Shortening Measurement Time without Hardware Investment

We propose a simple method to shorten the measurement time of experiments using a two-dimensional detector by exploiting spatial correlation in a smooth intensity distribution. We use kernel density estimation (KDE), a statistical tool to estimate the true probability density distribution from observed data, to exploit spatial correlation via a Gaussian kernel. The improvement of statistical quality (variance) by time accumulation and KDE was investigated using a short-time measurement dataset of silica nanoparticles. The time required to achieve a certain variance was reduced by one order of magnitude. For the case of anisotropic data where the strong noise reduction ability of radial averaging is not available, KDE can reduce measurement time by half.

Scientific research is experiencing a new paradigm called data-intensive science, or data-centric science. Materials informatics is one example of an emerging interdisciplinary research field that has arisen under the new paradigm. The aim of the field is to make materials development procedures more efficient by exploiting large or minimum datasets with the help of statistical and/or machine-learning analysis techniques. The first step in a research cycle in the field is data acquisition or preparation, and is often achieved by using existing databases or creating a new one experimentally or computationally. In this regard, large experimental facilities including the Photon Factory have the potential to make significant contributions to database preparation. Some studies have already used experimental data obtained at synchrotron facilities as the input for data-centric materials science studies owing to the capability of synchrotron facilities to generate a massive amount of data [1, 2]. In contrast, data from neutron facilities have not yet been used for such purposes presumably due to the low measurement throughput. In general, such a drawback can be overcome only by costly hardware upgrades to increase the beam flux or to reduce the overhead time for changing the sample. The Photon Factory and the Paul Scherrer Institute (PSI) have developed a method that, under certain conditions, reduces the measurement time by introducing a well-established tool in statistics without expensive hardware upgrades.

There are two types of correlation in a smooth two-dimensional (2D) intensity distribution with no sharp peaks: temporal correlation and spatial correlation. The former relates to the fact that the count rate per unit time at a pixel always lies in a certain range for multiple measurements following Poisson statistics. The latter means that the count rate of a pixel is similar to those of adjacent pixels because of smoothness. The statistical quality of many kinds of counting experiments has been guaranteed by a simple strategy using temporal correlation, that is, a strategy to reduce the uncertainty of observed counts by accumulating enough counts. On the other hand, a similar approach to improve the statistical quality of a 2D count distribution using spatially-correlated values, which is practically smoothing, has not been widely used, probably because such data processing is regarded as unfavorable data manipulation just to visually deceive other researchers. In fact, however, weighted-averaging based on spatial correlation is called kernel density estimation (KDE) in statistics and is widely used to estimate the true probability density distribution from observed values. Considering that the 2D intensity distribution is equal to the probability density distribution of detection events, it is natural to apply KDE to 2D count maps.

We collected one hundred small-angle neutron scattering (SANS) datasets of silica nanoparticles with low statistical quality by repeating a short measurement at SANS-I, SINQ (PSI). The measurement time was set so that the total counts of the one hundred datasets reached a conventional standard, 300k counts, which is thought to be enough counts to secure statistical quality. **Figure 1** shows raw 2D SANS datasets and the effect of KDE on them. It can be clearly seen that KDE makes the datasets with shorter measurement time close to the raw datasets with longer measurement time. As our datasets are isotropic, we can examine this effect by taking the variance of pixel intensities within concentric rings. The mean of the variances of fifty rings normalized by average counts for each ring is shown in **Fig. 2**. KDE significantly reduces the variance to the same level as that of raw datasets with ten-times longer measurement time, meaning that the measurement time required to obtain a certain level of variance in a 2D SANS dataset can be shortened by one order of magnitude. However, in practice, researchers often use radial-averaged 1D SANS intensity for isotropic data to perform model fitting, and radial averaging also has a strong ability to reduce variance. For this reason, we examined the effect of KDE on radial-averaged 1D intensity and sector-averaged 1D intensity used for anisotropic 2D SANS data. The results are shown in **Fig. 3**. We confirmed that, while the advantage of KDE is less significant for radially-averaged 1D SANS intensity, KDE is still effective on sector-averaged intensity for anisotropic cases, which means that KDE can shorten the measurement time for anisotropic SANS data.

Although it would depend on the samples and other conditions, we believe this method is valid for most 2D scientific data obtained with 2D pixel detectors. Simple ideas in other fields may improve research efficiency without huge investment. We believe there is still room for this type of improvement in synchrotron and neutron experiments.

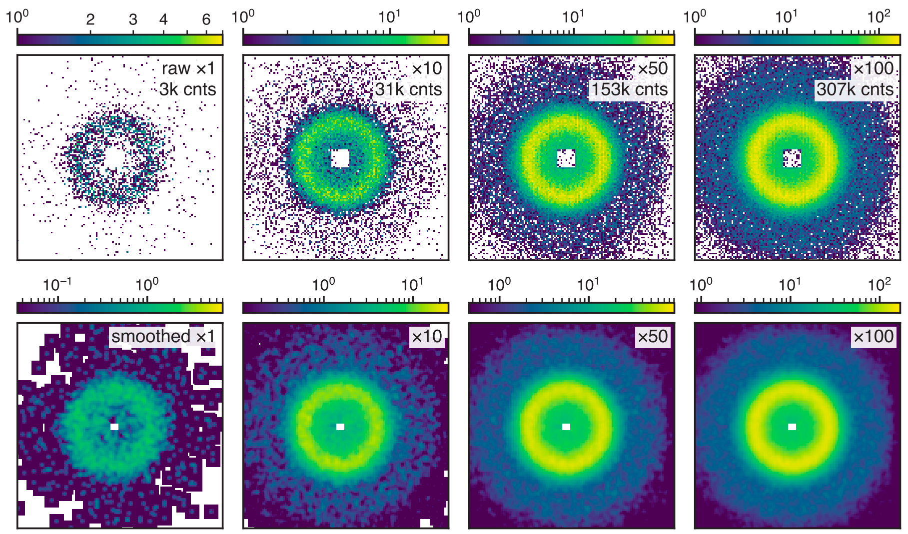

Figure 1: The effect of KDE on 2D SANS data. Raw (upper) and KDE-smoothed (lower) data. White pixels are null pixels [3].

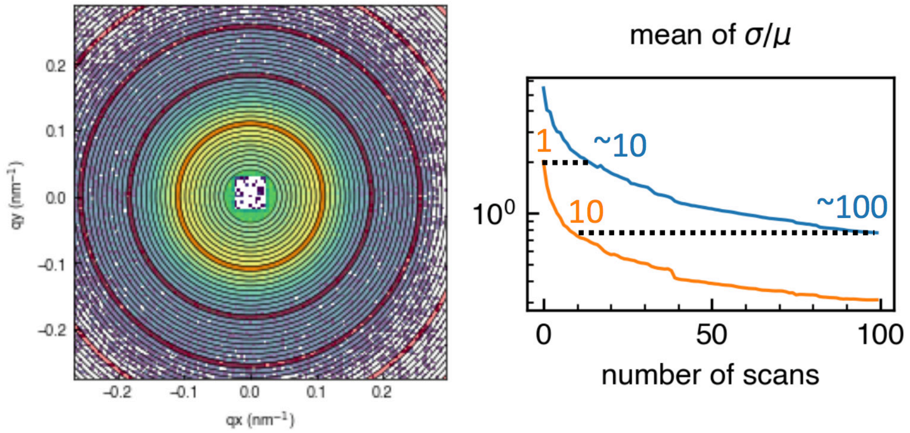

Figure 2: Mean of variances of pixel intensities in 50 concentric rings (left) normalized by the mean intensity of each ring. The blue line is time-averaged data and the orange line is KDE-smoothed data.

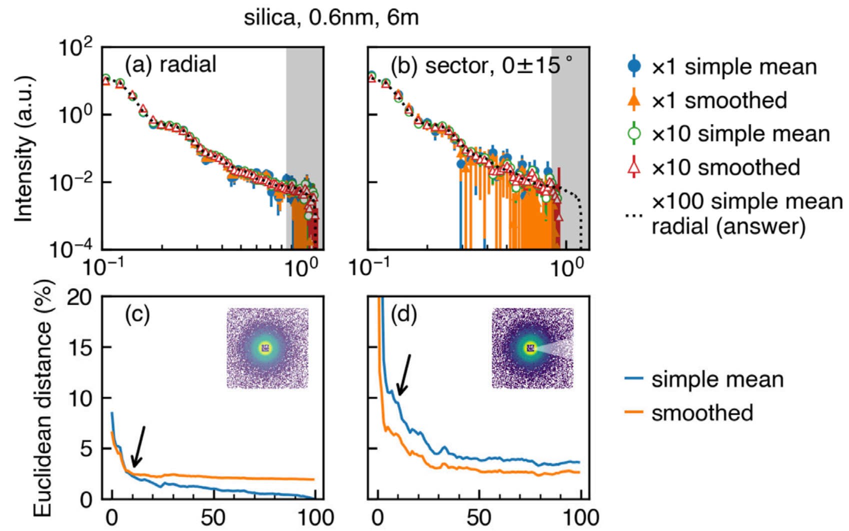

Figure 3: The effect of KDE on 1D intensity. Radial averaging (left column) and sector averaging (right column) [3].

REFERENCES

[1] F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers and A. Mehta, *Sci. Adv.* **4**, eaaq1566 (2018).

[2] I. Akai, K. Iwamitsu, Y. Igarashi, M. Okada, H. Setoyama, T. Okajima and Y. Hirai, *J. Phys. Soc. Jpn.* **87**, 074003 (2018).

[3] K. Saito, M. Yano, H. Hino, T. Shoji, A. Asahara, H. Morita, C. Mitsumata, J. Kohlbrecher and K Ono, *Sci. Rep.* **9**, 1526 (2019).

K. Saito[1] and K. Ono[2] ([1]PSI-LNS, [2]KEK-IMSS-PF)