

# Early-Stopping of Scattering Pattern Observation with Bayesian Modeling

**Akinori Asahara, Hidekazu Morita**

Hitachi Ltd.  
Tokyo, 100-8280, Japan

**Chiharu Mitsumata**

National Institute for Materials Science  
Tsukuba, 305-0047, Japan

**Kanta Ono**

High Energy Accelerator Research Organization  
Tsukuba, 305-0801, Japan

**Masao Yano, Tetsuya Shoji**

Toyota Motor Corporation  
Toyota, 471-8572, Japan

## Abstract

This paper describes a new machine-learning application to speed up Small-angle neutron scattering (SANS) experiments, and its method based on probabilistic modeling. SANS is one of the scattering experiments to observe microstructures of materials; in it, two-dimensional patterns on a plane (SANS pattern) are obtained as measurements. It takes a long time to obtain accurate experimental results because the SANS pattern is a histogram of detected neutrons. For shortening the measurement time, we propose an early-stopping method based on Gaussian mixture modeling with a prior generated from B-spline regression results. An experiment using actual SANS data was carried out to examine the accuracy of the method. It was confirmed that the accuracy with the proposed method converged 4 minutes after starting the experiment (normal SANS takes about 20 minutes).

## Introduction

Materials informatics (MI) is a field in which information technology is used to accelerate materials science research. Especially data mining techniques will be used to easily find very small features of measurement data. As a such application, we propose an approach to shorten the time for experiments by making predictions based on machine learning. Imagine an experiment that usually takes 10 minutes, where half of the results are obtained in 5 minutes. If the full results of the experiment can be predicted with the data obtained in 5 minutes, the experiment can be finished immediately; namely, the experiment can be made twice as fast. Moreover, if the prediction requires only 1 minute, it will be ten times faster. Generally speaking, measurements on materials requires a special facility and cost a lot – if it is based on high-energy physics, the cost is comparable with hundreds of millions of dollars. Accordingly, such precious experiment time have to be efficiently used.

This study focuses on small-angle neutron scattering (SANS) experiments (Higgins and Benoît 1994). SANS is a scattering experiment that is a popular method for observing the microstructures of materials. There are similar various scattering experiments such as x-ray scattering, ion-beam scattering, etc. Their difference lies just in the particles to be scattered. The solution for the problem in SANS can be

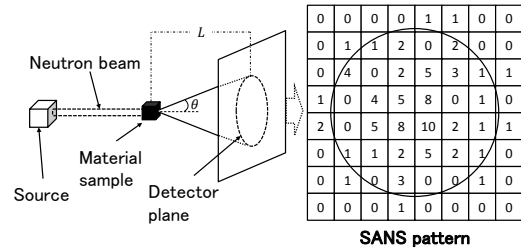


Figure 1: SANS Experiment

expected to apply for these experiments also. Thus the problem is crucial enough to be solved.

We accordingly propose a method to predict a SANS pattern in a short time. The method is based on the Gaussian mixture model (GMM). The GMM can be fitted to the SANS pattern through the variational Bayesian (VB) approach. In particular, our method can estimate the accuracy of the prediction, and thereby, the SANS experiment can be stopped once the prediction is deemed accurate enough. If it stops before convergence of the SANS pattern, the experiment will be faster.

## Problem Setting

### Overview of SANS

Figure 1 shows an illustration of the experimental instruments. A neutron beam incident upon sample interacts with the microstructures therein. The directions of the neutrons thus are changed due to the interactions. The angle  $\theta$  between a straight beam and the changed direction of the scattered beam depends on the interaction.

The neutrons detectors are arranged on a plane. When the distance  $L$  between the sample and the plane is large enough, the coordinate values on the plane  $\mathbf{x} = (x, y)$  are approximately in proportion to  $L \sin \theta \simeq L\theta$ . The probability density function (PDF)  $P(\mathbf{x})$  of neutron detection corresponds to the probability  $P(\theta)$  that neutron goes in the direction of  $\theta$ , which is related to the microscopic structures.

Hence,  $P(\mathbf{x})$  is crucial for understanding the microstructure of the material. When  $n$  detection events denoted as  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  are obtained, the “true” PDF  $P(\mathbf{x})$  is to be estimated from the event data. Note that the event coor-

ordinates are discrete, depending on the density of the detectors. Therefore, the event data are usually handled as event counts in a certain time duration  $t$  for each detector (denoted as  $\{v_{i,j}(t)\}$ , where  $(i, j)$  is discretized  $\mathbf{x}$ ), as shown by the “SANS pattern” in the figure. Furthermore,  $\{v_{i,j}(t)\}$  is considered to be a histogram. Hence,  $v_{i,j}(t)/\sum v_{i,j}(t)$  will converge to the true probability  $P_{i,j}$  that is derived by taking the integral of  $P(\mathbf{x})$  over the cell  $(i, j)$ . Accordingly, the true PDF  $P(\mathbf{x})$  is estimated from the histogram  $\{v_{i,j}\}$  after a sufficiently long time.

### Fast SANS with Prediction

The prediction-based early-stopping for SANS (hereafter, called fast SANS) is proposed in this paper. For the early-stopping, the SANS pattern  $\{v_{i,j}(t)\}$  is periodically posted to a computer to predict  $P_{i,j} = \lim_{t \rightarrow \infty} v_{i,j}(t)/\sum v_{i,j}(t)$  with  $\{v_{i,j}(t)\}$  and tests whether the prediction result has converged. If it has converged, the computer sends a termination signal to the experimental instrument. Otherwise, the computer waits for the next SANS pattern. Note that the prediction and convergence test have to be fast to terminate the experiment earlier. If it takes more time than the conventional SANS, the fast SANS cannot be said to be not fast. For example, if the conventional SANS takes 20 minutes, the time for the process should be less than 20 minutes. Even if the time for the process is 10 minutes, there is just one chance to send the termination signal.

According to the discussion, the problem to be solved is that of predicting the true PDF obtained as  $P_{i,j} = \lim_{t \rightarrow \infty} v_{i,j}(t)/\sum v_{i,j}(t)$  under the following constraints.

1. Only  $\{v_{i,j}(\tau)\}$  where  $t > \tau$  are available for the prediction
2. The convergence of the PDF can be evaluated
3. The processing time to obtain the PDF should be short
4. The PDF formula may be smooth, but must be general

In this paper, probabilistic modeling is discussed as an approach to solve the problem.

### Related Works

Scientific data is one of the targets of study in the data engineering field. Compressed sensing (Donoho 2006) (Bourgignon, Carfantan, and Böhm 2007) is a popular technology for reconstruction by sparse modeling of a few data. Compressed sensing assumes that the signal sources are sparsely distributed. Under the constraint of sparsity, a compressed sensing model can be accurately estimated without many sensing data. In such case, we can reduce the sensing data without loss of accuracy. Here, compressed sensing has been used to accelerate the measurements (Lustig et al. 2008). On the other hand, while smoothness is suitable for fast SANS, sparsity might be unsuitable for it.

The goal of the prediction is also similar to those of image reconstruction and restoration methods. A popular method for image restoration is the random Markov field (Zhang 1993). It might be applied to the SANS pattern, though the result is not likely to be so accurate because the restoration is developed for visually improving the images.

Kernel density estimation (KDE) (Silverman 1986) (Simonoff 1996) is also a well-known method to estimate a smooth distribution for point-like events. In KDE, every observed point is inferred to have a probability around it. The effect is determined by the “kernel function”, of which the parameter is the distance from the observed point  $\mathbf{x}_n$ . The PDF  $P(\mathbf{x})$  is formed as the sum of the contributions as follows.

$$P(\mathbf{x}) = \frac{1}{hN} \sum_n K\left(\frac{|\mathbf{x} - \mathbf{x}_n|}{h}\right), \quad (1)$$

where  $\mathbf{x}_n$  is the  $n$ th observed point,  $h$  is a bandwidth parameter to indicate the effective area of a point, and  $N$  is the number of points. A Gaussian function  $\mathcal{N}(\mathbf{x}|\mathbf{x}_n, hI)$  is often used as the kernel function for generality, where  $I$  is the unit matrix and  $x_n$  is the  $n$ th observation.

A criterion to determine termination of the experiment is required for fast SANS. To evaluate the convergence of the PDF, a simple criterion is how much the PDF was changed. For KDE, the difference between the latest PDF and the previous one should be calculated as the criterion. The criterion  $C_{\text{KDE}}(t)$  at time  $t$  is derived as follows.

$$C_{\text{KDE}}(t) = \int |P(\{\mathbf{x}_1, \dots, \mathbf{x}_t\}) - P(\{\mathbf{x}_1, \dots, \mathbf{x}_{t-1}\})| d\mathbf{x}. \quad (2)$$

A disadvantage of this criterion is the timing for detecting convergence. Namely, the criterion can detect convergence only after the PDF has converged. For a fast SANS experiment, the detection of the convergence should be simultaneous with convergence. Convergence of the parameter might be used as a criterion for GMM modeling, instead of a PDF. However, in this case as well, convergence is detected after convergence.

### GMM-based prediction

#### Bayesian Early-stopping

In this study, Bayesian estimation approach is taken to obtain a criterion for the convergence of the PDF. In the Bayesian approach, the parameters of the PDF, denoted as  $\{\eta\}$ , are also probabilistic, in addition to observations  $\{\mathbf{x}\}$ . The PDF of the parameters  $\{\eta\}$  when  $\{\mathbf{x}_n\}$  are given is denoted as  $P(\{\eta\}|\{\mathbf{x}_n\})$  hereafter. The basic idea of the proposed method is that the variance of  $P(\{\eta\}|\{\mathbf{x}_n\})$  is used as the criterion of experiment termination because the PDF indicates the uncertainty of  $\{\eta\}$ .

Two requirements, i.e., high accuracy and short processing time, should be satisfied for a fast SANS experiment. Both accuracy and processing time depend on the prior setting. If a prior similar to the  $P(\{\eta\}|\{\mathbf{x}_t\})$  is used,  $P(\{\eta\}|\{\mathbf{x}_t\})$  is considered to converge earlier. In the proposed method,  $P(\mathbf{x})$  is roughly estimated before the Bayesian estimation, and the prior is constructed along the roughly estimated  $P(\mathbf{x})$ .

#### Variational Bayesian Estimation for GMM

The proposed method is based on variational Bayesian (VB) estimation (Waterhouse, MacKay, and Robinson 1996) (Attias 1999). VB estimation is frequently used to estimate

$P(\{\eta\}|\{\mathbf{x}_n\})$ . VB estimation approximately assumes that the prior of the GMM parameters is independent from the distribution of the latent parameters. Under this assumption, the PDF estimation is separated into two steps: estimation of the latent parameter and improvement of the PDF of the parameters. The whole process is called the Bayesian EM algorithm (Bishop 2006).

The PDF of the GMM for fast SANS is as follows.

$$P(\mathbf{x}|\{\pi_k\}, \{\mu_k\}, \{\Lambda_k\}) = \sum_k \pi_k \mathcal{N}(\mathbf{x}|\Lambda_k, \mu_k), \quad (3)$$

where  $\mathbf{x}$  is a point on the detection plane,  $\mathcal{N}(\mathbf{x}|\Lambda_k, \mu_k)$  is the  $k$ -th two-dimensional Gaussian function with a correlation matrix denoted by  $\Lambda_k$  and an average denoted by  $\mu_k$ .  $\pi_k$  is the mixing ratio of the  $k$ th component. The priors for the parameters  $\{\pi_k\}$ ,  $\{\mu_k\}$ , and  $\{\Lambda_k\}$  are set

$$\begin{aligned} P(\{\pi_k\}, \{\mu_k\}, \{\Lambda_k\} | W_0, \nu_0, \beta_0) \\ = \text{Dir}(\{\pi_k\}|\alpha_0) \prod_k \mathcal{N}(\mu_k | (\beta_0 \Lambda_k)^{-1}, \mathbf{m}_0) \mathcal{W}(\Lambda_k | W_0, \nu_0). \end{aligned}$$

$\alpha_0, W_0, \nu_0$ , and  $\beta_0$  are the given parameters of the prior (called hyper-parameters).  $\mathcal{W}(W, \nu)$  is the function called the Wishart distribution with the deviation  $W$  and degrees of freedom  $\nu$ .  $\text{Dir}(\{\pi\}|\alpha)$  is the Dirichlet distribution with parameter  $\alpha$ .

For the Bayesian estimation, a posterior with new observations  $\{\mathbf{x}_n\}$  is calculated. If the prior mentioned above is used, the formula of the posterior is the same as that of the prior. That is to say, the hyper-parameters  $\alpha_0, \beta_0, \mathbf{m}_0, W_0$  are revised through Bayesian estimation.

$$\alpha_k = \alpha_0 + N_k, \quad \beta_k = \beta_0 + N_k, \quad (4)$$

$$\mathbf{m}_k = \frac{1}{\beta_k} (\beta_0 \mathbf{m}_0 + N_k \bar{\mathbf{x}}_k), \quad \nu_k = \nu_0 + N_k, \quad (5)$$

$$W_k^{-1} = W_0^{-1} + S_k N_k + \frac{N_k \beta_0}{N_k + \beta_0} (\mathbf{m}_0 - \bar{\mathbf{x}}_k)(\mathbf{m}_0 - \bar{\mathbf{x}}_k)^T, \quad (6)$$

where  $N_k$ ,  $\bar{\mathbf{x}}_k$ , and  $S_k$  are respectively calculated as

$$N_k = \sum_n r_{n,k}, \quad \bar{\mathbf{x}}_k = \frac{1}{N_k} \sum_n r_{n,k} \mathbf{x}_n, \quad (7)$$

$$S_k = \sum_n r_{n,k} (\mathbf{x}_n - \bar{\mathbf{x}}_k)(\mathbf{x}_n - \bar{\mathbf{x}}_k)^T. \quad (8)$$

$r_{n,k}$ , which is called "responsibility", is calculated as

$$r_{n,k} = \frac{\rho_{n,k}}{\sum_n \rho_{n,k}}, \quad (9)$$

$$\rho_{n,k} \equiv \tilde{\pi}_k |\tilde{\Lambda}_k|^{\frac{1}{2}} e^{\left[-\frac{\beta_k}{2} - \frac{1}{2}(\mathbf{x}_n - \mathbf{m}_k)^T W_k (\mathbf{x}_n - \mathbf{m}_k)\right]}, \quad (10)$$

$$\ln \tilde{\pi}_k \equiv \ln \psi(\alpha_k) - \ln \psi\left(\sum_k \alpha_k\right), \quad (11)$$

$$\ln |\tilde{\Lambda}_k| \equiv \ln \psi\left(\frac{\nu_k}{2}\right) + \ln \psi\left(\frac{\nu_k - 1}{2}\right) + \ln 4 |W_k|, \quad (12)$$

where  $\psi(\cdot)$  is a digamma function. The responsibility  $r_{n,k}$  indicates the contribution to each component labeled by  $k$ . The calculation of the responsibility requires the hyper-parameters; however, the hyper-parameter estimation also

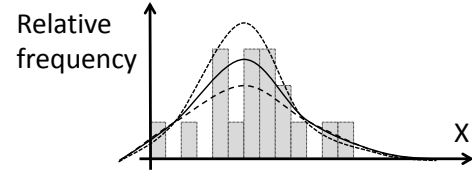


Figure 2: PDF distribution

requires the responsibility. So, the above-described processes are iterated to optimize the hyper-parameters. After iterating the calculation, the optimized hyper-parameters are obtained.

The probability distribution is estimated with marginalization of every hidden parameter,

$$\begin{aligned} P(\mathbf{x}) = \int \sum_z P(\mathbf{x}|\mu_k, \Lambda_k) P(z = z_k | \{\pi_k\}) \\ \times P_k(\mu_k, \Lambda_k | \mathbf{m}_k, \beta_k, W_k, \nu_k) P(\{\pi_k\} | \{\alpha_k\}) d\pi d\Lambda d\mu. \end{aligned} \quad (13)$$

The integration can be carried out as follows.

$$P(x) = \frac{\sum_k \hat{\alpha}_k \text{St}(x | \hat{m}_k, L(\beta_k, \nu_k, W_k), \hat{\nu}_k - 1)}{\sum_{k'} \hat{\alpha}_{k'}}, \quad (14)$$

where

$$L(\hat{\beta}_k, \hat{\nu}_k, \hat{W}_k) \equiv \frac{(\hat{\nu}_k - 1) \hat{\beta}_k}{1 + \hat{\beta}_k} \hat{W}_k. \quad (15)$$

## Criteria for Terminating the Experiment

In the Bayesian approach, the PDF of the parameters is estimated and predicted SANS pattern is derived with marginalization of it, as shown above. The relation between probability and the observations of the SANS pattern is illustrated in Fig. 2. If the observations shown as a histogram are input, the PDF of the parameters is determined. Various PDFs of new observations can be estimated depending on the PDF of the parameters. Through marginalization, the average of such PDFs is calculated for predicting  $\mathbf{x}$  (drawn as the solid line). That is, the expectation value of  $P(\mathbf{x})$  shown in formula (13) is written as

$$P(\mathbf{x}) = \mathbb{E}_{\pi \Lambda \mu} \left( \sum_z P(\mathbf{x}|\mu_k, \Lambda_k) P(z = z_k | \{\pi_k\}) \right) \quad (16)$$

where  $\mathbb{E}_{\pi \Lambda \mu}$  is the expectation value over  $\pi \Lambda \mu$ . However, the other PDFs such as those indicated by the dashed line and the dotted line in the figure are possible estimates. The variance of the probability, therefore, is considered to be a criterion for predicting convergence.

The variance of the probability is similarly calculated as  $\mathbb{E}(P(\mathbf{x})^2) - \mathbb{E}(P(\mathbf{x}))^2$ . The first term is written as a sum of components  $F_{i,j}$ .

$$\begin{aligned} \mathbb{E}(P(\mathbf{x})^2) = \sum_{i,j} \int P(\mathbf{x}|\mu_i, \Lambda_i) P(\mathbf{x}|\mu_j, \Lambda_j) P(z = z_i | \{\pi_i\}) \\ \times P(z = z_j | \{\pi_j\}) d\pi d\Lambda d\mu \equiv \sum_{i,j} F_{i,j}. \end{aligned} \quad (17)$$

The integrations in  $F_{i,j}$  are categorized into two types. When the two indices are the same, as in  $F_{k,k}$ , we have

$$F_{k,k} = \frac{\alpha_k(\alpha_k + 1)}{\sum_i \alpha_i \sum_j (\alpha_j + 1)} \times \frac{1}{4\pi} |W|^{\frac{1}{2}} (\nu - 1) \text{St}(\mathbf{x}|\hat{\mathbf{m}}_k, L(\frac{1}{2}\hat{\beta}_k, \hat{\nu}_k + 1, 2\hat{W}_k), \nu_k). \quad (18)$$

If the two indices are different, we have

$$F_{i,j} = \frac{\alpha_i \alpha_j}{\sum_{l,m} \alpha_l (\alpha_m + 1)} \text{St}(\mathbf{x}|\hat{\mathbf{m}}_i, L_i, \hat{\nu}_i - 1) \text{St}(\mathbf{x}|\hat{\mathbf{m}}_j, L_j, \hat{\nu}_j - 1). \quad (19)$$

Moreover,  $\mathbb{E}(P(\mathbf{x}))^2$  is similarly described as a sum of components.

$$\mathbb{E}(P(\mathbf{x}))^2 = \sum_{i,j} \frac{\hat{\alpha}_i \hat{\alpha}_j}{(\sum_k \hat{\alpha}_k)^2} \text{St}(\mathbf{x}|\hat{\mathbf{m}}_i, L(\beta_i, \nu_i, W_i), \hat{\nu}_i - 1) \times \text{St}(\mathbf{x}|\hat{\mathbf{m}}_j, L(\beta_j, \nu_j, W_j), \hat{\nu}_j - 1) = \sum_{i,j} \tilde{F}_{i,j} \quad (20)$$

The formulas of  $F_{i,j}$  and  $\tilde{F}_{i,j}$  are very similar. Actually, if  $N_k \rightarrow \infty$ , we have

$$F_{i,j} \simeq \frac{\alpha_i \alpha_j}{(\sum_i \alpha_i)^2} \mathcal{N}(\mathbf{x}|\hat{\mathbf{m}}_i, L(\beta_i, \hat{\nu}_i, \hat{W}_i)) \times \mathcal{N}(\mathbf{x}|\hat{\mathbf{m}}_j, L(\beta_j, \hat{\nu}_j, \hat{W}_j)) \simeq \tilde{F}_{i,j}. \quad (21)$$

Thus,  $\tilde{F}_{i,j} - F_{i,j} \rightarrow 0$  when  $N_k$  is large enough. Because  $N_k$  becomes larger when the data increase, this proves that the variance will be 0 when  $t \rightarrow \infty$ .

### B-spline-based Rough Estimation of Priors

As mentioned above,  $\{\mathbf{m}_0, \beta_0, W_0, \nu_0, \alpha_0\}$  are the hyper-parameters for the GMM. If there are enough observations, these hyper-parameters are determined optimally. However for fast SANS, the number of iterations of the VB estimation is limited because the process should finish before the next SANS patterns arrive. Thus, a rough estimate of the PDF is used to generate more suitable priors in the proposed method.

The rough estimation is a regression estimation with multinomial functions, i.e., B-spline curves (Silverman 1985). A B-spline curve is a combination of multinomial functions, determined by given parameters that correspond to points, called control points. Thus, the control points should be optimized to fit the PDF to the SANS patterns.

The estimated PDF  $P(\mathbf{x})$  with B-spline curves is used for generating the prior for the VB estimation. If the gradient of the estimated function is higher, more Gaussian functions of GMM will be required. So in the proposed methods, more components of the Gaussian prior  $\mathbf{m}_0$  should be generated in the area where the gradient is high. Namely  $\mathbf{m}_0$  is selected randomly with the probability that is in proportion to the strength of the gradient  $|\nabla P(\mathbf{x})|$ . Even if the number of components is limited due to processing time, we can select a prior suitable for the posterior.

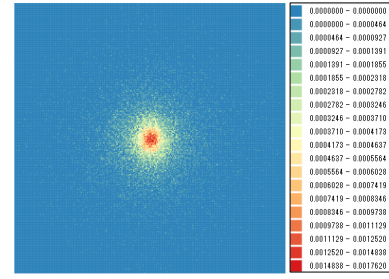


Figure 3: SANS pattern after 1 minute

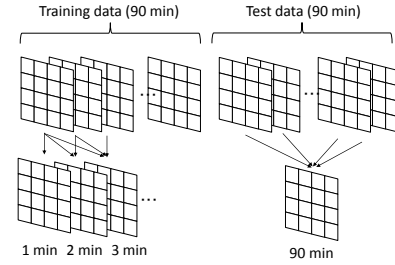


Figure 4: Experimental data setting

## Experiments

### Experimental Settings

Experiments using actual SANS data were carried out to examine the accuracy of the proposed method, that is, GMM with B-spline-based prior. The data for the experiment were from an actual SANS experiment lasting 180 minutes (called the “180min SANS pattern” here). The sample for the experiment was a polycrystalline metallic sample, for which gives SANS patterns are intense at the center. The SANS pattern consists of  $256 \times 256$  cells with intensity. The 180min SANS pattern was divided into 180 patterns that included the observations accumulated in 1 minute. Figure 3 shows one of these patterns (called a “1min SANS pattern”). By aggregating  $x$  1-min SANS patterns, we can generate an “ $x$  min SANS pattern” freely. The measurement normally takes 20 minutes to obtain accurate SANS patterns. For the experiments, the 1-min SANS patterns were grouped into two sets of 90 patterns, as illustrated in Fig 4. The 90 patterns of one set were aggregated to make a “true” data. The patterns of the other set were used as inputs to the prediction methods.

The accuracies of KDE and GMM with the VB estimation (called simply GMM) were evaluated against those of the proposed method, that is, GMM with B-spline-based prior (called BSGMM). The logarithmic likelihood, expressed as  $\sum_t \ln P(x_t)$ , was used as the criterion for accuracy. Hyper-parameters of the methods have to be determined before the evaluation. For the VB estimation of GMM, they were set naive values, that is,  $\mathbf{m}_0 = \mathbf{0}$ ,  $\beta_0 = 1.0$ ,  $W_0 = I$ ,  $\nu_0 = \mathbf{0}$ ,  $\alpha_0 = 0$ , and the number of mixed components was 50. The BSGMM estimation was carried out with the same settings except  $\mathbf{m}_0$ .

The bandwidth parameters  $h$  of KDE should be determined experimentally. As the result of testing  $h = 2, 4, 6$ ,

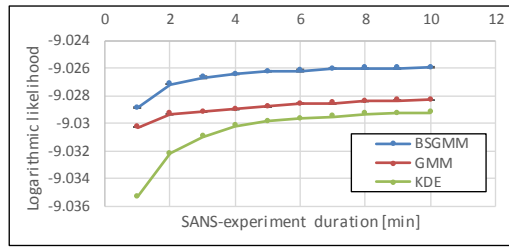


Figure 5: Accuracy in experiment 2

and 8, the accuracy of  $h = 4$  was better than those of  $h = 6$  and 8. The accuracy of  $h = 2$  results were lower than  $h = 4$  before 4 min SANS pattern was input. However, the initial behavior is crucial for fast SANS because the idea is for the experiment to finish earlier. Thus,  $h = 4$  was considered most reasonable for fast SANS.

Generally speaking, the iteration processes for the VB estimation should continue until convergence. However processing time is limited for the fast SANS. Thus to keep the processing time sufficient for fast SANS, the number of the iteration was set to constant. So the number of the iterations in VB estimation was adjusted to 400 manually to finish the process in 1 minute. Additionally, KDE was also tuned. If  $|\mathbf{x} - \mathbf{x}_n| > 3\sigma$ , the kernel function was set to 0. The effects in such case were quite small, so the calculation can be omitted to finish the calculation in 1 minute.

The CPU of the computer used in the experiment was Intel Core i7-3770 3.40 GHz and the RAM was 8GB. The methods were implemented in Java using JDK 1.8.0 update 144 and multithreading.

## Experimental Results

Figure 5 plots the accuracy of the prediction by BSGMM, GMM and KDE ( $h = 4$ ). The vertical axis indicates the logarithmic likelihood and the horizontal axis is the duration to obtain SANS pattern. BSGMM and GMM are calculated 10 times with changing the random seeds and the averages are plotted in the figure. The error bars indicating the standard errors among them are drawn but they are invisibly small. The accuracy of BSGMM is highest; that of GMM is second; that of KDE is lowest.

For comparison, the accuracy of the conventional method was evaluated. The relative frequency of the SANS pattern is used as the probability, where the probability is set to  $10^{-200}$  at points of which the probability is 0 to avoid infinity to avoid  $\infty$  due to  $\log 0$ . The accuracy at 1 minute was -138.5 and that at 10 minutes was -36.1. With the conventional method, it is generally considered that it takes around 20 minutes for convergence. In contrast, the BSGMM and GMM almost converged at only 4 minutes. This indicates that GMM and BSGMM can shorten the SANS experiment duration from 20 minutes to 4 minutes.

Figure 6 plots the average of the processing time for each method. The process was carried out 10 times and the averages are plotted with the error bar indicating the standard error. All of the calculation were finished in 60 seconds be-

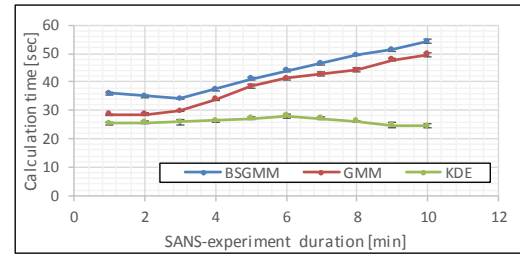


Figure 6: Processing Time in experiment 2

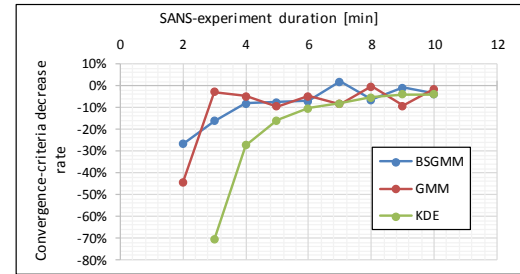


Figure 7: The differential convergence criteria

cause of the tuning. The calculation time of GMM and BSGMM are rising up after 4 minutes, but it keeps around 50 seconds. Thus the processes can be carried out in 1 minute, which is period of SANS pattern. The result shows that the GMM and BSGMM are feasible for the fast SANS.

Additionally the criteria of convergence for each method were evaluated. For KDE,  $C_{KDE}$  defined as formula (2) is suitable. As the discretized and normalized version of it, the average difference between the  $t$  and  $t - 1$  results were the criteria. For GMM and BSGMM, formula (20) was the criteria. Figure 7 plots decrease rates of them (difference between the criteria at  $t$  and  $t - 1$  was divided by the criteria at  $t = 0$ ). The criteria are basically correlated to the accuracies; however according to its criteria, KDE did not converge until around 8 minutes. In contrast, the GMM criteria were almost constant after 3 or 4 minutes. Therefore, the criteria are feasible for judging convergence, and the GMM and BSGMM criteria are the best.

## Discussion

Figures 8 show the predicted patterns using 1-min SANS pattern and the original SANS pattern. The GMM and BSGMM results are smoother than the KDE result. This implies that the number of data is too small for KDE to give an estimate. To obtain smoother results with KDE, the bandwidth has to be larger, but a larger bandwidth flattens the SANS pattern. As shown, GMM and BSGMM are comparably smooth, though BSGMM result is not symmetric. This asymmetry comes from the asymmetry of the prior.

In SANS experiments, changes in the logarithmic cross sections along the radius are significant. Hence, the differential logarithmic cross sections are plotted in Fig. 9. The horizontal axis is  $\log_{10} r$ , where  $r = |x - 128|$  and the ver-

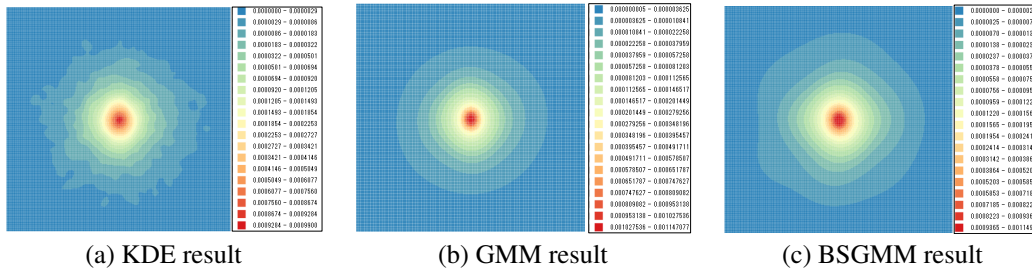


Figure 8: Predicted SANS patterns

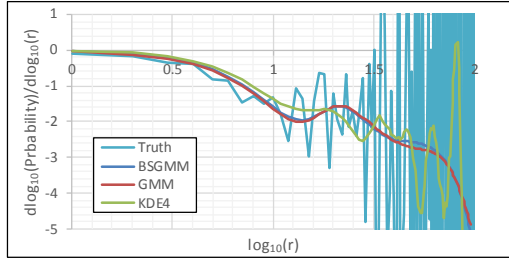


Figure 9: Differential Cross Section at  $y = 128, 128 < x$

tical axis is the differential of  $\log_{10} P$  as follows.

$$\frac{\log_{10} P(\mathbf{x}_n) - \log_{10} P(\mathbf{x}_0)}{\log_{10} |\mathbf{x}_n - \mathbf{x}_c| - \log_{10} |\mathbf{x}_0 - \mathbf{x}_c|}, \quad (22)$$

where  $\mathbf{x}_n = (n + 128, 128)$  and  $\mathbf{x}_c = (128, 128)$ . Because KDE is based on smoothing, the results of KDE tend to be flat. Thus, the differential values of KDE are nearer to zero than those of GMM and BSGMM. The truth is actually similar to the GMM and BSGMM results.

The truth frequently oscillates at high  $\log_{10} r$ . The oscillation is caused by the inaccuracy of the SANS pattern. KDE similarly oscillates because the PDF is generated from the detected events. In contrast, GMM and BSGMM do not oscillate unlike KDE. As the result, we can see a peak around  $\log_{10} r = 1.3$ , which implies that there are effects from the microstructure around the scale. GMM and BSGMM are hence superior in this aspect.

As discussed above, both GMM and BSGMM are applicable for fast SANS. However remember that BSGMM is more accurate than GMM as shown Fig. 5. Therefore it is considered that BSGMM is more appropriate than GMM.

## Conclusion and Future work

A GMM with B-spline-based prior for fast SANS experiments was proposed in this paper. An experiment using actual data confirmed that the GMM-based prediction of SANS patterns requires only a few minutes worth of SANS data to predict the results that will be obtained later. The accuracy converged 4 minutes after starting the experiment; this compares favorably with the around 20 minutes for conventional SANS experiments. Therefore, we conclude that the proposed method shortens the SANS experiment to 1/5th of its usual duration.

SANS is just one of many scattering experiments. That is, there are many scattering-like experiments in materials science: light scattering (such as X-ray scattering and laser scattering), beam scattering (such as ion-beam scattering and electron-beam scattering), and so on. The proposed method may be applied to such experiments. Not only that, the problem that spatial distribution of events should be estimated fast is common in statistics survey. Modifying it for those applications will be a future work.

## References

- Attias, H. 1999. Inferring parameters and structure of latent variable models by variational bayes. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, 21–30. Morgan Kaufmann Publishers Inc.
- Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer.
- Bourguignon, S.; Carfantan, H.; and Böhm, T. 2007. Sparspec: a new method for fitting multiple sinusoids with irregularly sampled data. *Astronomy & Astrophysics* 462(1):379–387.
- Donoho, D. L. 2006. Compressed sensing. *IEEE Transactions on information theory* 52(4):1289–1306.
- Higgins, J. S., and Benoît, H. 1994. *Polymers and neutron scattering*. Clarendon press Oxford.
- Lustig, M.; Donoho, D. L.; Santos, J. M.; and Pauly, J. M. 2008. Compressed sensing mri. *IEEE Signal Processing Magazine* 25(2):72–82.
- Silverman, B. W. 1985. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)* 1–52.
- Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC.
- Simonoff, J. S. 1996. *Smoothing methods in statistics*. Springer.
- Waterhouse, S. R.; MacKay, D.; and Robinson, A. J. 1996. Bayesian methods for mixtures of experts. In *Advances in neural information processing systems*, 351–357.
- Zhang, J. 1993. The mean field theory in em procedures for blind markov random field image restoration. *IEEE Transactions on Image Processing* 2(1):27–40.