International Tables for Crystallography (2012). Vol. F, Section 11.4.10.3, p. 292.

11.4.10.3. Merging and signal validation

Symmetry-related scaled measurements $I(hkl)$ and their uncertainty estimates σ_E are used to obtain merged intensities by a standard weighted averaging formula:

$$\langle I\rangle = \frac{\sum_j I_j/\sigma_{E,j}^2}{\sum_j 1/\sigma_{E,j}^2}. \qquad (11.4.10.6)$$

This allows for calculations of validation statistics, called goodness-of-fit or normalized χ^2, for each unique index:

$$\chi^2 = \left(\frac{1}{n-1}\right)^{1/2} \sum_j \frac{(I_j - \langle I\rangle)^2}{\sigma_{E,j}^2}, \qquad (11.4.10.7)$$

where n represents the number of observations of a given unique index. This χ^2 statistic is then averaged in resolution shells or over intensity bins or batch number. If the error model accounts properly for all effects, the χ^2 statistic should fluctuate around a value of unity. If χ^2 values depart from this expectation it may indicate a number of possibilities, *e.g.* various problems at earlier stages (poorly edited beam-stop shadow, hardware failures, mistakes in processing or other source of outliers *etc.*), inadequacy of the error model or variations in the structure factors within the symmetry-related observations. The instrumental problems or mistakes in processing should be corrected. The effects that cannot be corrected may be handled by adjusting the error model. However, if the more detailed analysis eliminates the obvious source for such problems, then the most likely source of discrepancies between symmetry-related measurements results from violation of Friedel symmetry. *SCALEPACK* calculates merging statistics both for the Bijvoet pairs merged together and separately. Differences in χ^2 values between these two merging outputs are very reliable estimates of anomalous signal strength. When a more detailed analysis eliminates the obvious reasons for high χ^2 values, the most likely remaining source of error is non-isomorphism (Borek *et al.*, 2007, 2010).

11.4.11. Detector diagnostics

The *HKL* package has a number of tools that can detect possible detector or experimental setup problems (Otwinowski & Minor, 1997; Otwinowski *et al.*, 2003). Visual inspection of the image may provide only a very rough estimate of data quality. A check of the analogue-to-digital converter can provide rough diagnostics of detector electronics. Examination of the background can provide information about detector noise, especially when uncorrected images can be examined in the areas exposed to X-rays and areas where pure read-out noise can be observed. *DENZO* provides several diagnostic tools during the integration stage, as the crystallographer may observe crystal slippage, a change of unit-cell parameters or a change of the values of positional and angular χ^2 during the refinement.

Even more tools are provided at the data-scaling stage. By observing scale factors, poor crystal alignment can be detected. Other tools may help diagnose X-ray shutter malfunction, spindle-axis alignment and internal detector-alignment problems. The final inspection of outliers may again provide valuable information about detector quality. The clustering of outliers in one area of the detector may indicate a damaged surface; if most outliers are partials, it may indicate a problem with spindle backlash or shutter control. The zoom mode may be used to display the area around the outliers to identify the source of a problem: for example, the existence of a satellite crystal or single pixel spikes due to electronic failure. Sometimes, even for very strong data, a histogram of the pixel intensities may stop below the maximum valid pixel value, indicating saturation of the data-acquisition hardware or software.

11.4.12. *HKL-2000* and *HKL-3000*

DENZO and *SCALEPACK* form the numerical processing core of the *HKL* package. These programs can be used directly by editing commands in input scripts, but most of the time they are run through *HKL-2000* and its more expanded version *HKL-3000*.

The basic mode of *HKL-2000* reads in previously collected data as input and produces scaled and merged reflections as output. This use can be extended in three directions: control of the data-collection process by *HKL-2000* and *HKL-3000* (Minor *et al.*, 2002, 2006), incorporation of methods of structure solution by *HKL-3000* (Minor *et al.*, 2006), and storage of critical intermediate results in an external database (Grabowski *et al.*, 2007).

Crystallographic structure determinations encompass a wide range of project dynamics. There are many projects where all the steps are executed serially, while others involve substantial iterative improvements, where one or a few stages are repeated a number of times. *HKL-2000* and *HKL-3000* are designed to make both types of project more effective. To accomplish this, information is automatically propagated between various stages of analysis, and many necessary data transformations are performed to accommodate the interface requirements of many programs and beamline controls. At the same time, crystallography requires the experimenter to be actively engaged in decision making, depending on the nature of a particular project and types of problems encountered. The experimenter may need to assess the quality of a set of crystals, decide how to collect data sets from a chosen subset, determine the symmetry of each diffraction pattern, reassess the crystal quality based on the integration and merging steps, and then solve the structure using an appropriate method. Not all programs are fully automatic, and the experimenter may need to be involved in defining how a particular procedure should be executed. As a consequence, both *HKL-2000* and *HKL-3000* have a multiplicity of interface screens (accessed by tabs in the graphical control centre), each of them designed to control a particular process.

The versions of *HKL-2000* and *HKL-3000* that interface with data-collection systems can coordinate all parameters of the diffraction experiment. This facilitates interactive experiments in which data analysis is done online and results are automatically updated when new data are collected. In such experiments, it is possible to adjust the data-collection strategy to guarantee the desired result, particularly with regard to data completeness. The strategy takes into account limitations arising from radiation damage. Radiation damage can be estimated from past experience with similar crystals, by theoretical calculations of decay based on beam intensity, and by evaluating scale- and B-factor changes in real time.

The graphical control centre of *HKL-2000* (and *HKL-3000*) consists of three components: an internal database (optionally connected to an external one), a transition-state engine and a graphical user interface (GUI). The internal database stores all the information about data processing and data collection. It can describe not only the data already collected, but also those being collected and even those planned for collection. Each datum entered or program executed, including the data-collection interface, induces a change in the database by the transition-state engine. One of the main functions of the GUI is input to and

references