

The *LaueUtil* toolkit for Laue photocrystallography. II. Spot finding and integration

Jarosław A. Kalinowski,^{a,b*} Bertrand Fournier,^a Anna Makal^{a*} and Philip Coppens^a^aChemistry Department, University at Buffalo, State University of New York, Buffalo, NY 14260-3000, USA, and ^bPhysical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. E-mail: jak@kalinowscy.eu, annamaka@buffalo.edu

A spot-integration method is described which does not require prior indexing of the reflections. It is based on statistical analysis of the values from each of the pixels on successive frames, followed for each frame by morphological analysis to identify clusters of high value pixels which form an appropriate mask corresponding to a reflection peak. The method does not require prior assumptions such as fitting of a profile or definition of an integration box. The results are compared with those of the seed-skewness method which is based on minimizing the skewness of the intensity distribution within a peak's integration box. Applications in Laue photocrystallography are presented.

Keywords: photocrystallography; X-ray diffraction; X-ray detection; CCD; Laue method; spot integration; background estimate; statistical analysis; image filtering.

1. Introduction

Treatment of Laue diffraction data has frequently attracted attention, in particular since the development of time-resolved pump–probe diffraction techniques, in which the Laue method has specific advantages (Ren *et al.*, 1999; Anderson *et al.*, 2004; Kamiński *et al.*, 2010). The renewed interest in the Laue technique has resulted in the development of several computer programs. Notable are the Daresbury Laue Software Suite (Helliwell *et al.*, 1989), *LaueView* and *Precognition* (Ren, 2010; Šrajcar *et al.*, 2000), and *LaueGui* (Messerschmidt & Tschentscher, 2008).

In the commonly used Laue data processing procedures the reflection's position on the detector is predicted and a box is established that encloses the area where the diffraction signal is expected. Subsequent integration of the identified spots is performed either by two-dimensional profile fitting (Helliwell *et al.*, 1989; Šrajcar *et al.*, 2000; Moffat, 2001) or by selecting a mask on the detector surface using statistical criteria as in the seed-skewness method (Bolotovskiy *et al.*, 1995; Bolotovskiy & Coppens, 1997). Although the latter does not in principle require information on the predicted spot position, the practical implementation of the method relies heavily on this information (Messerschmidt & Tschentscher, 2008). The above processing sequence cannot be accomplished without *a priori* knowledge of crystal orientation with respect to a diffractometer-based coordinate system; in other words, without successful indexing of the Laue pattern, which may be time-consuming.

In Laue crystallography the Ewald sphere is replaced by a shell of varying thickness, often referred to as the Ewald region. As a result the diffraction spot profile is affected not

only by the sample's mosaic spread and incident beam divergence but also by the X-ray bandwidth. Therefore reflection 3D profile reconstruction is not possible with existing software and integration is typically accomplished on a frame-by-frame basis, followed by appropriate corrections, scaling and wavelength deconvolution procedures (Šrajcar *et al.*, 2000; Ren & Moffat, 1995).

The RATIO method (Coppens *et al.*, 2009), as used in pump–probe photocrystallography, avoids the spectral deconvolution step. As the interest is in the ratios of the light-ON and light-OFF intensities, collected sequentially in consecutive frames, wavelength-dependent effects such as the spectral distribution of the beam and the absorption and detector response effects are essentially eliminated. Small shifts in spot positions may occur if the cell dimensions are affected by the excitation, an effect that may be pronounced when conversion percentages are appreciable. However, for the low conversion percentages of ~6% or less achieved in many studies including our own (Benedict *et al.*, 2011; Makal *et al.*, 2011; Collet *et al.*, 2012), cell dimension changes are not significant.

In the previous paper on the *LaueUtil* toolkit we proposed a method for a rapid orientation matrix determination in Laue crystallography (Kalinowski *et al.*, 2011). Here we focus on the task of efficient Laue data integration. The presented method combines identification of the diffraction spots and their integration into a single algorithm. The complete set of images is treated as a single dataset with 'frame by frame' or 'pixel by pixel' views used in tandem. The method explicitly uses a statistical approach and requires temporal stability of the source and some simple assumptions on the noise distribution in the background. No prior information about the crystal

orientation, sample cell parameters or their stability in the course of experiments is required.

2. The method

2.1. Assumptions and outline

Our integration method does not imply uniformity of the background across the surface of a diffraction spot. It requires a series of diffraction images at subsequent values of the φ scan angle. For application to data collected at X-ray free-electron laser sources it would require at least modifications to allow for scaling of subsequent frames.

The method consists of three steps. First, for each of the pixels, intensity values are collected from all frames, leading to one-dimensional arrays which are statistically analyzed to estimate the background contributions for each of the individual pixels. The idea is illustrated in Fig. 1. In the second step the mask is defined for each of the frames from the pixel background values previously estimated, and optimized. Finally, the processed masks are analyzed to determine the footprints of each of the reflections which are subsequently integrated to obtain the corresponding intensities.

2.2. Pixel statistics

The measured values of a pixel on all frames can be regarded as a sample of the pixel background intensity which can include an unknown number of outliers corresponding to the presence of spots. The estimate of the background on the frames is then achieved by exclusion of the outliers in per-pixel samples of values. The simplest approach is to assume that a certain percentage (usually about 20–30%) of the highest values along a pixel line (*i.e.* all values of a specified pixel on the successive frames) represent spot contributions. Fig. 2 shows an example of such a pixel-sample analysis. The remaining values are used to estimate, for each pixel, the

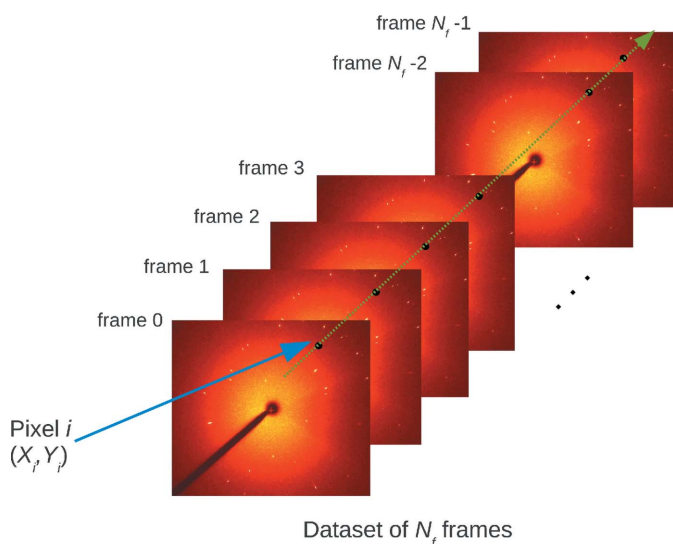


Figure 1 Construction of a statistical sample of values for a chosen pixel: the values for a given pixel are collected for all frames in data collection sequence.

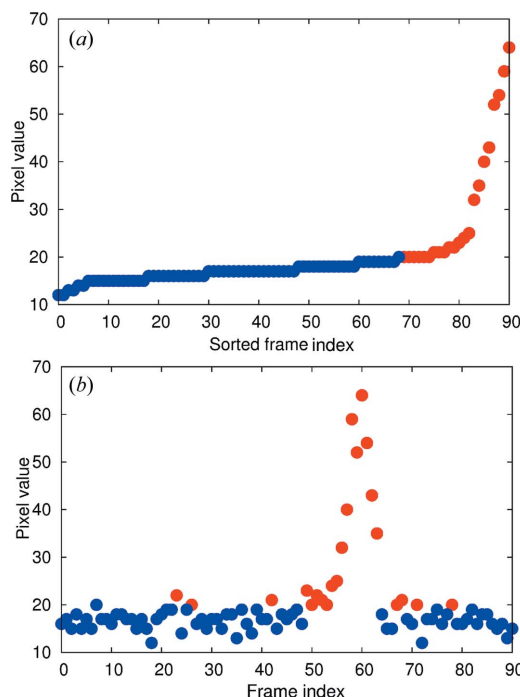


Figure 2 Example of statistical analysis performed using the constant-fraction method on the pixel (657, 1014) of dataset 2; (a) sorted pixel values; (b) a plot of pixel values as a function of frame indices in the dataset. In all images, blue dots correspond to values identified by the method as background contributions, and red dots as outliers.

parameters of its background distribution such as the mean, the variance, the median or the interquartile distance, the latter being defined as the difference between the 25th-percentile and the 75th-percentile statistics of a pixel. Then, from the knowledge of the pixel-by-pixel background characteristics, we use a simple condition to identify which pixels contribute to spots and thereby define a raw mask on each frame. For the j th image, the condition applied to the i th pixel is

$$I_{(i,j)} - \langle I_i \rangle_{\text{background}} > c \sigma(I_i)_{\text{background}}, \quad (1)$$

where $I_{(i,j)}$ is the i th pixel value in the j th image, $\langle I_i \rangle_{\text{background}}$ and $\sigma(I_i)_{\text{background}}$ are, respectively, the mean value and the standard deviation of the i th pixel background, and c is a positive adjustable parameter with a default value of 3.0.

This simple approach, henceforth referred to as the constant-fraction approach, though reasonably successful, suffers from two drawbacks. The first is that the background tends to be underestimated for pixels which contribute to reflections only in a few frames. The second is that some intense reflections located close to the origin of reciprocal space can be present on many adjacent frames, in some cases spanning a range of over 25° in the φ scan angle and contributing to the values of some pixels on a significant portion of the frames which shows up as spot-like features in the reconstruction of the background as discussed in §3.1. As all the statistical descriptors are available for immediate inspection in the output file, a user can decide on the trade-off

between a global underestimate of the background due to too large a removed fraction *versus* a local background overestimate due to an incomplete filtering of the signal.

2.3. Advanced pixel statistics for redundant measurements

We consider a pump–probe experiment in which, for each goniometer setting, a series of repeated measurements is made, both with or without the laser pump pulses (referred to below as the ON and OFF frames). In our experiments, measurements are made ten times in both situations, giving for each pixel a block of 20 frames for a single goniometer setting. This allows an improved statistical analysis. We consider the 20 values in each block as independent statistical samples. We then test whether all block samples are from the same background distribution, or whether some of them, so-called ‘block outliers’, are from distributions with higher median value as a result of the presence of diffraction spots.

We use the non-parametric Kruskal–Wallis (K–W) test (Kruskal & Wallis, 1952; Corder & Foreman, 2009) which checks whether samples originate from the same distribution (see Appendix A). For each pixel, all blocks of 20 values are sorted by their median. Then the one with the highest median value is recursively eliminated until the remaining samples pass the K–W test using the $\chi^2_{\alpha;n-1}$ approximation for K–W test critical values, where n is the number of remaining blocks and α is a user-selected significance, with a default value of 5%. Fig. 3 illustrates the K–W analysis performed on a given pixel’s

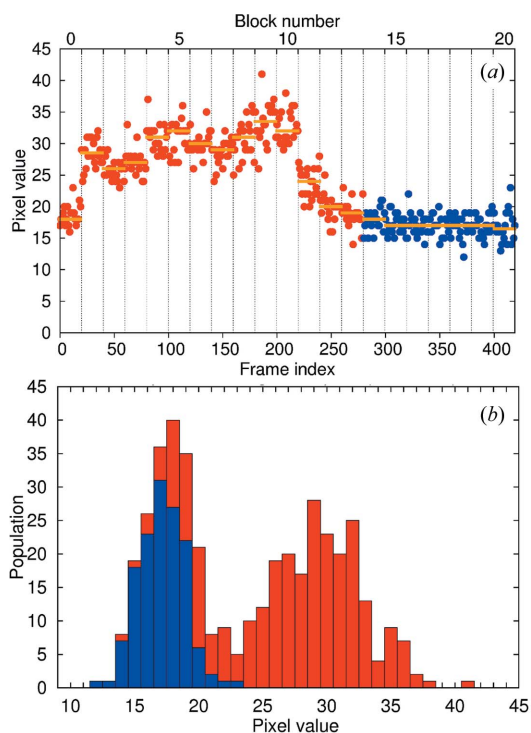


Figure 3

Example of statistical analysis performed using the K–W method on the pixel (679, 1039) of dataset 1 (subset 1): (a) plot of pixel values as a function of frame indices in the dataset; (b) the pixel values histogram. In all images, dots and histogram bars in blue correspond to values identified by the method as background contributions, and in red as outliers. In (a) the orange segments correspond to the block median values.

series of blocks. Like in the constant-fraction method, for each pixel the remaining samples are combined and used to estimate the distribution parameters listed earlier. This method allows efficient detection of block outliers. However, it is not designed to identify erratic singular outliers which are related to noise. To avoid any bias in the subsequent mask definition caused by these outliers, we apply a similar criterion as in the constant-fraction method but using the estimated median and interquartile distance, which are more robust to erratic outliers.

For the i th pixel the criterion for acceptance as part of a mask is

$$I_{(i,j)} - \text{median}(I_i)_{\text{background}} > d \text{iq}(I_i)_{\text{background}}, \quad (2)$$

where $I_{(i,j)}$ is the i th pixel value in the j th image, $\text{median}(I_i)_{\text{background}}$ and $\text{iq}(I_i)_{\text{background}}$ are, respectively, the estimated median value and interquartile distance of the i th pixel background, and d is a positive adjustable parameter with a default value of 3.0.

2.4. Filtering masks

The procedure described above is carried out on a pixel-by-pixel basis and is completely insensitive to the spatial relationships of the pixels. We expect, however, that spots have a non-negligible size and some definite shape though not necessarily as simple as circular or elliptical. The next step is therefore based on a frame-by-frame analysis and mask filtering applying binary morphological operations in two steps (Pierre, 2003). First, erosion operations are applied to remove isolated pixels or lines from a mask. Then, dilation operations are used to add back some relevant pixels lost during erosion operations and also to add a margin to spot footprints. Assuming no bias in the background estimate, we expect that the only adverse effect of the latter operation on the resulting intensities may be an increase in their variances. The benefit of applying dilations is a possible correction of pixel omissions from the previous phase, which may occur at the edge of the spots where the increase of intensity is below the threshold for the pixel-by-pixel outlier detection (Fig. 4). The numbers of erosion and dilation operations are adjustable parameters with default values of 1 and 2, respectively.

The erosion-dilation method produces an improved mask for each frame. These masks can be used directly in the integration step, or can be merged per ON/OFF pair or per block by performing a logical OR operation to obtain a mask shared by the corresponding pair or block of frames. This prevents possible bias due to differences in spot footprints in calculating the ON/OFF intensity ratios. Masks resulting from both types of merging are compared in Fig. 5.

The number of spot candidates produced by the method is generally larger than the number of Bragg reflections that could occur on a given frame. This results from high-intensity noise that may be indistinguishable from spots or from possible splitting of very elongated spots by the erosion operation. We do not consider an excessive number of spots a significant issue as they are effectively filtered by subsequent

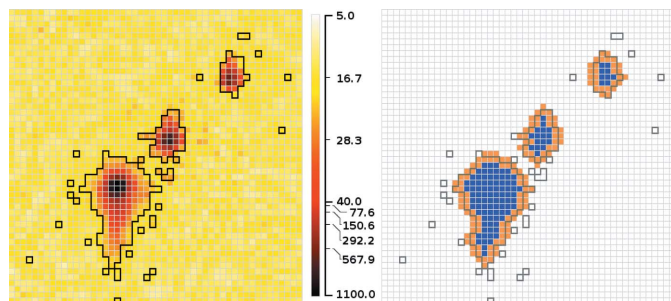


Figure 4 Application of binary morphological operations on the initial spot masks: on the left a magnified fragment of frame 366 from dataset 1 (subset 1) containing the registered signal and an outline of an initial reflection mask drawn in black; on the right the same outline of the initial mask is drawn in dark grey; pixels remaining in the spot mask after erosion operations are drawn in blue; pixels added to the reflection mask after dilation operations are drawn in orange. Singular pixels included in the initial spot mask have been removed by the morphological operations.

application of the *LaueUtil* indexing routine. It should also be noted that in some cases too many dilations can cause merging of particularly extended spots; therefore it is advisable to visually inspect the resulting masks to verify the choice of morphological operations.

2.5. Final integration and implementation details

The resulting frame masks are scanned to identify reflection footprints which are assigned a unique numerical label *l*. Their integrated intensities are calculated as

$$I_l = \sum_{(i,j) \in M_l} [I(i,j) - B(i,j)], \quad (3)$$

where M_l is a list of pixels belonging to spot *l*, $I(i,j)$ is the intensity of pixel (i,j) and $B(i,j)$ is the estimated mean, or median in the K–W method, of the background intensity distribution.

The presented method does not directly provide an estimate of intensity errors (σI) for each of the reflection intensities. Instead, sample statistics based on the redundant measurements are used to estimate the errors.

Following the general design of the *LaueUtil* suite, integration results are stored in HDF5 files together with masks and collected statistics. This choice allows for easy inspection and analysis of the data using general purpose HDF5 data visualization programs like *HDFView* (The HDF Group, 2010) or advanced statistical toolkits like *R* (R Development Core Team, 2010).

The automatic data integration procedure, including retrieval of the experimental data from frames and storing them in HDF5 format (with compression), takes from 5 min for 90 frames, 0.7 GB standard data collection, 20 min for a short-diagnostic 20° scan with ten ON/OFF frame pairs per angle (420 frames, 3.3 GB), up to about 90 min for a full 90° photocrystallographic dataset with ten ON/OFF frame pairs per angle (1800 frames, 14.5 GB). Tests were performed on a standard desktop computer, *i.e.* a Linux machine using a single core of an AMD Phenom II X6 1090T processor. The running times should be compared with the time required to perform sole data compression with standard system tools (*tar -z*)

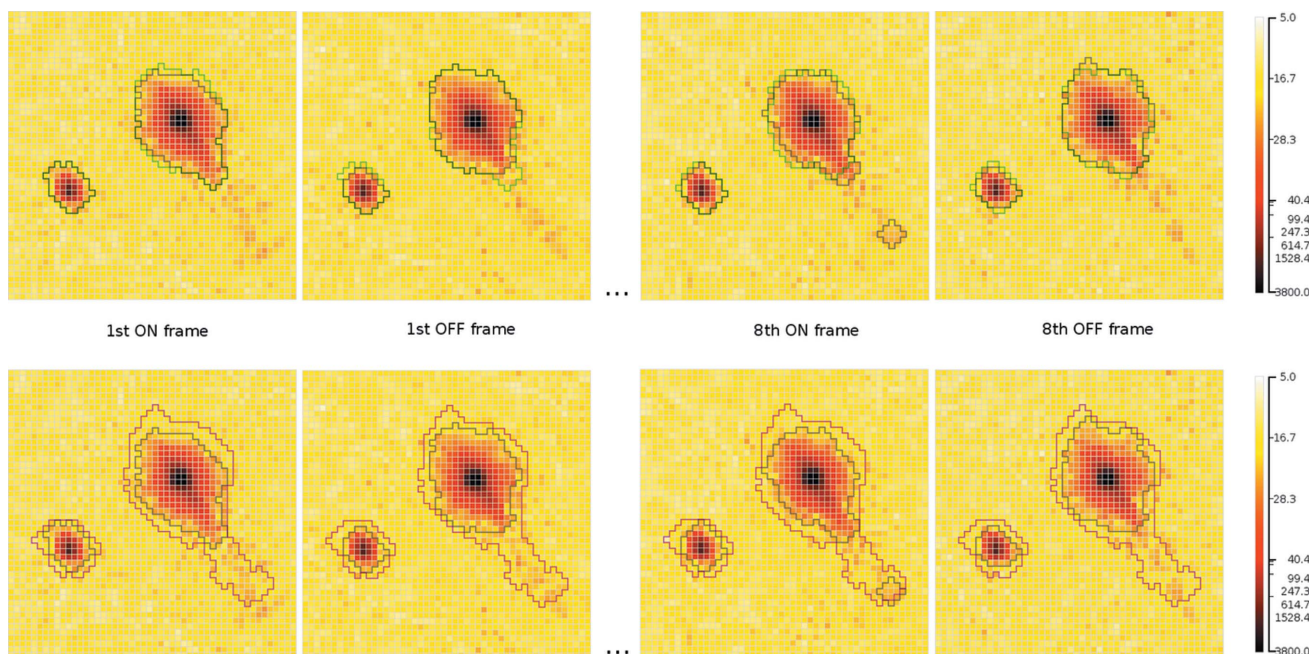


Figure 5 Results of obtaining common masks for a set of raw frames. The 1st and 8th pair of frames from the 18th block of dataset 1 (subset 1), *i.e.* frames 340, 341, 354 and 355, were selected for the presentation; the detector region was $X: 760\text{--}810$; $Y: 808\text{--}858$ pixels. Raw images are colored by intensity; outlines of filtered masks for each frame are presented in black; outlines of common masks for each ON/OFF pair are overlaid in green in the upper row; outlines of a common mask for each block are overlaid in magenta in the lower row. Certain variation of the final mask shape between the frames is visible. A common mask per block allows for incorporation of a reflection tail, too weak to be recognized as a signal on some frames.

which are, respectively, 1 min, 4 min, 21 min and network transfer times: 19.3 min for 14.5 GB at a theoretical maximum speed of 100 Mbps Ethernet connection.

The algorithm is suitable for reimplementing in a high-throughput processing system, with several optimizations envisageable, including parallelization, direct connection with data collection software, in-memory processing of whole datasets, limited usage of compression and storage of intermediate data. However, such developments would exceed our current needs and would require sufficient support from the computing infrastructure used for data collection.

3. Results and discussion

Application of the method is illustrated with two datasets, representative of the data collected in our photocrystallographic experiments.

Dataset 1 is a set of five short 21° scans (subsets 1–5) consisting of 21 blocks of ten ON/OFF frame pairs collected at 1° φ spacing. All five scans have the exact same φ range. Such scans are typically acquired in order to test crystal response to various laser powers or pump–probe delay times. The former was tested in the case presented here.

Dataset 2 consists of 90 OFF frames, routinely collected at 1° φ spacing in order to assess general crystal quality. It is also a typical example of a standard crystallographic dataset.

The datasets were collected for crystals of two solvates of Cu(I) organometallic complexes listed in Table 1. All data were collected at the 14-ID beamline of the BioCARS station at the Advanced Photon Source with an undulator setting of 15 keV and a MARCCD-165 CCD detector at fixed position, used in several of our photocrystallographic Laue experiments.

Datasets were processed in parallel with the *LaueUtil* and *LaueGui* integration software, the latter utilizing the seed-skewness method (Bolotovskiy *et al.*, 1995; Bolotovskiy & Coppens, 1997; Coppens *et al.*, 2010).

3.1. Background reconstruction for a complete dataset

Background mean values obtained for dataset 2 with the constant-fraction method are presented in Fig. 6. These values correctly reconstruct the beamstop shadow as well as the increased background at low scattering angles and a slight conical shadow of the copper mount, which partly obstructs the X-ray beam. The background decreases at higher scattering angles. Also visible are differences between the background on the four quadrants of the detector. The background estimate even in the constant-fraction approach will correct for variations in noise intensity of the different detector regions and can serve as a diagnostic tool for the characteristics of the detector.

Table 1
Description of test datasets.

1: complex $[\text{Cu}^{\text{I}}(\text{phen})(\text{PPh}_3)_2][\text{BF}_4][\text{EtOH}]$; 2: complex $[\text{Cu}^{\text{I}}(\text{phen})(\text{PPh}_3)_2][\text{BF}_4]$.

Dataset	Crystallographic system	Space group	Cell parameters (\AA , $^\circ$)	Temperature (K)	ON/OFF data
1	Monoclinic	$P2_1/n$	12.0520 (10), 21.1930 (18), 17.2507 (14), 90, 93.952 (2), 90	180	YES
2	Triclinic	$P\bar{1}$	12.8340 (11), 17.4719 (15), 19.3926 (17), 106.466 (2), 99.421 (2), 95.440 (2)	180	NO

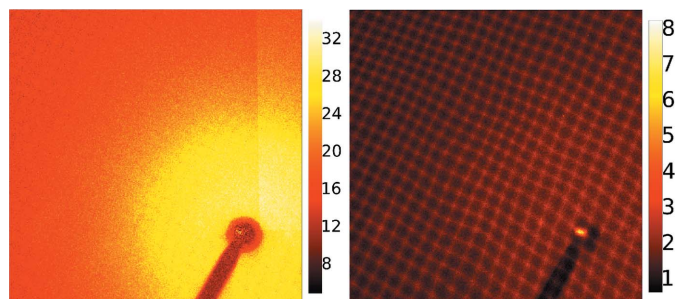


Figure 6

The background mean values (left) and standard deviations (right) around the detector center obtained using the constant-fraction method for dataset 2 and mapped on the detector plane. A Moiré pattern is exposed in the visualization of the standard deviation. A magnified view containing a 600×600 pixel subset close to the detector center is presented for the sake of clarity.

The background variance for each pixel is similarly presented in Fig. 6. In addition to the radial dependence observed for the mean or median and the difference between separate CCD quadrants, the statistic reproduces the Moiré pattern predicted for the variance of the background pixel values for CCD detectors with optical taper, as described by Waterman & Evans (2010). As noted in their work, this effect can significantly bias integration routines utilizing profile fitting.

Comparison of the background estimates with the constant-fraction and K–W approaches can be best illustrated for any subset of dataset 1. These subsets have a relatively short angular range and a block of repetitive frames at each φ angle. As a result, the contributions from the reflections constitute a significant part of the counts for certain pixels. Application of the constant-fraction approach in such a case leads to ‘spot-like’ contamination on the reconstructed background, as evident in Fig. 7(a). The more sophisticated K–W method effectively reconstructs the proper background (Fig. 7b). The difference in detection of outliers is shown for a pixel located in the spot-like feature [Figs. 7(c) and 7(d)].

3.2. Statistical distribution of the background intensities

A Poisson distribution of the background signal is often assumed in X-ray data integration algorithms to calculate standard deviations (Bolotovskiy *et al.*, 1995). The sources of background noise result from specifics of the detector design, as well as external sources such as diffuse scattering by air. Assuming that the fluctuations of the noise are independent of

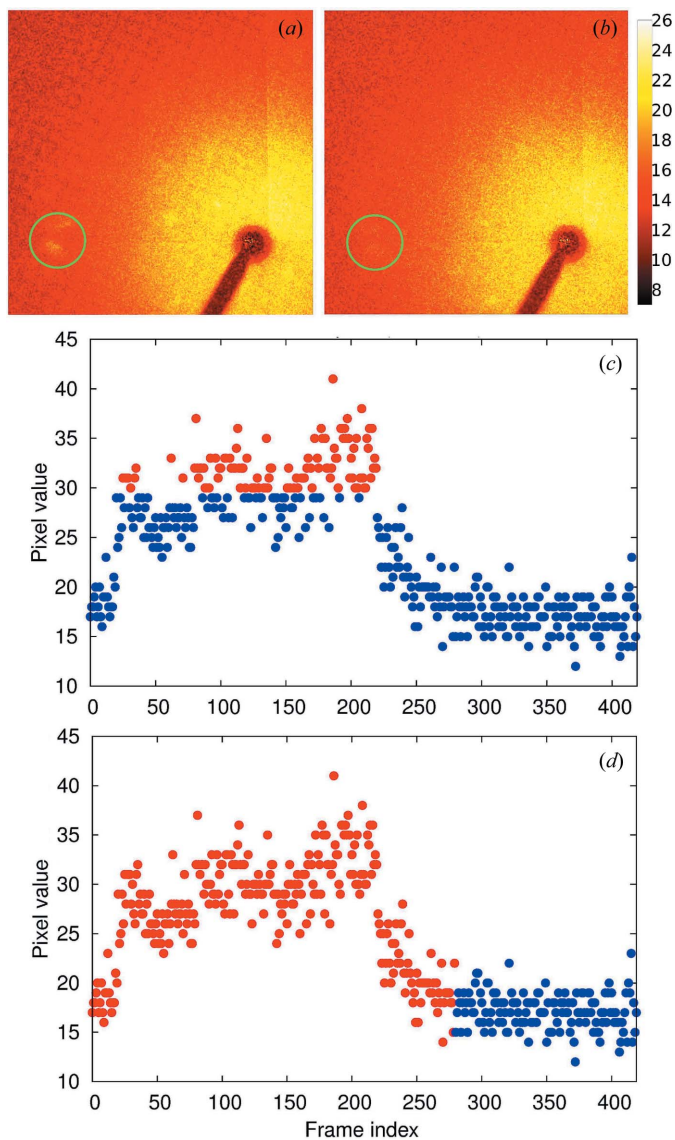


Figure 7 Background obtained using the constant-fraction method (a) and (c), and the K-W method (b) and (d). (a) and (b) represent the background mean on the fragment of the detector, while (c) and (d) represent the pixel (679, 1039) value for each consecutive frame of dataset 1 (subset 1). The location of the pixel is specified by a green circle. Blue dots correspond to values identified by each method as background contributions, and red ones as outliers.

time and there is no Bragg signal, the average event rate for each pixel unambiguously belonging to the background will be constant, and simple counting statistics should be applicable. In order to verify the nature of the background signal distribution in the current case, the statistical distributions of the pixel background values in dataset 2 were examined after the *LaueUtil* processing with the constant-fraction approach. The background samples tend to exhibit a symmetrical almost-Gaussian distribution (Fig. 8). The assumption of a Poisson distribution can be tested by examining the relation between the mean values of the background intensities and the corresponding variances, selecting only pixels which are not part of spots on any of the frames. Fig. 9 shows that for almost all of these pixels the estimated intensity variances are smaller than

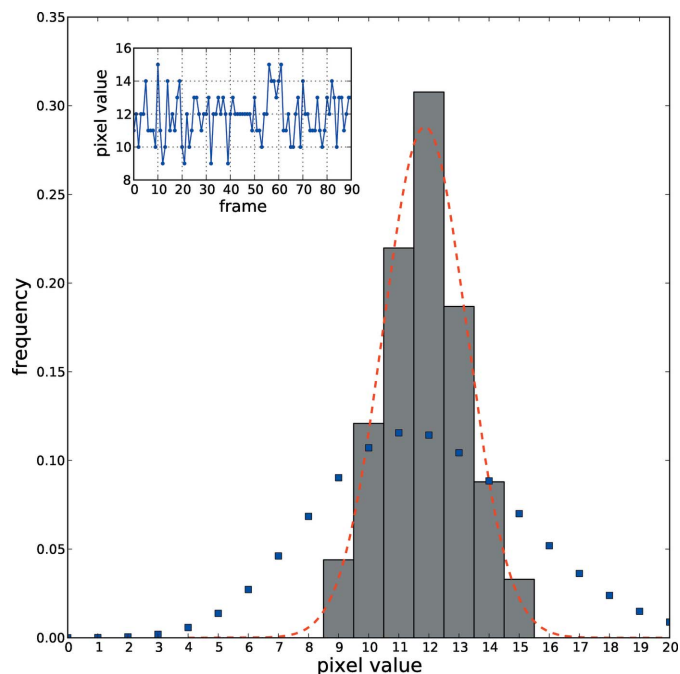


Figure 8 A histogram of background pixel values for a selected background pixel (310, 400) from dataset 2. (The inset presents pixel values in the order in which they were acquired.) The experimental distribution is plotted by a dashed red line, and the best-fitting Poisson distribution is presented by a dotted blue line.

their estimated intensity means. There is no obvious dependence of the variances on the average intensities, whereas the Poisson law implies the definite relationship of $\sigma^2(I_{\text{background}}) = \langle I \rangle_{\text{background}}$. A possible explanation is that the numerical data collected on raw frames are not necessarily an exact representation of events at the detector surface (Waterman & Evans, 2010). The information from the detector is processed to convert optical signals into electronic ones and enhanced prior to storage. This preliminary step can lead to non-

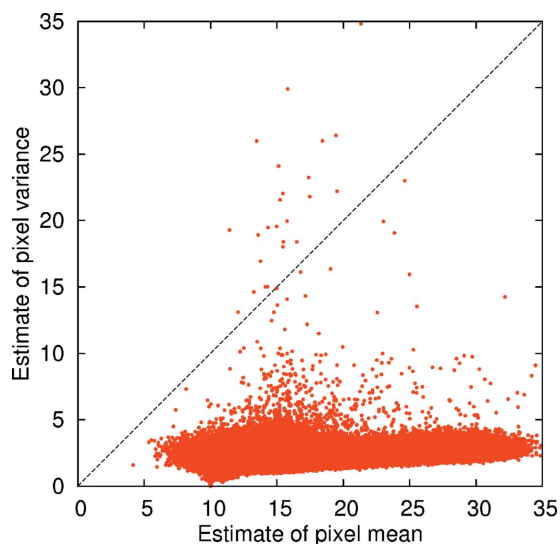


Figure 9 Variance of the background for each pixel of dataset 2 plotted against its estimated mean value. The dashed line has a 45° slope.

Poisson-distributed numerical data, depending on the detector specifications and setting.

3.3. Comparison with the seed-skewness integration

Reflection intensities resulting from *LaueGui* software (Messerschmidt & Tschentscher, 2008; Peters, 2003) were compared with the outcome of the constant-fraction *LaueUtil* results on dataset 2 and K–W results for dataset 1. In both instances a simple linear relationship exists between the intensities processed by the two methods, which therefore in principle should not affect the response ratios. The intensities from the *LaueGui* method are systematically lower than those obtained with the current approach. Fig. 10 illustrates the correlation of intensities integrated using *LaueUtil* and *LaueGui*. A magnification of the low-intensity range ($I \leq$

10000) is also plotted and confirms the linear relation between the *LaueUtil* and *LaueGui* intensities. A major source of the discrepancy between the two groups of intensities is due to the difference in the background calculations. Fig. 11 illustrates the discrepancies between the two methods. Each reflection is represented by a mark colored as a function of the differences between the average background intensities (normalized per-pixel) estimated by the *LaueUtil* and *LaueGui* programs. In a first approximation, the closer the reflection to the beam center the more significant the difference between the two methods. The *LaueUtil* algorithm provides a lower estimate of the background in the majority of cases. This difference of average background can be significant for reflections less than half way from the beam center to the edge of the frame. The

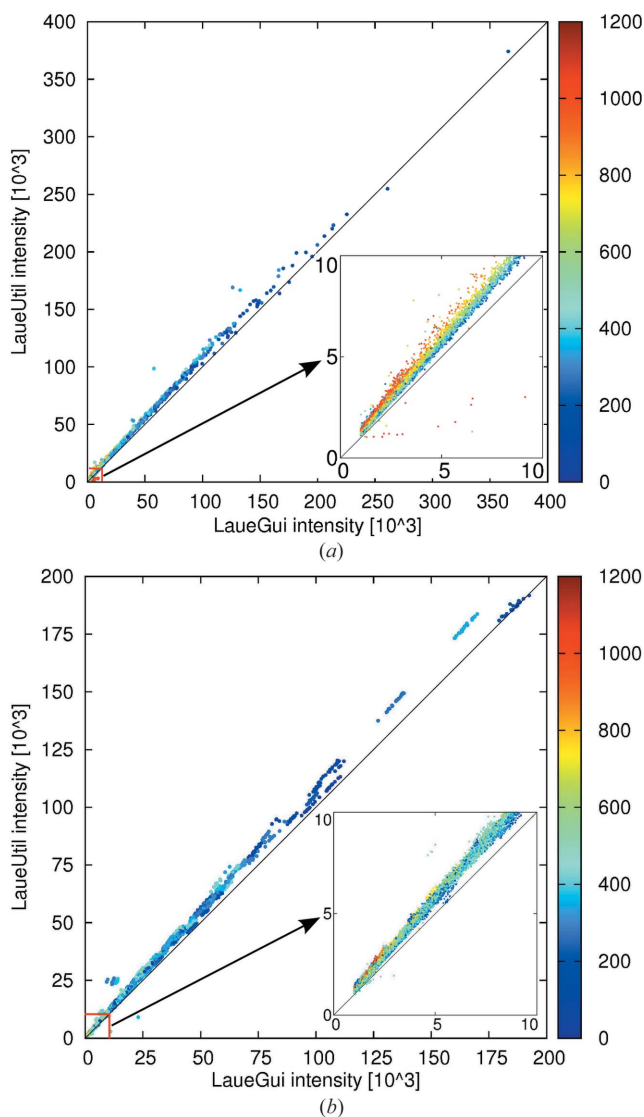


Figure 10 Correlation plot of intensities from (a) dataset 2 and (b) dataset 1 (subset 1), integrated by *LaueUtil* and *LaueGui*. Intensity dots are colored according to the distance of the spot from the beam center on the detector surface. Distances are given in pixel side length. The insets illustrate the low-intensity region.

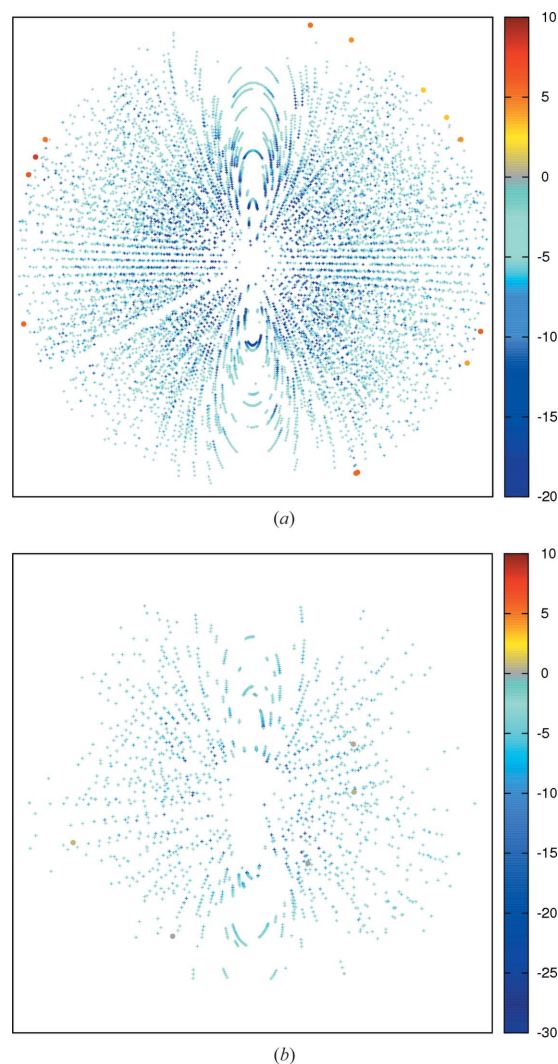


Figure 11 Differences between the average background intensities (normalized per-pixel) processed with *LaueUtil* and the *LaueGui* integration methods illustrated at the reflection's position on the detector (a) on the dark dataset 2 and (b) on the ON/OFF dataset 1 (subset 1). Only intensities from the first frame of each block are shown for the purpose of clarity. Negative differences are marked as crosses and positive differences as circles. The greatest negative discrepancies can be observed for a few strong low-order reflections. The color scale was truncated for negative values at the level of -20 (a) or -30 (b).

maximum difference, excluding the edge of the detector, is about 79%.

The discrepancy likely results from the lack of accuracy of the footprint definition in *LaueGui*. Part of the reflection tail is sometimes included in the background count, as illustrated in Fig. 12, which shows the relation between the normalized background difference and the surface of the reflection footprint. Almost all spots with background differences larger than 20 are very strong and located in the vicinity of the beam center, with the *LaueGui* background being larger. Only a few reflections on the edge of the detector have opposite background differences. The background estimate method used in *LaueGui* allows inclusion of pixels located outside the actual active detector area in the background calculation, which explains this background underestimate relative to *LaueUtil*. In all instances the background estimate in the *LaueUtil* method appears more reliable.

3.4. Prompt signal analysis during pump–probe experiments

No information on the unit-cell parameters or crystal orientation is required for data integration with the *LaueUtil* tool. As a result, data can be integrated without the time-consuming indexing of the Laue pattern. It allows prompt evaluation of the light-induced signal. The intensities for any spot on any frame can be analyzed (plotted or otherwise processed) to ascertain whether or not there are systematic ON versus OFF differences. The method is especially advantageous when unit-cell parameters are not known, or when the sample is twinned. The only necessary condition is that cell parameters do not change significantly upon laser exposure. This can be immediately verified by analysis of the spots positions on consecutive ON and OFF frames. Table 2 presents the experimental ON to OFF ratios for selected spots

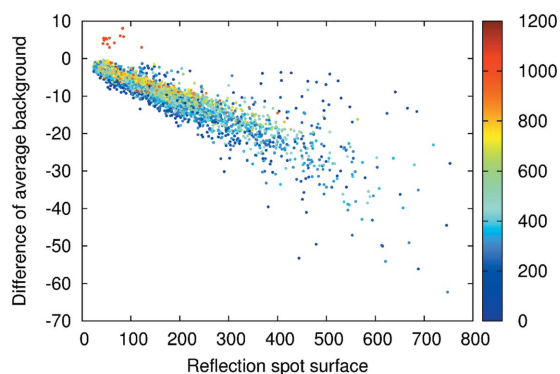


Figure 12 Differences between the average background yielded by *LaueUtil* and the *LaueGui* integration method with respect to the spot area as estimated in *LaueUtil* (dataset 2). Intensity dots are colored according to the distance of the spot from the beam center on the detector surface.

Table 2

Examples of reflections with I_{ON}/I_{OFF} experimental ratios (R) and their variation with the laser power applied during ON exposure.

The X-ray to laser pulse delay time was 100 ps.

	Laser power ($\mu\text{J pulse}^{-1}$)				
	24.7	37.1	49.5	61.9	74.2
Spot position (1102, 984)					
Label	221	217	216	207	214
R_{ave}	1.18	1.44	1.4	1.2	1.08
$\sigma(R)$	0.1	0.12	0.12	0.07	0.05
$\eta/\sigma(\eta)$	1.89	3.8	3.3	2.7	1.66
Spot position (1162, 1070)					
Label	248	244	244	228	209
R_{ave}	1.07	1.22	1.18	1.13	1.07
$\sigma(R)$	0.03	0.07	0.08	0.07	0.06
$ (R - 1) /\sigma(R)$	2.39	3.1	2.3	1.9	1.17

in the five subsets of dataset 1, which differ in applied laser power. Data were integrated using a common mask for each 20 frames in a block. Ratios of intensities were obtained for all ON/OFF frame pairs, their averages calculated and standard deviations estimated from the ten repeated measurements. Fig. 13 shows a reasonable agreement between ratios obtained applying *LaueUtil* and *LaueGui* software on dataset 1 (subset 1).

4. Conclusions

A new approach to Laue X-ray data integration from CCD detectors is presented. The method uses simple statistical tools for identification of the background values for a given pixel on all frames in a scan.

Two particular approaches are described, their applicability depending on the X-ray measurement strategy. The constant-fraction approach is best suited for conventional data collection strategies, in which the crystal orientation and the

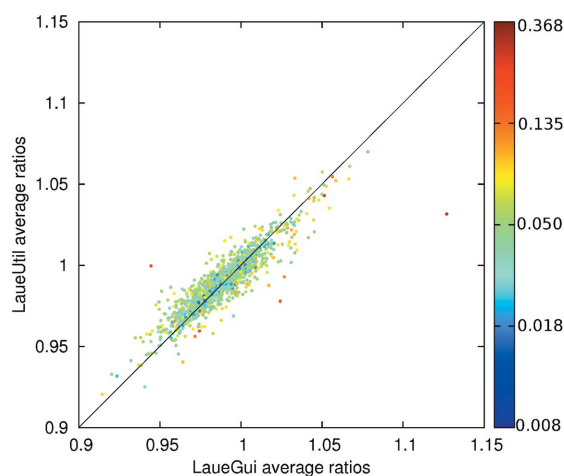


Figure 13 Correlation plot of ratios averaged within constant φ angle calculated from dataset 1 (subset 1) using *LaueUtil* and *LaueGui*. The dots are colored according to the largest standard deviation in the two methods, $\max(\sigma_{LaueUtil}; \sigma_{LaueGui})$, the *LaueGui* standard deviation being generally larger.

resulting pattern are being changed from frame to frame. In the photocrystallographic experiments, the total range of crystal orientations may be limited. However, at each crystal orientation a batch of frames can be measured. In such case the K–W method is more suitable and yields superior results in terms of identification of the pixels belonging to the background and therefore leads to more reliable values of the Bragg intensities.

As the method is strictly based on statistical analysis of the pixel values, it does not depend on a data indexing routine, and thus allows monitoring X-ray intensities and light-induced changes even when cell parameters are not known. In the case of ultra-fast photocrystallographic experiments in which cell parameters, and hence reflection positions, do not vary significantly between ON and OFF exposures, the response ratios can be calculated promptly. As no integration box is to be defined or profile is to be fitted, the method is also more suitable for dealing with reflections of elongated shape, as often observed in such experiments.

In its current implementation the method presented here is applicable mainly to data from photocrystallographic synchrotron experiments, although conventional data can also be processed with the constant-fraction approach. The only limitation of the method is the stability of the background levels during the experiments, depending on the X-ray source stability and diffuse scattering from the crystal support. When the stability criterion is fulfilled, the method yields prompt response ratios for subsequent photocrystallographic analysis.

APPENDIX A The K–W test

The K–W test is a non-parametric method for testing whether two or more independent samples of values share a similar population distribution. In contrast to the one-way analysis of variance, ANOVA, the K–W test does not require an assumption about the nature of sample distribution, such as normality. The null hypothesis of this test is that the populations from which the samples originate share the same probability distribution. Like many other non-parametric tests the K–W test is based on the calculation of sample ranks. The test consists of five steps.

Let us assume a series of M samples with N_i the number of values in the i th sample and N the total number of values in this set:

(a) Sort all values from all samples together, and assign a rank to each value from 1 to N . If there are subsets of tied values, the average value of their ranks must be calculated and assigned to all of them.

(b) For each i th sample, calculate its average rank \bar{R}_i ,

$$\bar{R}_i = \frac{\sum_{k=1}^{N_i} r_{(i,k)}}{N_i}, \quad (4)$$

where $r_{(i,k)}$ is the rank of the k th value of the i th sample.

(c) Deduce the statistical test coefficient K defined as

$$K = (N - 1) \frac{\sum_{i=1}^M N_i (\bar{R}_i - \bar{R})^2}{\sum_{i=1}^M \sum_{j=1}^{N_i} [r_{(i,j)} - \bar{R}]^2}, \quad (5)$$

with $\bar{R} = (N + 1)/2$ the average rank of the full set.

K can be rewritten as follows,

$$K = \left[\frac{12}{N(N + 1)} \sum_{i=1}^M N_i \bar{R}_i^2 \right] - 3(N + 1). \quad (6)$$

(d) If there are some tied values in the total set, divide the K value by a tie-correction factor T to obtain $K_{\text{corrected}}$. This factor is given by

$$T = 1 - \frac{\sum_{k=1}^{L} p_k^3 - p_k}{N^3 - N}, \quad (7)$$

where L is the number of tied value sequences and, for each sequence k , p_k is its size.

(e) Approximate $K_{\text{corrected}}$ by a χ^2 distribution with $M - 1$ degrees of freedom and calculate the corresponding probability p -value $\text{Pr}(\chi^2_{M-1} \geq K)$. This approximation is reasonable if the samples are larger than 5. Depending on the desired α level, the null hypothesis is rejected if the p -value $\leq \alpha$.

Support of this work by the National Science Foundation (CHE0843922) is gratefully acknowledged. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant RR007707. The Advanced Photon Source is supported by the US Department of Energy, Office of Basic Energy Sciences, under Contract No.W-31-109-ENG-38.

References

- Anderson, S., Srajer, V., Pahl, R., Rajagopal, S., Schotte, F., Anfinrud, P., Wulff, M. & Moffat, K. (2004). *Structure*, **12**, 1039–1045.
- Benedict, J. B., Makal, A., Sokolow, J. D., Trzop, E., Scheins, S., Henning, R., Graber, T. & Coppens, P. (2011). *Chem. Commun.* **47**, 1704–1706.
- Bolotovskiy, R. & Coppens, P. (1997). *J. Appl. Cryst.* **30**, 244–253.
- Bolotovskiy, R., White, M. A., Darovsky, A. & Coppens, P. (1995). *J. Appl. Cryst.* **28**, 86–95.
- Collet, E., Moisan, N., Balde, C., Bertoni, R., Trzop, E., Laulhe, C., Lorenc, M., Servol, M., Cailleau, H., Tissot, A., Boillot, M.-L., Graber, T., Henning, R., Coppens, P. & Cointe, M. B.-L. (2012). *Phys. Chem. Chem. Phys.* **14**, 6192–6199.
- Coppens, P., Benedict, J., Messerschmidt, M., Novozhilova, I., Graber, T., Chen, Y.-S., Vorontsov, I., Scheins, S. & Zheng, S.-L. (2010). *Acta Cryst.* **A66**, 179–188.
- Coppens, P., Pitak, M., Gembicky, M., Messerschmidt, M., Scheins, S., Benedict, J., Adachi, S., Sato, T., Nozawa, S., Ichiiyanagi, K., Chollet, M. & Koshihara, S. (2009). *J. Synchrotron Rad.* **16**, 226–230.
- Corder, G. W. & Foreman, D. I. (2009). *Nonparametric Statistics for Non-Statisticians*. New York: John Wiley and Sons.
- Helliwell, J. R., Habash, J., Cruickshank, D. W. J., Harding, M. M., Greenhough, T. J., Campbell, J. W., Clifton, I. J., Elder, M., Machin, P. A., Papiz, M. Z. & Zurek, S. (1989). *J. Appl. Cryst.* **22**, 483–497.
- Kalinowski, J. A., Makal, A. & Coppens, P. (2011). *J. Appl. Cryst.* **44**, 1182–1189.
- Kamiński, R., Graber, T., Benedict, J. B., Henning, R., Chen, Y.-S., Scheins, S., Messerschmidt, M. & Coppens, P. (2010). *J. Synchrotron Rad.* **17**, 479–485.

- Kruskal, W. H. & Wallis, W. A. (1952). *J. Am. Stat. Assoc.* **47**, 583–621.
- Makal, A., Trzop, E., Sokolow, J., Kalinowski, J., Benedict, J. & Coppens, P. (2011). *Acta Cryst.* **A67**, 319–326.
- Messerschmidt, M. & Tschentscher, T. (2008). *Acta Cryst.* **A64**, C611.
- Moffat, K. (2001). *Chem. Rev.* **101**, 1569–1581.
- Peters, J. (2003). *J. Appl. Cryst.* **36**, 1475–1479.
- Pierre, S. (2003). *Morphological Image Analysis; Principles and Applications*. Berlin: Springer.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ren, Z. (2010). *Precognition User Guide*. RenZ Research Inc., Illinois, USA.
- Ren, Z., Bourgeois, D., Helliwell, J. R., Moffat, K., Šrajer, V. & Stoddard, B. L. (1999). *J. Synchrotron Rad.* **6**, 891–917.
- Ren, Z. & Moffat, K. (1995). *J. Appl. Cryst.* **28**, 482–494.
- Šrajer, V., Crosson, S., Schmidt, M., Key, J., Schotte, F., Anderson, S., Perman, B., Ren, Z., Teng, T., Bourgeois, D., Wulff, M. & Moffat, K. (2000). *J. Synchrotron Rad.* **7**, 236–244.
- The HDF Group (2010). *Hierarchical data format version 5*, <http://www.hdfgroup.org/HDF5>.
- Waterman, D. & Evans, G. (2010). *J. Appl. Cryst.* **43**, 1356–1371.