How to obtain the best data from XDS

The procedures for processing data with XDS have been described [4, 5] and are not repeated here. Instead, based on first hand experiences when processing datasets from my own group and helping others with their challenging datasets, I focus on those steps that are critical for data quality. For simplicity, we assume that a given dataset can be indexed in the correct space group. The overarching rules for data processing, in the order of their importance, are that

1. Sources of systematic error should be excluded if possible.
2. The impact of any remaining sources of systematic error on the data should be minimized.
3. The random error should be minimized.
4. The completeness of the data should be maximized.

Experience shows that goals (1), (2), and (3) are not conflicting, but can be met with the same set of processing parameters. Topic (4), however, requires a compromise. For instance, rejecting the final frames of a dataset, in order to minimize the impact of radiation damage, will reduce the completeness, or at least the multiplicity of the data. Likewise, too generous masking of shadowed detector regions might lead to rejection of well-measured reflections.

Analysis of the information provided by XDS (see below) may lead to deeper insight about the data collection experiment itself. Designing and evaluating an experiment is a genuinely scientific approach and can and should not be left to automatic procedures.

The goal of data processing is to best parameterize the data collection experiment. If the data processing is repeated with changed parameters, the magnitude of the systematic error should be monitored, using ISa. By optimizing (maximizing) ISa, indicators of data precision are usually enhanced along the way, mainly because the location and shape of the reflections on the frames can be predicted more accurately. Generally, if the systematic error in the data is reduced, the noise associated with it is converted to signal. In case of doubt about any specific aspect of data processing, the parameter value that maximizes ISa is usually the correct one.

General Approach

To discover problems associated with data processing, it is essential that in particular the files FRAME.cbf, INTEGRATE.LP, XDS_ASCII.HKL and CORRECT.LP are analyzed.

FRAME.cbf should be inspected (Fig. 1) to find out whether spot shapes are regular, or whether there is indication of splitting and multiple lattices. Irregular and split spots indicate problems in crystal growth or handling, and always compromise data quality due to higher random noise (because spots extend over more pixels) and higher systematic error (because the reflection profiles differ from the average). Furthermore, FRAME.cbf allows finding out if predicted and observed diffraction patterns match. If they do not, the space group or geometric parameters may be wrong which may either prevent data processing from giving useful data, or may lead to downstream problems in phasing and refinement. However, this is beyond the scope of this article. Finally, FRAME.cbf, which visualizes the last frame processed by INTEGRATE, should be checked for the presence of ice rings (see below).

The tables in INTEGRATE.LP should be inspected for jumps or large changes in frame-wise parameters like scale factors, mosaicity, beam divergence, or refined parameters like unit cell parameters, direct beam position, and distance (Fig. 2). Such changes should be understood as indicating a potential

source of systematic error. Scale factor jumps should be brought to the attention of the beamline manager; the other changes point to problems concerning the experiment parameterization, like crystal decay or slippage, and should trigger reprocessing after change of parameters like DATA_RANGE, DELPHI, and REFINE(INTEGRATE) until no further improvement can be obtained.

CORRECT.LP, among other statistics, reports on systematic error (ISa) and the precision of unmerged and merged intensities (Rmeas and CC1/2, respectively). It needs to be consulted to monitor the success of changes to parameters in XDS.INP, and of changes to the file XPARM.XDS describing the geometry of the experiment, which is used by INTEGRATE. It is useful to plot the quantities reported in CORRECT.LP as a function of resolution, and as a function of the upper frame range (Fig. 3).

XDSSTAT, a program that analyzes XDS_ASCII.HKL, should be run and its output diverted to XDSSTAT.LP, to be visualized with a plotting program. In addition, the control images written by XDSSTAT offer a graphical way to inspect the projection of several quantities on the detector surface, most notably R-values, scale factors, and misfits (outliers identified during scaling) (Fig. 4).

Better processing may lead to a lower number of reflections rejected during scaling. A guideline for the acceptable number of outliers is the following: provided that the average multiplicity is 2 or higher, up to 1 % of the observations (the default that XDS employs) may be rejected as outliers. If the percentage is higher, the reason for this should be investigated, first by inspecting “misfits. pck” as obtained from XDSSTAT. If “misfits.pck” shows concentric rings of outliers, the high percentage appears justified, but the options for treating ice rings (see above) should be evaluated. Second, if specific frames have many outliers, as shown by XDSSTAT.LP, then these frames should possibly be omitted from processing, and the reason why they delivered outlier data should be investigated.

Shaded areas of the detector

Several parameters have to be manually set before the integration step of XDS to mask shaded detector areas. Since the keywords TRUSTED_REGION, UNTRUSTED_RECTANGLE, UNTRUSTED_ELLIPSE, and UNTRUSTED_QUADRILATERAL are not evaluated by the INTEGRATE and CORRECT steps, they have to be specified earlier, namely, for the INIT or DEFPIX steps. This requires graphical inspection of at least a single data frame.

The low resolution limit of the data should be set such that the shadow of the beam stop is completely excluded, using INCLUDE_RESOLUTION_RANGE. Contrary to the keywords mentioned before, this keyword can be specified at a later step (CORRECT). If the lower resolution limit is too optimistic (i.e., too low), many rejections and high $\chi 2$ values result in the low-resolution shell of the first statistics table available from CORRECT. If this is indeed observed, the lower resolution limit should be raised.

Ice rings, ice reflections and "Aliens"

Single ice reflections, which fall onto a predicted spot position, are usually automatically excluded by the default outlier rejection mechanisms in CORRECT, either because their symmetry does not obey that of the macromolecular crystal, or because they are much stronger (“aliens” in CORRECT.LP) than the other reflections in their resolution range. The positions of rejected reflections can be visualized by inspecting the file “misfits.pck” using XDS-Viewer or adxv.

Strong ice rings should be manually excluded using EXCLUDE_RESOLUTION_RANGE; weak ice rings

should be left to the automatic mechanisms for outlier rejection, because that results in higher completeness. To decide whether an ice ring should be considered strong or weak, the user should inspect the first statistics table in CORRECT.LP ("STANDARD ERROR OF REFLECTION INTENSITIES AS FUNCTION OF RESOLUTION"); ice rings are easily identified by a large number of rejections at resolution values near those of ice reflections (3.897, 3.669, 3.441, 2.671, 2.249, 2.072, 1.948, 1.918, 1.883, 1.721 Å for hexagonal ice, the form most often encountered). If the $\chi 2$ and R-values in these resolution ranges are much higher than in the other ranges, the user should consider to reject the ice rings, using EXCLUDE_RESOLUTION_RANGE. This should also be done if the control image "scales.pck" (written by XDSSTAT) shows a significant deviation of scale factors from the value of 100 % at resolution values close to those of ice rings, or if "rf.pck" shows high R-values.

At very high resolution, in shells with mean intensity approaching zero, the "alien" identification algorithm sometimes rejects very many reflections when using its default value of REJECT_ALIEN = 20. If this happens, the default should be raised to, say, 100 to prevent this from happening.

Specific procedure for optimizing data quality

Since the defaults in XDS.INP are carefully chosen and XDS has robust routines, very good data are usually obtained from a single processing run, in particular from good crystals. However, in case of difficult or very important datasets, the user may want to try and optimize the data processing parameters. This can be understood as minimizing or eliminating the impact of systematic errors introduced by the data processing step.

Three simple options should be tried:

- The globally optimized geometric parameter file GXPARM.XDS (obtained from CORRECT) may be used for another run of INTEGRATE and CORRECT. This operation may reduce the systematic error which arises due to inaccurate geometric parameters. It requires that the values of "STANDARD DEVIATION OF SPOT POSITION" and "STANDARD DEVIATION OF SPINDLE POSITION" in CORRECT.LP are about as high as the corresponding values printed out multiple times in INTEGRATE.LP, for each batch of frames. This option is particularly successful if the SPOT_RANGE for COLSPOT was chosen significantly smaller than the DATA_RANGE, because in that case the accuracy of geometric parameters from IDXREF may not be optimal.
- In XDS.INP, the averages of the refined profile-fitting parameters as printed out in INTEGRATE.LP, may be specified for another run of INTEGRATE and CORRECT. Essentially, this option attempts to minimize the error associated with poorly determined spot profiles. This is most effective if there are few strong reflections and/or large frame-to-frame variations between estimates of SIGMAR (mosaicity) and SIGMAB (beam divergence) as listed in INTEGRATE.LP.
- In XDS.INP, one may specify the keyword REFINE (INTEGRATE) with fewer (e.g., only ORIENTATION) or no geometric parameters, instead of the default parameters DISTANCE BEAM ORIENTATION CELL. This approach, which also requires at least one more run of INTEGRATE and CORRECT, is most efficient if the refined parameters, as observed in previous INTEGRATE runs, vary randomly around a mean value. Of course, preventing refinement of a parameter is not the correct approach if its change is required to achieve a better fit between observed and predicted reflection pattern. If removal of certain geometric parameters from geometry refinement in INTEGRATE indeed improves ISa, this indicates that the geometry refinement is not well enough determined to improve them beyond those obtained by the global refinement in IDXREF or CORRECT. This option thus reduces the systematic error due to poorly determined

geometry. An alternative to switching refinement off is to specify a larger DELPHI than the default (5°).

Ideally, each of the three options (a–c) should be tried separately. Those options that improve ISa can then be tried in combination, and the optimization procedure may be iterated as long as there is significant improvement (of, say, a few percent) in ISa.

In my experience, optimization may lead to significantly better data, as shown by improved high-resolution CC1/2 and improved merging with other datasets, particularly for poor datasets with high mosaicity and/or strong anisotropy.

Don't

Two possible ways of misusing XDS parameters should be mentioned.

First, it may be tempting to increase the number of outliers and thereby to “improve” (or rather “beautify”) the numerical values of quality indicators. This could in principle be achieved by lowering the WFAC1 parameter below its default of 1. However, the goal of data processing is to produce an accurate set of intensities for downstream calculations, not a set of statistical indicators that have been artificially “massaged.” Experience shows that reducing WFAC1 below its default almost always results in data with worse accuracy; conversely, raising WFAC1 may sometimes be a way to prevent too many observations to be rejected as outliers. Only if there is additional evidence for the validity of reducing WFAC1 should this quantity be lowered.

The second way to misuse XDS is to consider all the reflections listed as “aliens” in CORRECT.LP as outliers, and to place them into the file REMOVE.HKL to reject them in another CORRECT run. This is not appropriate; it should only be done if there is additional evidence that these reflections are indeed outliers. Such evidence could be the fact that the “aliens” occur at resolution values corresponding to ice reflections (see above).

From:
https://bsi.inscog.eu/ - **BSI wiki**

Permanent link:
https://bsi.inscog.eu/doku.php?id=crystallography:processing:xds_detailed_procedure&rev=1463581162

Last update: **2023/11/01 20:17**