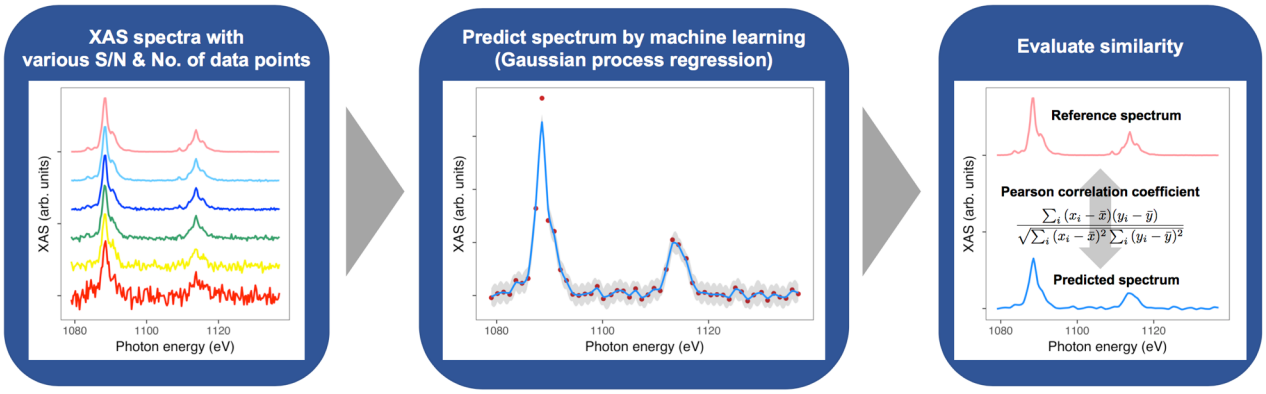

Figure. 1. Schematic of the concept of the present study.

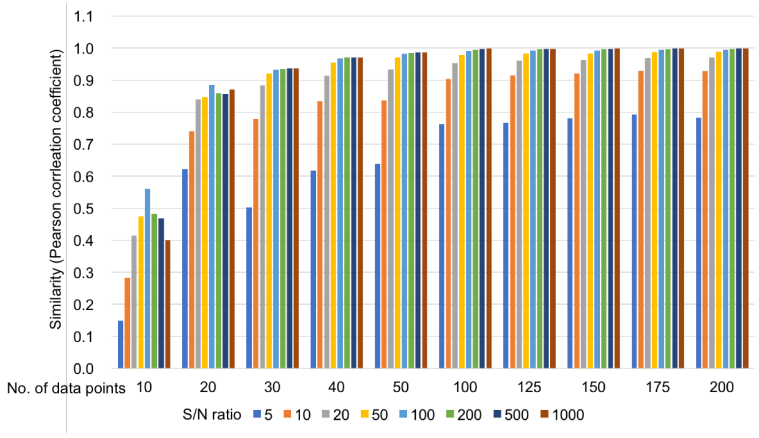

Figure. 2. Similarity values between GP-predicted spectra and the reference spectrum.

https://doi.org/10.1017/S1431927618013065 Published online by Cambridge University Press