A Serial Crystallography Error Model Robust to Outlier Observations

David W Mittan-Moreau[1], Vanessa Oklejas[1], Daniel W Paley[1], Nicholas K Sauter[1], Aaron S Brewster[1]

[1]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory

dwmoreau@lbl.gov

In serial crystallography, redundant measurements of each reflection are averaged with weightings of each measurement's uncertainty. A reflection's uncertainty consists of contributions from counting statistics, systematic, and random error. Counting statistics is well understood and is propagated to a reflection's final uncertainty. Systematic and random errors are challenging to model directly and are instead treated empirically. As developed by Brewster (2019), a parametric equation for uncertainty is used to scale counting statistics for gain error and introduces terms that represent random measurement error. This equation's parameters are optimized to explain the variance of the redundant measurements using a nonlinear least-squares approach. In least-squares minimization, a small number of outliers can have an outsized impact as the optimization tends to reduce their contributions instead of appropriately modeling the data. Within cctbx.xfel.merge, outlier reflections are typically caught before merging using a series of filters. This work introduces a complementary approach to where the optimization is made robust to outliers. This is done by replacing Brewster's least-squares minimization target function with a maximum log-likelihood approach that utilizes a student t-distribution.

{1} Brewster, A. S., Bhowmick, A., Bolotovsky, R., Mendez, D., Zwart, P. H. & Sauter, N. K. (2019). Acta Cryst. D75, 959–968.