

Maximum likelihood methods in DIALS

James Parkhurst¹, Graeme Winter¹, David Waterman², Richard Gildea¹, Luis Fuentes-Montero¹, Garib Murshudov³, Gwyndaf Evans¹
¹Diamond Light Source, Didcot, United Kingdom, ²CCP4, Didcot, United Kingdom, ³Laboratory of Molecular Biology, Cambridge, United Kingdom

E-mail: james.parkhurst@diamond.ac.uk

In macromolecular crystallography, integration programs - such as DIALS (Waterman et al. 2013) - are used to estimate the intensities of Bragg reflections recorded on a series of X-ray diffraction images. The reflection intensities are estimated using the following procedure. A model for the shape of the reflection profile is estimated from a set of strong reflections. This model is then applied to each reflection in order to estimate the size and shape of the reflection on the detector surface and to label each pixel as either foreground or background. The intensity of each reflection is then estimated (in the case of summation integration) by summing the total counts minus the estimated background counts in the foreground region. Since the background level under the reflection peak cannot be measured directly, it is estimated from the surrounding background pixels assuming a given model.

Long wavelength beamlines provide additional challenges to integration programs. For long wavelength beamlines, scattering of the photons by air between the crystal and the detector can have a significant effect, resulting in large background and consequently reduced signal to noise ratio. It is for this reason that beamlines such as I23 at Diamond Light Source (Wagner et al. 2016) have been constructed with the detector and sample being completely encased in a vacuum, thereby avoiding the issue of air scatter. This results in very low background on the detector pixels.

Due to the low background and weak diffraction data from long wavelength beamlines, special care needs to be taken when processing the data. Pixel values from photon counting detectors are Poisson distributed; simplifying assumptions of an approximate Normal distribution of pixel counts no longer hold for small pixel counts. Use of maximum likelihood based methods in integration allows the distribution of the pixel counts to be modelled explicitly by a Poisson distribution allowing a proper statistical treatment of the X-ray diffraction data.

Here, we present maximum likelihood-based algorithms for profile modelling, background estimation and profile fitting. Profile modelling employs a reciprocal space profile model whose parameters are estimated directly in detector space using maximum likelihood methods allowing better modelling of complex spot shapes (such as streaky or irregularly shaped spots). Background modelling is performed using a generalized linear model framework that performs unbiased outlier handling for Poisson distributed data. It is shown that this algorithm has better statistical properties than traditional methods which may introduce systematic bias into the reflection intensities.

[1] Waterman, D. G. et al. (2013). CCP4 newsletter on protein crystallography, 49, 16–19.

[2] Wagner, A. et al. (2016). Acta Crystallographica Section D Structural Biology, 72(3), 430–439.

Keywords: [Integration](#), [maximum likelihood](#), [modelling](#)