IFSC/CCP4

Macromolecular Crystallography School 2018

Data Reduction with POINTLESS, AIMLESS and TRUNCATE

Andrew GW Leslie, MRC LMB, Cambridge, UK

Integration

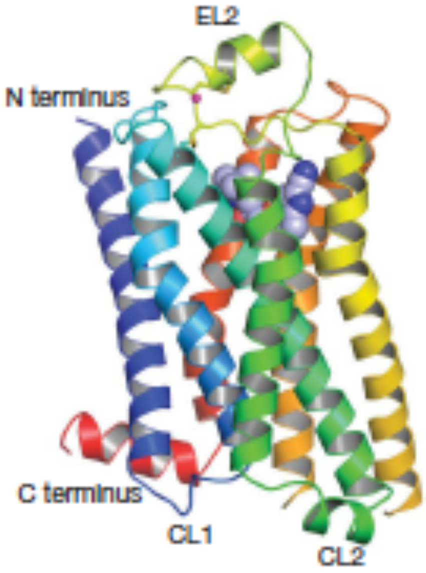
Images: 1-29

Parameter	Value	Fix
Beam x	119.51	☐
Beam y	120.02	☐
Distance	249.50	☐
Y-scale	1.0003	☐
Tilt	0.02	☐
Twist	0.11	☐
Tangential offset	-0.040	☐
Radial offset	-0.040	☐
RMS residual	0.043	
RMS res. (central)	0.031	
RMS res. (weighted)	0.540	

Parameter	Value	Fix
φ(x)	-0.04	
φ(y)	-0.11	
φ(z)	0.02	
a	58.43	☑
b	58.43	☑
c	156.01	☑
α	90.00	☑
β	90.00	☑
γ	120.00	☑
Mosaicity	1.024	☐

Parameter	Full	Partial
$\langle I/\sigma(I)\rangle$ (prf)	0.00	10.70
$\langle I/\sigma(I)\rangle$ (sum)	0.00	10.30
Reflections	0	728
$\langle I/\sigma(I)\rangle$ HR (prf)	0.00	4.80
$\langle I/\sigma(I)\rangle$ HR (sum)	0.00	4.80
Reflections HR	0	133

???



Starting point:
Unmerged (MTZ) file with a list of intensities and estimated standard deviations of all measured reflections on the processed images.

Required:

- A list of reflections reduced to the correct asymmetric unit.
- Partially recorded reflections have to be summed (if using 2D integration). Intensity estimates of symmetry related reflections should be averaged after appropriate scaling.
- The structure factor amplitude of each reflection is also required and estimates of the anomalous differences.
- Improved standard deviation estimates should also be obtained, and statistical tests can be applied to detect the presence of merohedral twinning.

Within the CCP4 suite, these tasks are carried out by POINTLESS (space group determination), AIMLESS (partial intensity summation, scaling and merging, improved standard deviation estimates) and TRUNCATE (reducing intensities to amplitudes, statistical tests for twinning).

Protocol for space group determination (program *POINTLESS* by Phil Evans)

Pointless reads the MTZ file output by MOSFLM (or other integration programs DIALS, XDS) before any scaling or averaging. It can be run on a (very) incomplete dataset (eg 5° of data).

1. From the unit cell dimensions, find the highest compatible lattice symmetry (within a tolerance). This may be higher than the symmetry used when integrating the data. **The input symmetry is ignored**.
2. Score each symmetry element (rotation) belonging to lattice symmetry using all pairs of observations related by that element.
3. Score combinations of symmetry elements for all possible sub-groups (Laue groups) of lattice symmetry group.
4. Score possible space groups from axial systematic absences (may fail for very incomplete datasets).

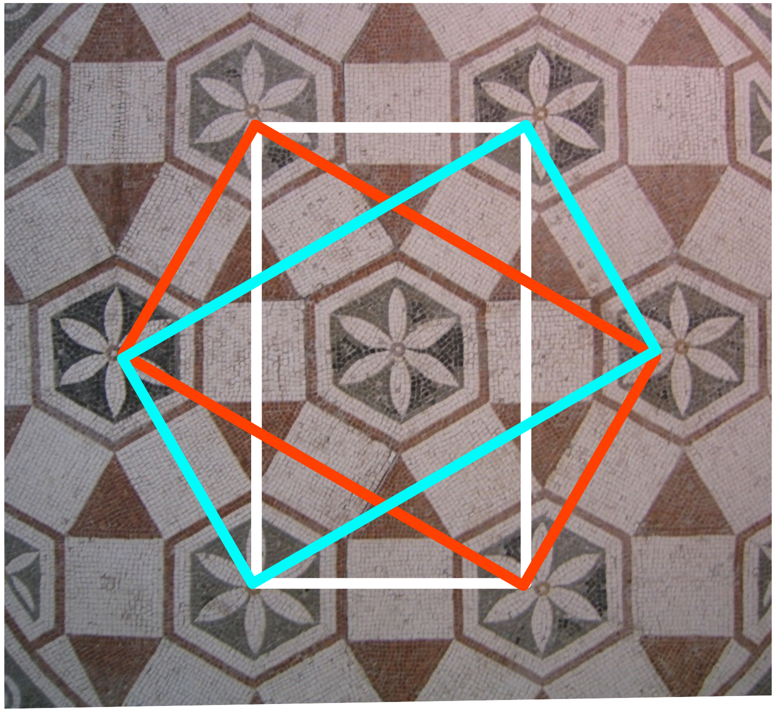
Scoring functions for rotational symmetry are based on ***correlation coefficients***, since these are relatively independent of the unknown scales. R_{meas} values are also calculated

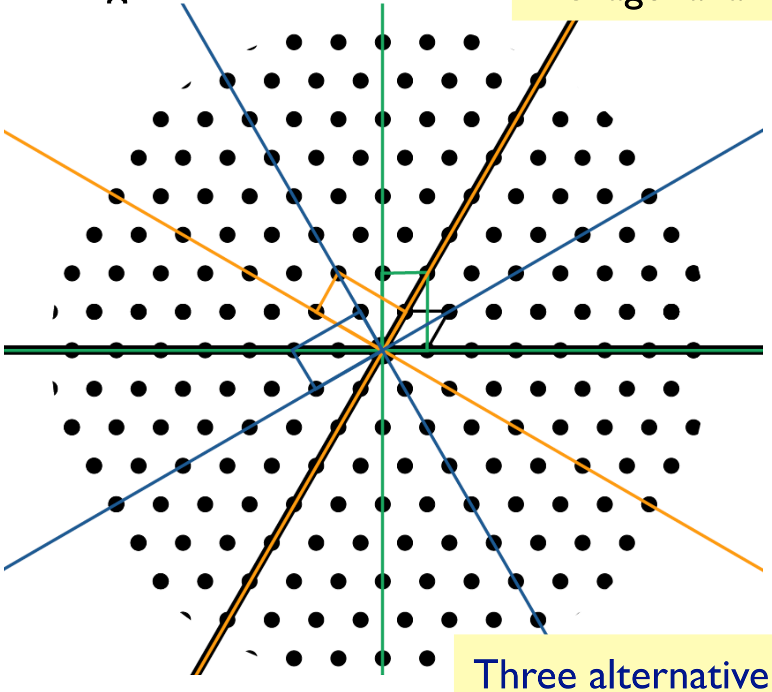
A confusing case in C222 with a pseudo-hexagonal cell:

Unit cell 74.72 129.22 184.25 90 90 90

This has $b \approx \sqrt{3}\,a$ so can also be indexed on a hexagonal lattice, and the indexing would probably select a trigonal/hexagonal solution as it only considers the shape of the unit cell.

Conversely, a hexagonal lattice may be indexed as C222 in three distinct ways, so there is a 2 in 3 chance of the indexing program choosing the wrong one if data from more than one crystal is being processed and merged.





Hexagonal axes (black)

Three alternative C-centred orthorhombic Lattices (coloured)

Results from running POINTLESS

POINTLESS will run in exactly the same way whether the data has been integrated as orthorhombic or hexagonal.

Score each symmetry operator in point group P622

"Likelihood"

Correlation coefficient on E^2

Z-score(CC)

Rfactor (multiplicity weighted)

```
Nelmt  Lklhd  Z-cc    CC        N  Rmeas    Symmetry & operator (in Lattice Cell)

  1   0.808   5.94   0.89    9313  0.115     identity
  2   0.828   6.05   0.91   14088  0.141 *** 2-fold l ( 0 0 1)  {-h,-k,+l}
  3   0.000   0.06   0.01   16864  0.527     2-fold   ( 1-1 0)  {-k,-h,-l}
  4   0.871   6.33   0.95   10418  0.100 *** 2-fold   ( 2-1 0)  {+h,-h-k,-l}
  5   0.000   0.53   0.08   12639  0.559     2-fold h ( 1 0 0)  {+h+k,-k,-l}
  6   0.000   0.06   0.01   16015  0.562     2-fold   ( 1 1 0)  {+k,+h,-l}
  7   0.870   6.32   0.95    2187  0.087 *** 2-fold k ( 0 1 0)  {-h,+h+k,-l}
  8   0.000   0.55   0.08    7552  0.540     2-fold   (-1 2 0)  {-h-k,+k,-l}
  9   0.000  -0.12  -0.02   11978  0.598     3-fold l ( 0 0 1)  {-h-k,+h,+l} {+k,-h-k,+l}
 10   0.000  -0.06  -0.01   17036  0.582     6-fold l ( 0 0 1)  {-k,+h+k,+l} {+h+k,-h,+l}
```

Only the orthorhombic symmetry operators are present

Combining symmetry elements shows a clear preference for Laue group Cmmm

Net Z(CC) scores are
Z+(symmetry in group) - Z-(symmetry not in group)

Likelihood allows for the possibility of pseudo-symmetry

Net Z(CC)

Likelihood

Cell deviation

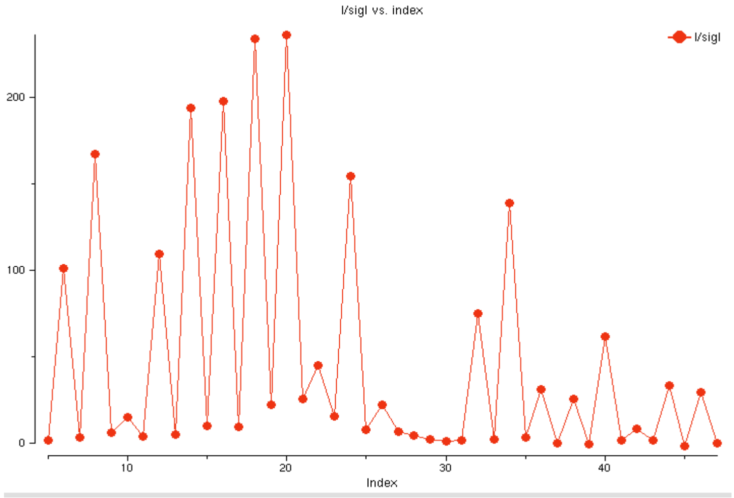
Correlation coefficient
& R-factor

Reindexing

```
    Laue Group         Lklhd   NetZc  Zc+   Zc-    CC    CC-  Rmeas   R-  Delta  ReindexOperator

> 1     C m m m  *** 0.991   6.00  6.12  0.12   0.93  0.02   0.12  0.56   0.1 [1/2h1/2k,3/2h+1/2k,l]
> 2   C 1 2/m 1      0.367   5.00  6.13  1.13   0.95  0.17   0.10  0.48   0.1 [3/2h+1/2k,-1/2h+1/2k,l]
> 3   C 1 2/m 1      0.365   4.55  6.04  1.49   0.95  0.22   0.09  0.46   0.1 [1/2h-1/2k,3/2h+1/2k,l]
> 4   P 1 2/m 1      0.250   4.88  5.99  1.11   0.91  0.17   0.14  0.49   0.0 [1/2h+1/2k,l,1/2h-1/2k]
  5          P -1    0.031   4.27  5.94  1.67   0.89  0.25   0.12  0.44   0.0 [-1/2h+1/2k,-1/2h-1/2k,l]
  6   C 1 2/m 1      0.000   2.45  4.18  1.73   0.08  0.26   0.54  0.44   0.1 [3/2h-1/2k,1/2h+1/2k,l]
  7   C 1 2/m 1      0.000   1.62  3.40  1.79   0.08  0.27   0.56  0.43   0.1 [-1/2h-1/2k,3/2h-1/2k,l]
  8   C 1 2/m 1      0.000   0.60  2.55  1.95   0.01  0.29   0.56  0.42   0.0 [-k,h,l]
  9   C 1 2/m 1      0.000   0.57  2.52  1.96   0.01  0.29   0.53  0.43   0.0 [h,k,l]
 10          P -3    0.000   0.75  2.68  1.93  -0.02  0.29   0.60  0.42   0.1 [1/2h-1/2k,1/2h+1/2k,l]
 11     C m m m      0.000   2.60  3.80  1.20   0.44  0.18   0.38  0.47   0.1 [-1/2h-1/2k,3/2h-1/2k,l]
=12     C m m m      0.000   0.94  2.59  1.65   0.26  0.25   0.42  0.46   0.0 [h,k,l]
 13        P 6/m     0.000   0.83  2.54  1.70   0.24  0.26   0.45  0.44   0.1 [1/2h-1/2k,1/2h+1/2k,l]
 14    P -3 m 1      0.000   0.72  2.46  1.74   0.24  0.26   0.45  0.44   0.1 [1/2h-1/2k,1/2h+1/2k,l]
 15    P -3 1 m      0.000  -0.57  1.79  2.36   0.10  0.35   0.52  0.39   0.1 [1/2h-1/2k,1/2h+1/2k,l]
 16   P 6/m m m      0.000   2.09  2.09  0.00   0.25  0.00   0.44  0.00   0.1 [1/2h-1/2k,1/2h+1/2k,l]
```

Screw axis along 00l shows space group is $C222_1$

Screws detected by Fourier analysis of I/σ



PeakHeight from Fourier analysis
1.0 is perfect screw

"Probability" of screw

```
       Zone                 Number  PeakHeight  SD  Probability ReflectionCondition

 1   screw axis 2(1) [c]     109     0.878    0.083    0.747  00l: l=2n

Spacegroup           TotProb SysAbsProb      Reindex         Conditions

 <C 2 2 21> ( 20)     1.063   0.747                          00l: l=2n (zones 1)
 ..........
  <C 2 2 2> ( 21)     0.360   0.253
```

This relies on having observations for axial reflections along a*,b* and c*. Results may be unreliable for incomplete data, or when one axis is aligned along the rotation axis.

Alternative indexing

If the true point group is lower symmetry than the lattice group, alternative valid but non-equivalent indexing schemes are possible, related by symmetry operators present in lattice group but not in point group (these are also the cases where merohedral twinning is possible)

eg if in space group P3 there are 4 different schemes
(h,k,l) or (-h,-k,l) or (k,h,-l) or (-k,-h,-l)

This can be a problem when merging data from multiple crystals.

- For the first crystal, you can choose any indexing scheme.
- For subsequent crystals, the autoindexing will randomly choose one setting, and we need to make it consistent.
- *POINTLESS* will do this for you if the MTZ files for all crystals are input together, by testing the different indexing schemes for the second and subsequent MTZ files and selecting the one that gives the best agreement with the data from the first MTZ file.

Data Processing: Scaling

The scaling step involves several operations:

- Determining a (resolution dependent) scale factor for each image
- Adding together the individual components of partially recorded reflections (for 2D integration programs like MOSFLM)
- Averaging (merging) symmetry related reflections
- Rejecting outliers
- Adjusting the standard deviations

The scaling and merging step is important because it provides the main diagnostics of data quality and provides an objective way of judging if the data collection and processing are satisfactory.

Because of this diagnostic role, it is important that data are scaled as soon as possible after collection, ideally while other crystals are still available in case extra data need to be collected.

In CCP4, scaling is best performed with the program AIMLESS, a replacement for the original SCALA program.

Choices

- What scaling model?
  - the scaling model should reflect the experiment

  (as a user you do not have much control over this)
- Is the dataset any good?
  - should it be thrown away immediately?
  - are there bits which should be discarded (bad images, radiation damage) ?
  - what is the real resolution?

Why are reflections on different scales?

Various physical factors lead to observed intensities being on different scales. Scaling models should if possible reflect the experiment so different experiments may require different models.

Understanding the effect of these factors allows a sensible design of correction and an understanding of what can go wrong.

Factors that need to be considered are:

- Those related to incident beam and the rotation camera
- Those related to the crystal and the diffracted beam
- Those related to the detector

1) Factors related to incident X-ray beam and rotation camera

- Incident beam intensity: variable on synchrotrons and not normally measured. Assumed to be constant during a single image, or at least varying smoothly and slowly (relative to exposure time). If this is not true, the data will be poor.

- Illuminated volume: changes with ϕ if the beam is smaller than the crystal.

- Absorption in primary beam by crystal: indistinguishable from illuminated volume changes.

- Variations in rotation speed and shutter synchronisation: These errors are disastrous, difficult to detect, and impossible to correct for: we **assume** that the crystal rotation rate is constant and that adjacent images exactly abut in ϕ. Shutter synchronisation errors lead to partial bias that may be **positive**, unlike the usual negative bias, but this is no longer an issue with shutterless data collection.

2) Factors related to crystal and diffracted beam

- Absorption in secondary (diffracted) beam - serious at long wavelength (including CuKα), worth correcting for SAD/MAD data, especially sulphur SAD.

- Radiation damage - serious on all modern high brilliance synchrotron sources. Not correctable unless small as the structure is changing.
Extrapolation to zero (quarter) dose successful in some cases (Kay Diederich).

The relative B-factor is largely a correction for radiation damage (but it can also soak up other errors).

3) Factors related to the detector

- The detector should be properly calibrated for spatial distortion and sensitivity of response, and should be stable. Problems with this are difficult to detect from typical diffraction data, but can be seen in cases of very high symmetry (cubic).

- The useful area of the detector should be calibrated or told to the integration program

  – Calibration should flag defective pixels, hot pixels and dead regions eg between tiles

  – The user should tell the integration program about shadows from the beamstop, beamstop support or cryocooler (define bad areas by circles, rectangles, arcs etc)

Determination of scales

What information do we have?

Scales are determined by comparison of symmetry-related reflections, ie by adjusting scale factors to get the best internal consistency of intensities. Note that we do not know the true intensities and an internally-consistent dataset is not necessarily correct. *Systematic errors will remain*.

$$\text{Minimize } \Phi = \Sigma_{hl}\, w_{hl}\, (I_{hl} - 1/k_{hl}\langle I_h \rangle)^2$$

I_{hl} l' th intensity observation of reflection **h**

k_{hl} scale factor for I_{hl} $\langle I_h \rangle$ current estimate of I_h

k_{hl} is a function of the parameters of the scaling model

$g_{hl} = 1/ k_{hl}$ is a function of the parameters of the scaling model

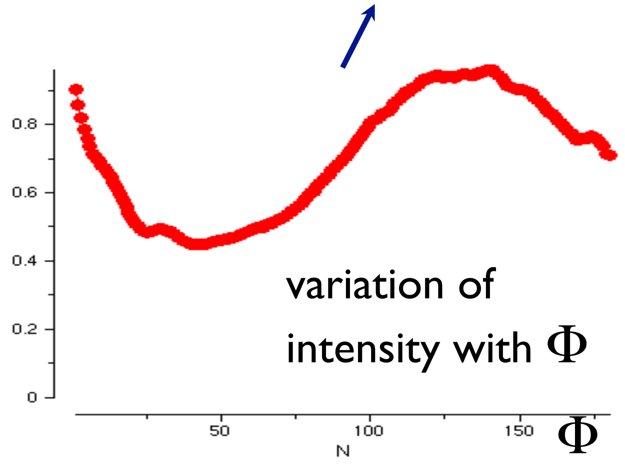
$g_{hl} = g(\phi$ rotation/image number) . g(time) . g(s) …. other factors

Primary beam s_o B-factor Absorption

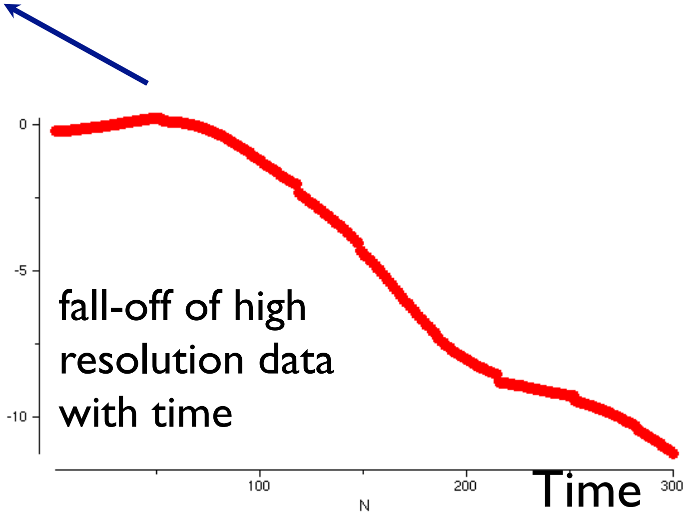
Scaling function

g_{hl} = g(Φ rotation/image number) . g(*time*) . g(**s**) ...*other factors*

Primary beam s_0 *B-factor* *Absorption* *eg "tails"*



variation of intensity with Φ



fall-off of high resolution data with time

scale is smooth function of spindle rotation (Φ)

or discontinuous function of image (batch) number (usually less appropriate)

g(time) = $\exp[+2B(\text{time}) \sin^2\theta / \lambda^2]$

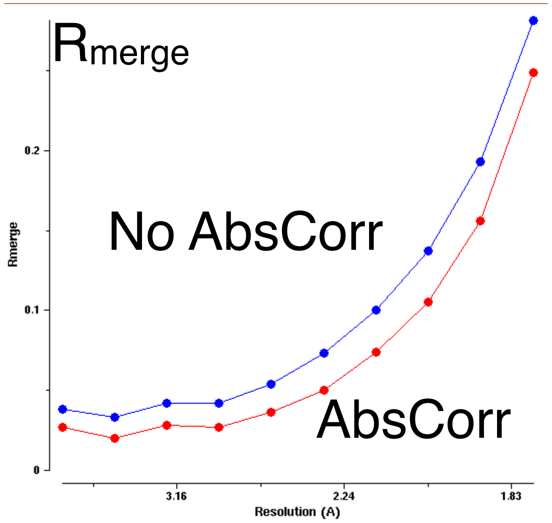
essentially a time-dependent radiation damage correction

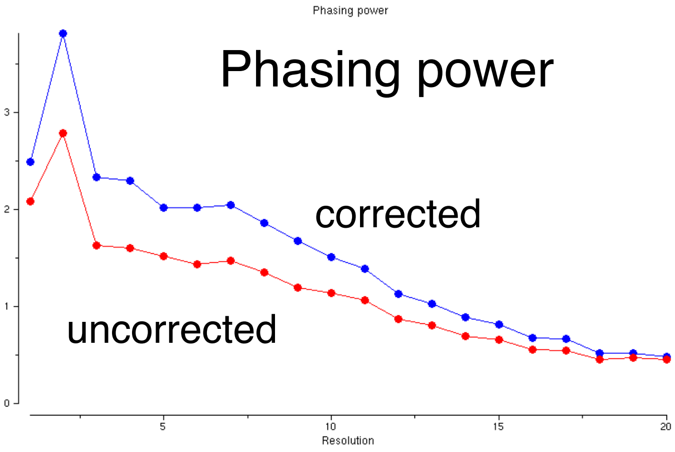
Secondary beam correction (absorption)

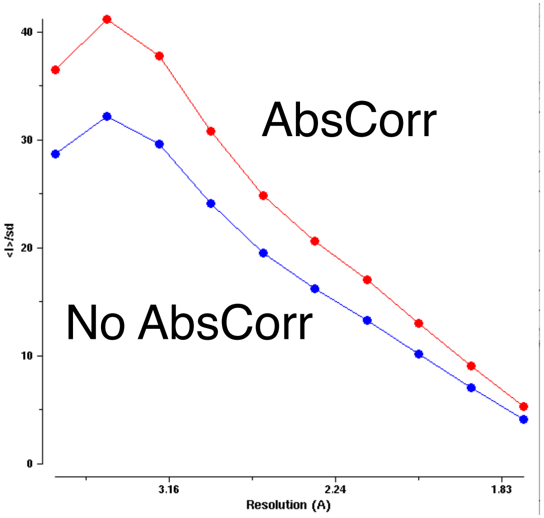
scale as function of secondary beam direction (θ,ϕ)

expressed as sum of spherical harmonics $g(\theta,\phi) = \Sigma_l \Sigma_m C_{lm} Y_{lm}(\theta,\phi)$

Correction improves the data







Sample dataset: Rotating anode (RU200, Osmic mirrors, Mar345) Cu Kα (1.54Å) 100 images, 1°, 5min/°, resolution 1.8Å

Scaling datasets together

For multiple-wavelength datasets, it is best to scale all wavelengths together simultaneously. This is then a ***local*** scaling to minimise the difference between datasets, reducing the systematic error in the anomalous and dispersive differences which are used for phasing

Other advantages of simultaneous scaling:-

- rejection of outliers with much higher reliability because of higher multiplicity
- correlations between ΔF_{anom} and ΔF_{disp} indicate the reliability of the phasing signal

What to look at ?

How well do equivalent observations agree with each other ?

1. R-factors

(a) R_{merge} (R_{sym}) = $\Sigma \mid I_{hl} - \langle I_h \rangle \mid / \Sigma \mid \langle I_h \rangle \mid$

This is the traditional measure of agreement, but it increases with higher multiplicity even though the merged data is better.

(b) $R_{meas} = R_{r.i.m.} = \Sigma \sqrt{(n/n-1)} \mid I_{hl} - \langle I_h \rangle \mid / \Sigma \mid \langle I_h \rangle \mid$

The multiplicity-weight R-factor allows for the improvement in data with higher multiplicity. This is particularly useful when comparing different possible point-groups (although this is normally done with POINTLESS).

*Diederichs & Karplus, Nature Structural Biology, **4**, 269-275 (1997)*
*Weiss & Hilgenfeld, J.Appl.Cryst. **30**, 203-205 (1997)*

2. Intensities and standard deviations: what is the real resolution ?

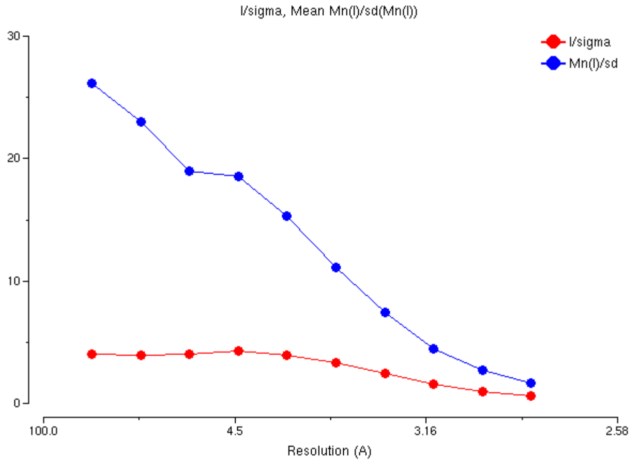
(a) Looking at the mean $I/\sigma(I)$ as a function of resolution.

First, modify the standard deviation estimates from the integration program:

Corrected $\sigma'(I_{hl})^2 = \mathrm{SDfac}^2\,[\sigma^2 + \mathrm{SdB}\,\langle I_h\rangle + (\mathrm{SdAdd}\,\langle I_h\rangle)^2]$

where the three parameters SDfac, SdB and SdAdd are determined automatically so that the modified sigmas (σ') reflect (on average) the actual differences between symmetry related reflections.

The corrected $\sigma'(I)$ is compared with the intensities: the most useful statistic is $\langle \langle I\rangle / \sigma(\langle I\rangle) \rangle$ (labelled Mn(I)/sd in table) as a function of resolution.



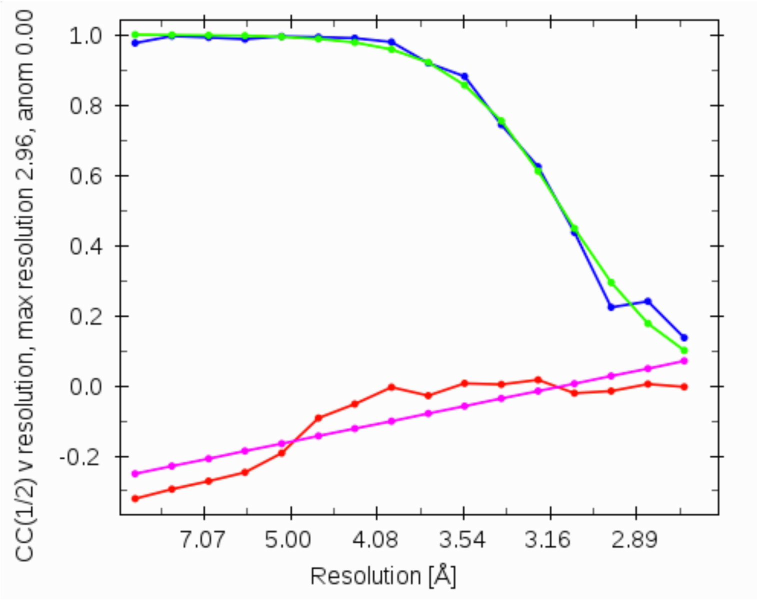
This statistic shows the improvement of the estimate of <I> with multiple measurements. It is the best indicator of the true resolution limit

$\langle \langle I\rangle / \sigma(\langle I\rangle) \rangle$ greater than 1.5 or 2.0

Maybe lower for anisotropic data, 1.5 to 1.0

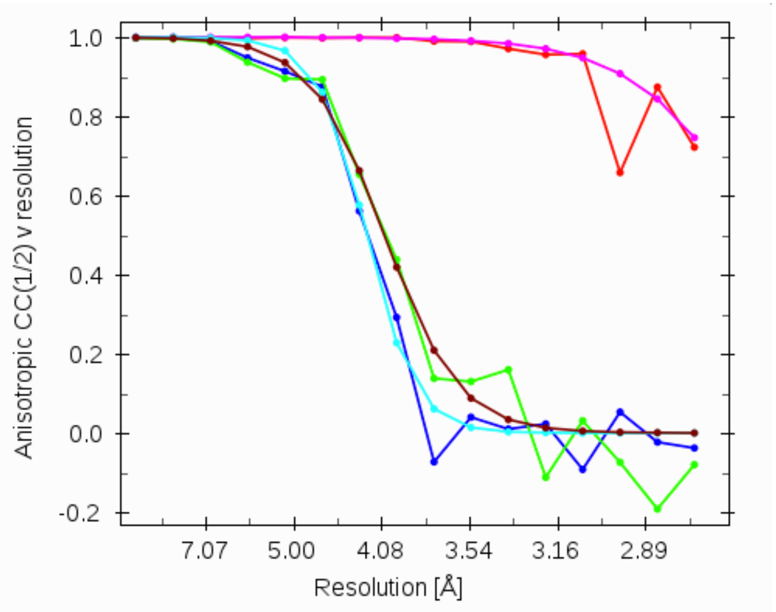
(b) Correlation between half datasets (random halves)

Divide the reflections (randomly) into two half-datasets and calculate a correlation coefficient between these two half-datasets as a function of resolution:



Overall

(green and pink lines are smoothed fits)



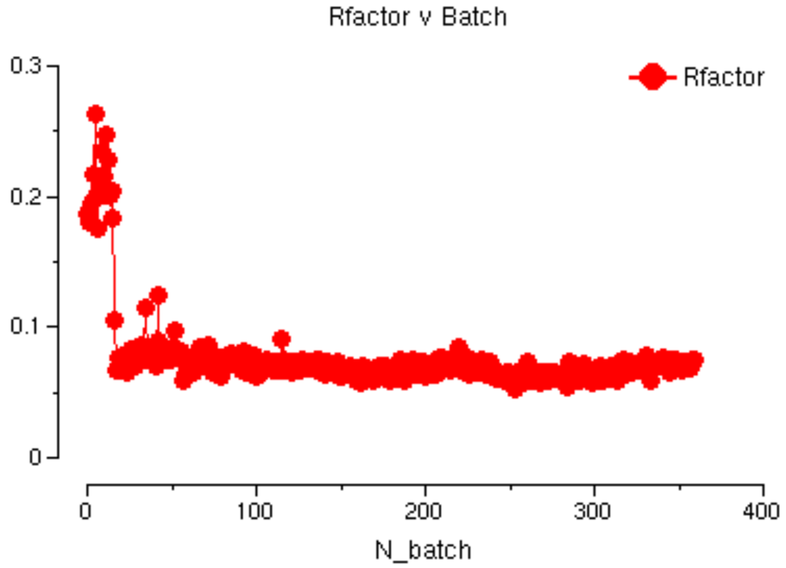
Anisotropic

(light blue, dark brown and pink lines are smoothed fits)

Resolution limit can be defined as the resolution at which CC(1/2) drops to 0.3

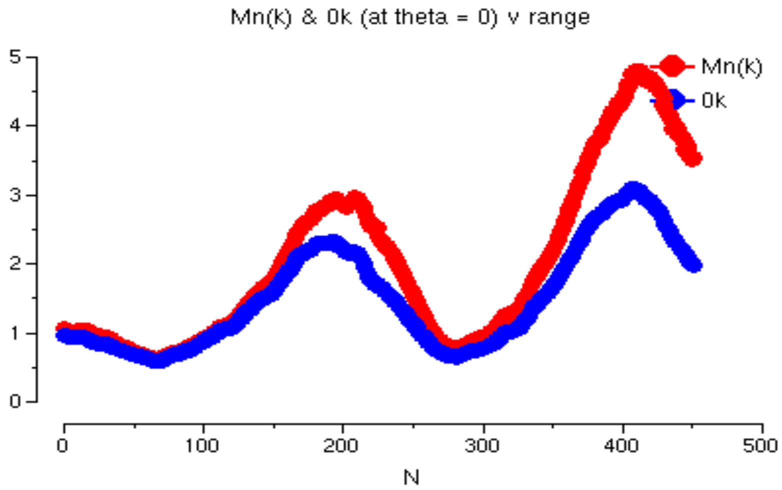
Are some parts of the data bad ?

Analysis of R_{merge} against batch number gives a very clear indication of problems local to some regions of the data. Perhaps something has gone wrong with the integration step, or there are some bad images

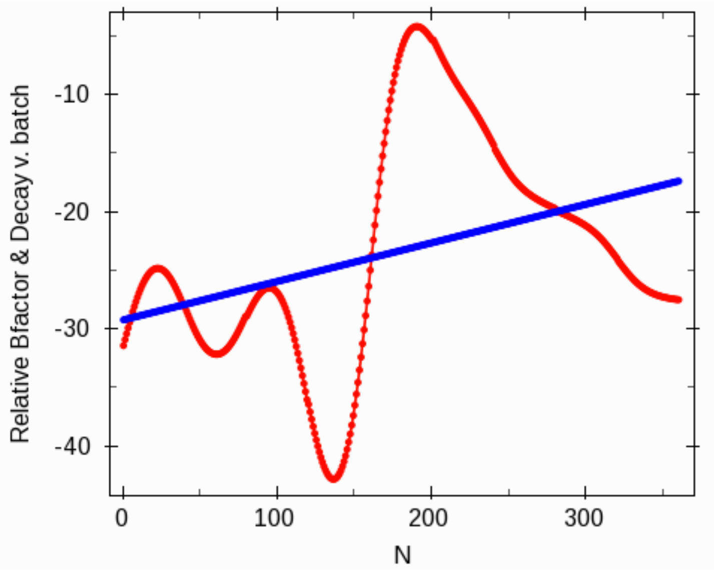


Here the beginning of the dataset is wrong due to problems in integration

Do the parameters (k, B etc) make physical sense ?



These scale factors follow a reasonable curve for variations in illuminated volume.



These B-factors are not sensible. This can happen if there are serious outliers. Try reducing the outlier cutoff, and inspect the ROGUES file.

Partial bias

This measures the systematic difference between fulls and summed partials (if there are any fully recorded observations) or between partials over “n” images (called I_{full}) and partials over “n+1” images (called $I_{partial}$). It is only calculated for 2D integration programs like MOSFLM (not for DIALS/XDS).

$$\text{Fractional Bias} = \Sigma\,(<I_{full}> - I_{partial}) \,/\, \Sigma <I>$$

Typically, its value is negative, ie the summed partials are bigger than the fulls, due to truncation of diffuse scattering tails on fulls (a partially-recorded observation is recorded over at least twice the angular range of a full)

Negative bias greater than ~3% can indicate that the mosaicity has been underestimated. If only observed at low resolution the mosaic block size may be too big. For very weak (high resolution) data it can be large even for correct processing.

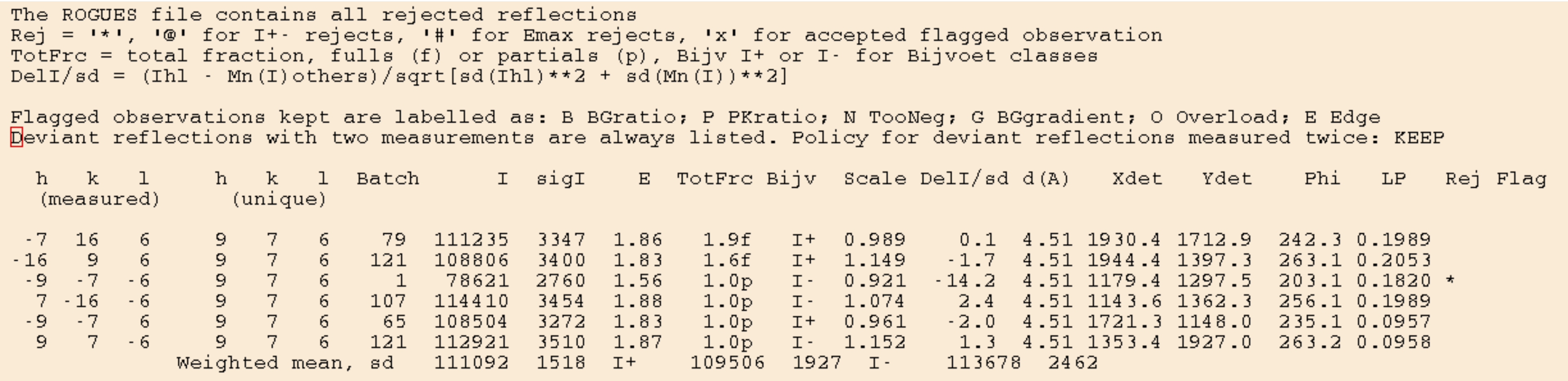
Outliers

Detection of outliers is easiest if the multiplicity is high

Removal of spots behind the backstop shadow does not work well at present: usually it rejects all the good ones, so **mask out the backstop shadow !**

It is also possible to define regions of the detector that are to be ignored (rectangles and arcs of circles)

Inspect the ROGUES file to see what is being rejected (at least occasionally)

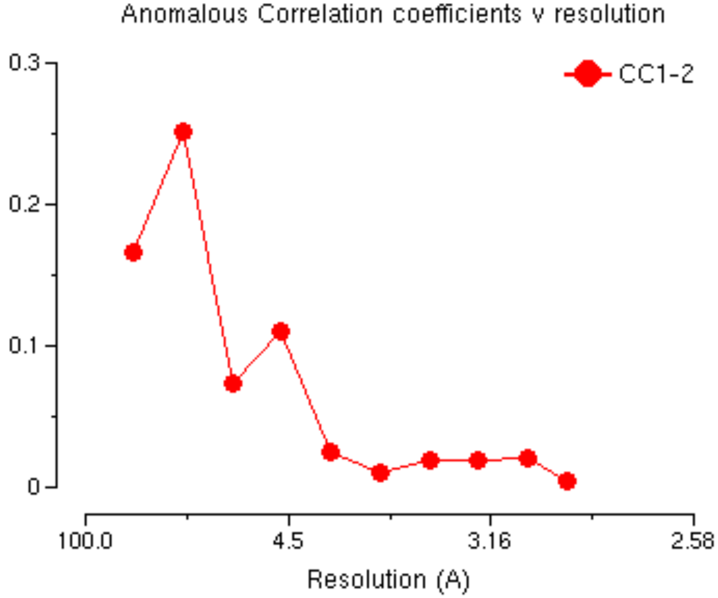


```
The ROGUES file contains all rejected reflections
Rej = '*', '@' for I+- rejects, '#' for Emax rejects, 'x' for accepted flagged observation
TotFrc = total fraction, fulls (f) or partials (p), Bijv I+ or I- for Bijvoet classes
DelI/sd = (Ihl - Mn(I)others)/sqrt[sd(Ihl)**2 + sd(Mn(I))**2]

Flagged observations kept are labelled as: B BGratio; P PKratio; N TooNeg; G BGgradient; O Overload; E Edge
Deviant reflections with two measurements are always listed. Policy for deviant reflections measured twice: KEEP

  h   k   l      h   k   l  Batch       I  sigI     E  TotFrc Bijv  Scale DelI/sd d(A)    Xdet    Ydet     Phi    LP   Rej Flag
  (measured)      (unique)

 -7  16   6      9   7   6     79  111235  3347  1.86    1.9f   I+   0.989    0.1  4.51 1930.4 1712.9   242.3 0.1989
-16   9   6      9   7   6    121  108806  3400  1.83    1.6f   I+   1.149   -1.7  4.51 1944.4 1397.3   263.1 0.2053
 -9  -7  -6      9   7   6      1   78621  2760  1.56    1.0p   I-   0.921  -14.2  4.51 1179.4 1297.5   203.1 0.1820 *
  7 -16  -6      9   7   6    107  114410  3454  1.88    1.0p   I-   1.074    2.4  4.51 1143.6 1362.3   256.1 0.1989
 -9  -7   6      9   7   6     65  108504  3272  1.83    1.0p   I+   0.961   -2.0  4.51 1721.3 1148.0   235.1 0.0957
  9   7  -6      9   7   6    121  112921  3510  1.87    1.0p   I-   1.152    1.3  4.51 1353.4 1927.0   263.2 0.0958
          Weighted mean, sd    111092  1518  I+    109506  1927  I-     113678  2462
```

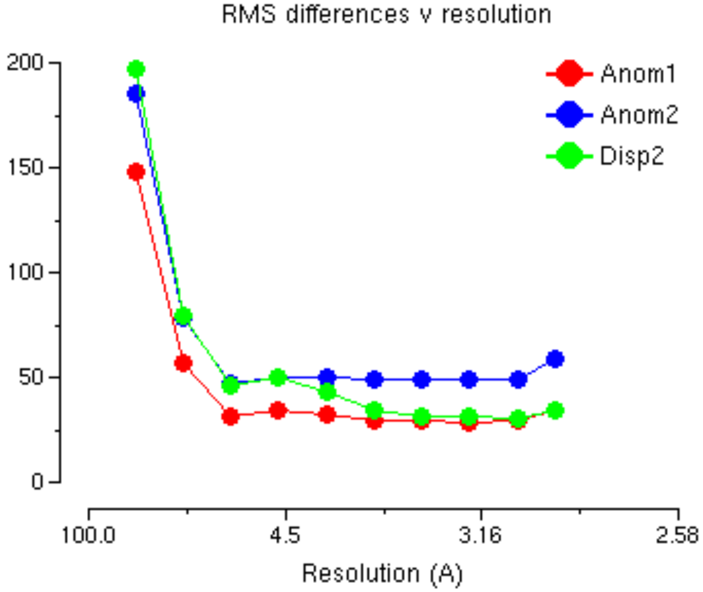
If more than 1% of reflections are being rejected, something is wrong (could be the wrong symmetry) !

Anomalous signal correlation coefficient

If different datasets are scaled together (eg MAD data), correlations between the anomalous and dispersive differences can be calculated between datasets. The same analysis can be applied to two “halves” of a single λ dataset if the multiplicity is high enough.



Anomalous Correlation coefficients v resolution



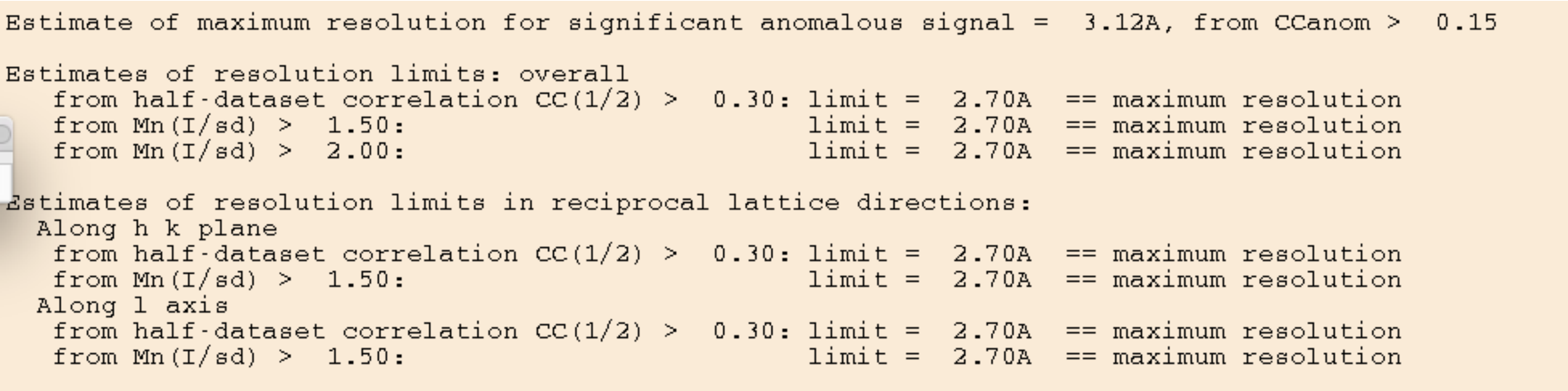
RMS differences v resolution

In this case there is little anomalous signal beyond about 6Å resolution (Hg derivative, two wavelengths)

Estimating the “true” resolution of the data

AIMLESS gives estimates of the resolution of the data based on both the Mn(I/sd) (cutoffs of 1.5 and 2.0) and the half dataset correlation CC(1/2).

Both overall values and values in crystallographically independent directions are given, which indicates any anisotropy (the example below is for hexagonal data). These estimates will differ if the corrected standard deviations do not agree with the actual differences between symmetry related reflections, in which case the CC(1/2) values are generally more reliable.



```
Estimate of maximum resolution for significant anomalous signal =   3.12A, from CCanom >  0.15

Estimates of resolution limits: overall
   from half-dataset correlation CC(1/2) >  0.30: limit =   2.70A  == maximum resolution
   from Mn(I/sd) >  1.50:                         limit =   2.70A  == maximum resolution
   from Mn(I/sd) >  2.00:                         limit =   2.70A  == maximum resolution

Estimates of resolution limits in reciprocal lattice directions:
  Along h k plane
   from half-dataset correlation CC(1/2) >  0.30: limit =   2.70A  == maximum resolution
   from Mn(I/sd) >  1.50:                         limit =   2.70A  == maximum resolution
  Along l axis
   from half-dataset correlation CC(1/2) >  0.30: limit =   2.70A  == maximum resolution
   from Mn(I/sd) >  1.50:                         limit =   2.70A  == maximum resolution
```

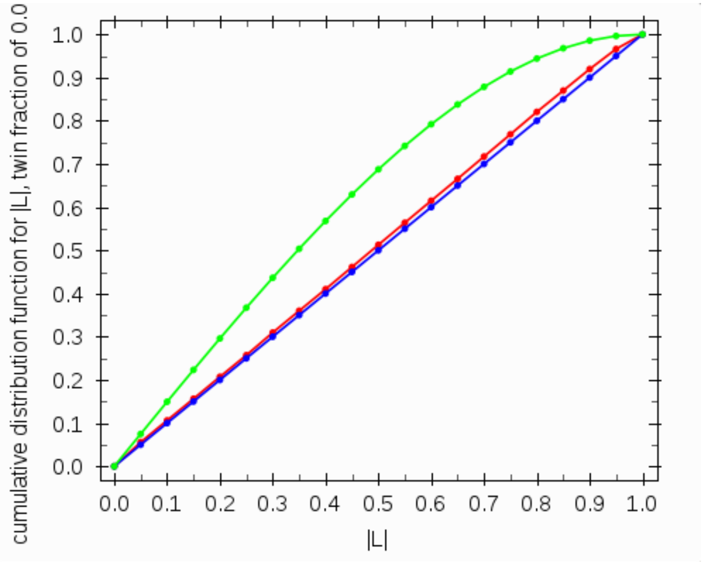
“maximum resolution” is the resolution to which the images have been integrated.

Converting Intensities to Amplitudes (Truncate)

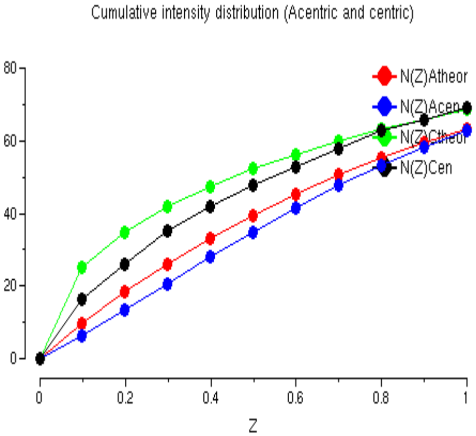
1) Gives best estimate of amplitude for reflections where the measured intensity is negative.
2) Provides an estimate of Wilson B factor (how rapidly amplitudes fall of with resolution).
3) Detects anisotropy in diffraction.
4) Check for merohedral twinning: L test and N(z) test:

The L test

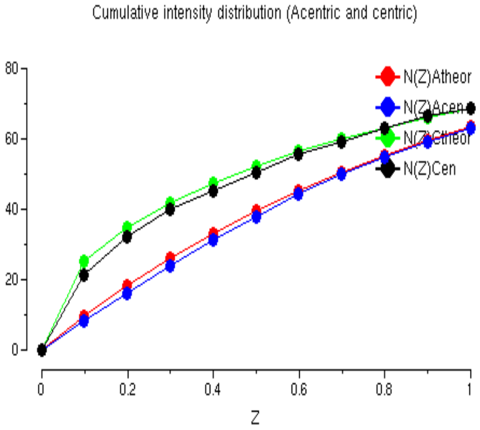
Cumulative intensity plot (N(z))



Green: Twinned (twin faction 0.5)
Blue: Untwinned
Red: Data

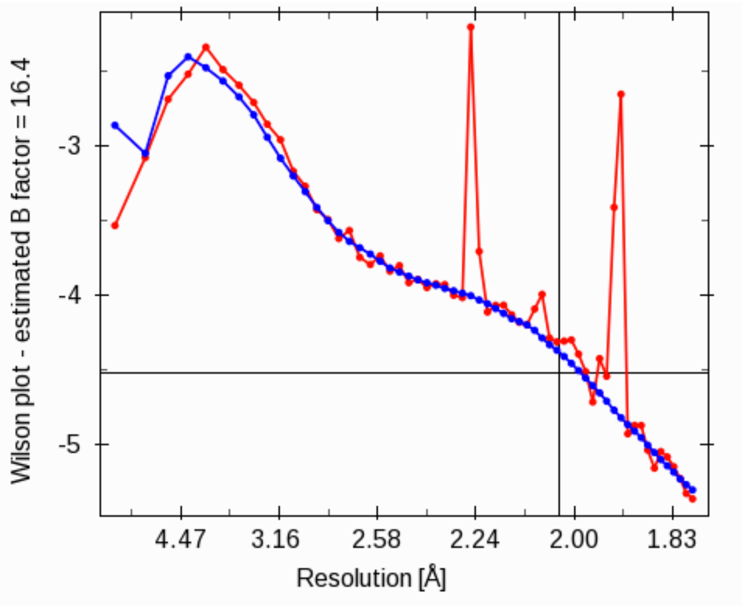


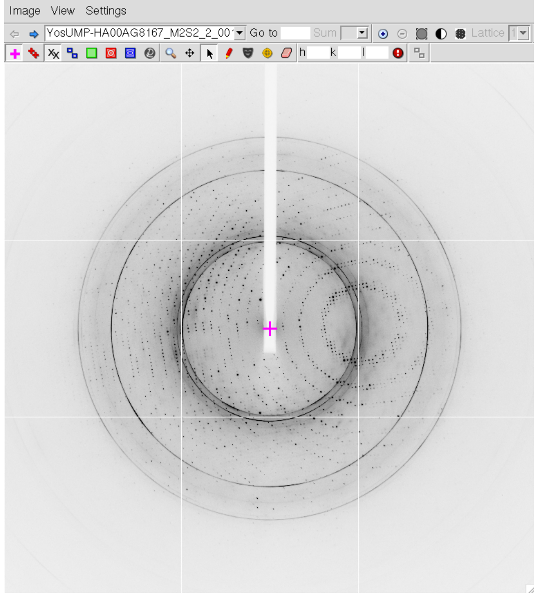
Twinned

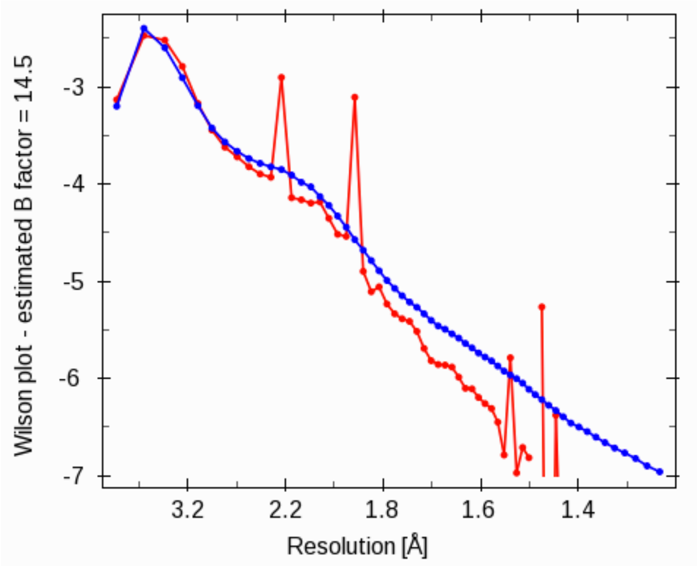


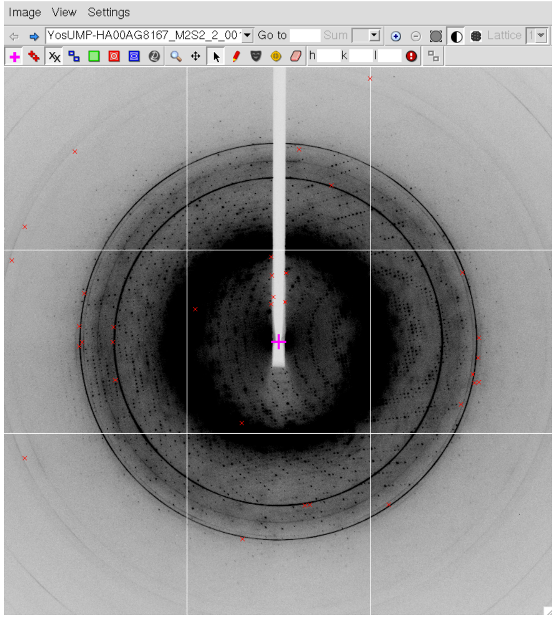
Untwinned

The Wilson plot shows problems in integration, eg due to ice:







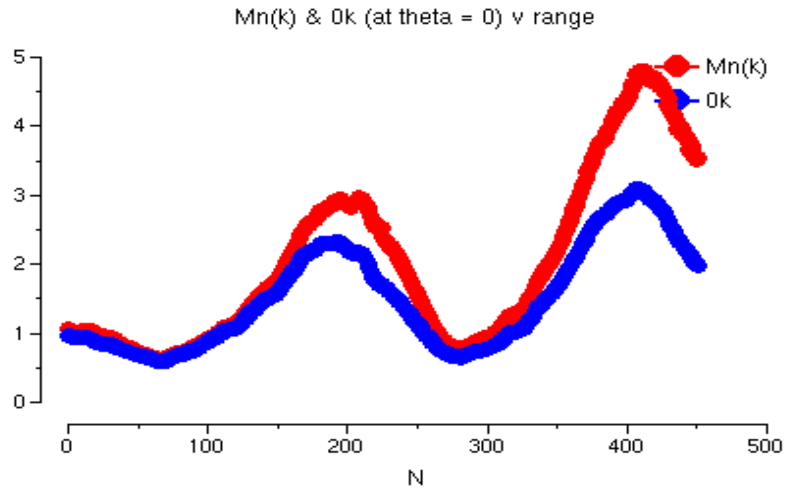


More examples in the tutorials ….

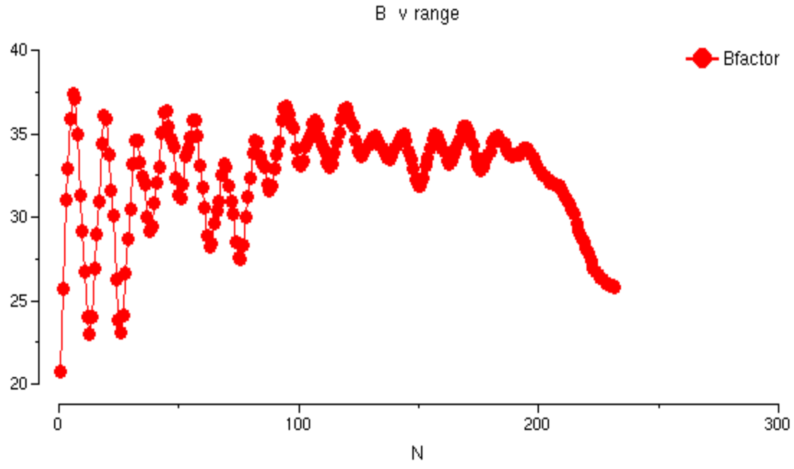
Thanks to Phil Evans:

The End

Do the parameters (k, B etc) make physical sense ?



These scale factors follow a reasonable absorption curve

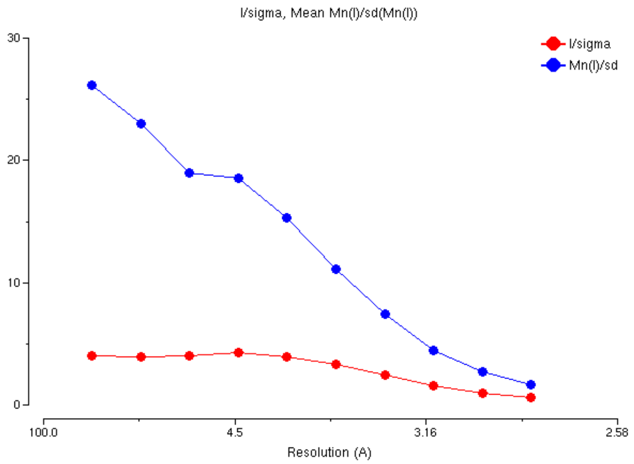


These B-factors are not sensible
As well as being highly variable, they are also **positive**: Bfactors should be negative (ie sharpening later observations)

2. **Intensities and standard deviations: what is the real resolution ?**

(a) Corrected $\sigma'(I_{hl})^2 = \mathrm{SDfac}^2\,[\sigma^2 + \mathrm{SdB}\,\langle I_h\rangle + (\mathrm{SdAdd}\,\langle I_h\rangle)^2]$

The corrected $\sigma'(I)$ is compared with the intensities: the most useful statistic is $\langle \langle I\rangle / \sigma(\langle I\rangle) \rangle$ (labelled Mn(I)/sd in table) as a function of resolution

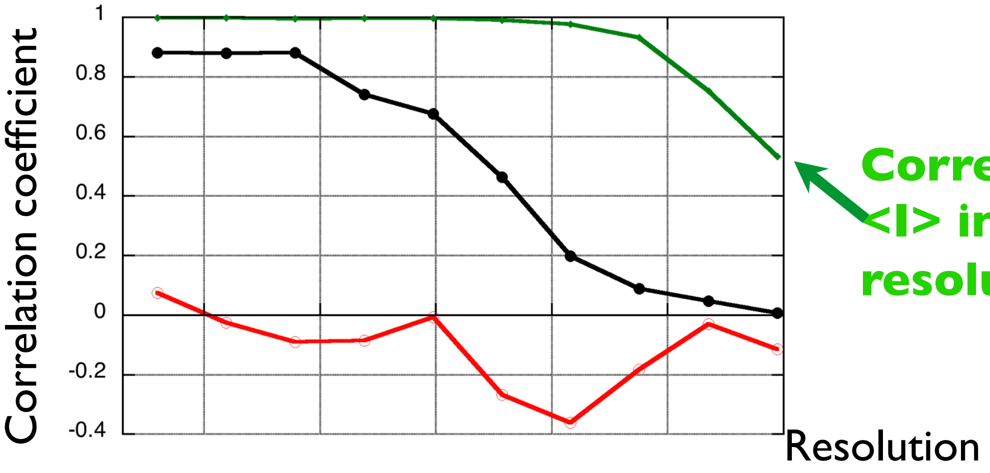


This statistic shows the improvement of the estimate of <I> with multiple measurements. It is the best indicator of the true resolution limit

$\langle \langle I\rangle / \sigma(\langle I\rangle) \rangle$ greater than 1.5 or 2.0

Maybe lower for anisotropic data, 1.5 to 1.0

(b) Correlation between half datasets (random halves)
Resolution limit can be defined as the resolution at which CC drops to 0.3



Correlation of <I> indicating a resolution limit

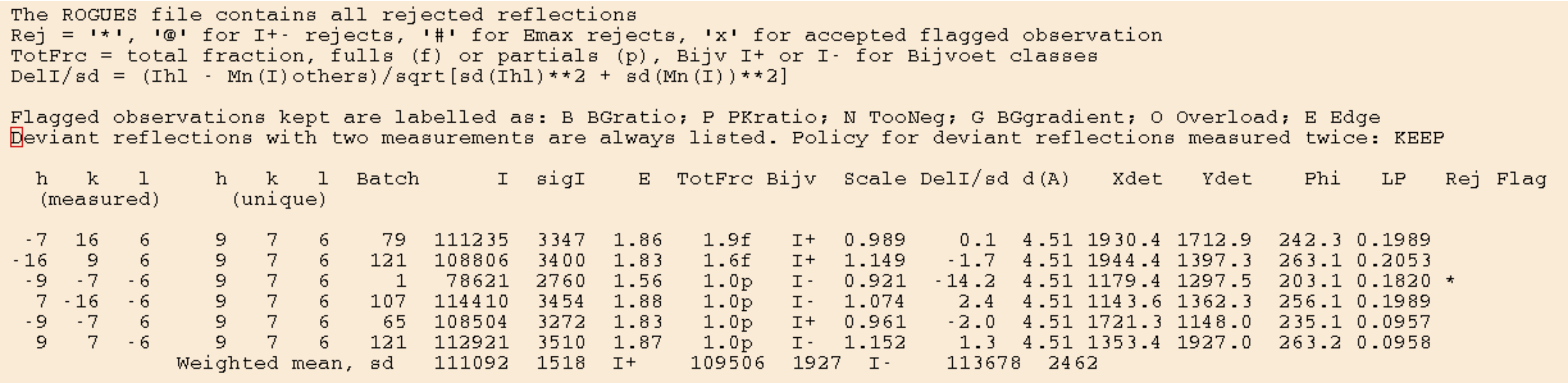
Outliers

Detection of outliers is easiest if the multiplicity is high

Removal of spots behind the backstop shadow does not work well at present: usually it rejects all the good ones, so **tell Mosflm where the backstop shadow is !**

It is also possible to define regions of the detector that are to be ignored (rectangles and arcs of circles)

Inspect the ROGUES file to see what is being rejected (at least occasionally)



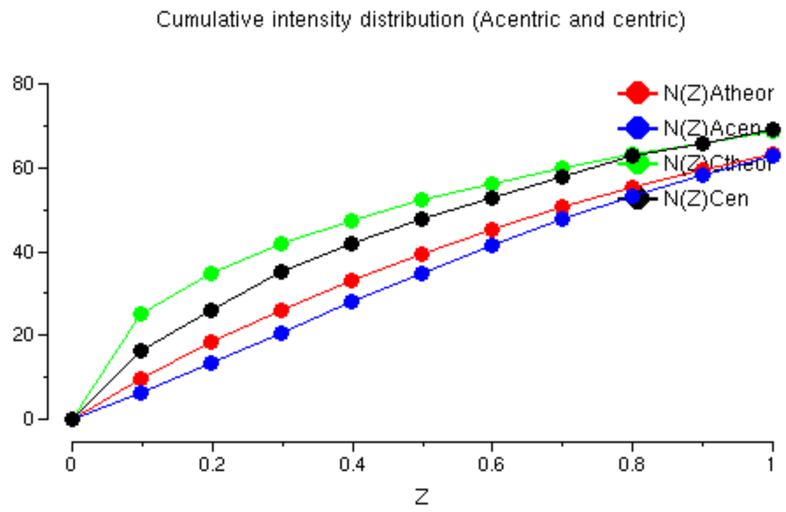
```
The ROGUES file contains all rejected reflections
Rej = '*', '@' for I+- rejects, '#' for Emax rejects, 'x' for accepted flagged observation
TotFrc = total fraction, fulls (f) or partials (p), Bijv I+ or I- for Bijvoet classes
DelI/sd = (Ihl - Mn(I)others)/sqrt[sd(Ihl)**2 + sd(Mn(I))**2]

Flagged observations kept are labelled as: B BGratio; P PKratio; N TooNeg; G BGgradient; O Overload; E Edge
Deviant reflections with two measurements are always listed. Policy for deviant reflections measured twice: KEEP

  h   k   l     h   k   l  Batch      I  sigI    E  TotFrc Bijv  Scale DelI/sd d(A)   Xdet   Ydet    Phi   LP   Rej Flag
  (measured)     (unique)

 -7  16   6     9   7   6    79  111235  3347  1.86   1.9f   I+  0.989    0.1  4.51 1930.4 1712.9  242.3 0.1989
-16   9   6     9   7   6   121  108806  3400  1.83   1.6f   I+  1.149   -1.7  4.51 1944.4 1397.3  263.1 0.2053
 -9  -7  -6     9   7   6     1   78621  2760  1.56   1.0p   I-  0.921  -14.2  4.51 1179.4 1297.5  203.1 0.1820 *
  7 -16  -6     9   7   6   107  114410  3454  1.88   1.0p   I-  1.074    2.4  4.51 1143.6 1362.3  256.1 0.1989
 -9  -7   6     9   7   6    65  108504  3272  1.83   1.0p   I+  0.961   -2.0  4.51 1721.3 1148.0  235.1 0.0957
  9   7  -6     9   7   6   121  112921  3510  1.87   1.0p   I-  1.152    1.3  4.51 1353.4 1927.0  263.2 0.0958
             Weighted mean, sd   111092  1518  I+    109506  1927  I-    113678  2462
```

If more than 1% of reflections are being rejected, something is wrong !

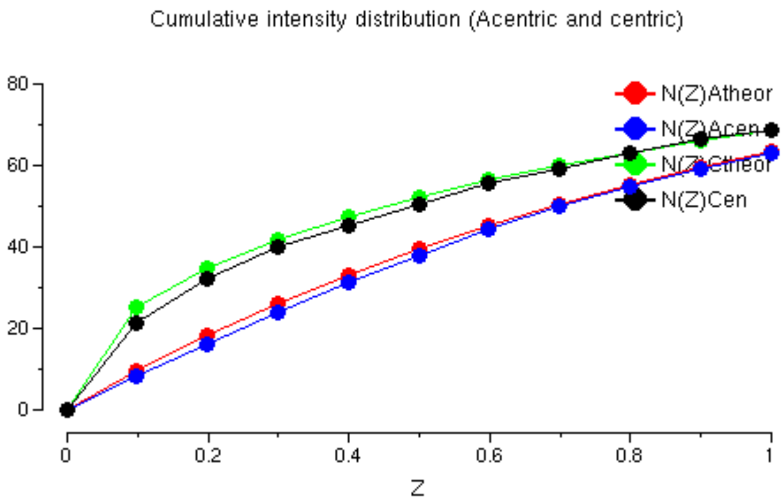
Converting Intensities to Amplitudes (Truncate)

1) Gives best estimate of amplitude for reflections where the measured intensity is negative.
2) Provides an estimate of Wilson B factor (how rapidly amplitudes fall of with resolution).
3) Detects anisotropy in diffraction.
4) Check for merohedral twinning: L test and N(z) test:

Cumulative intensity plot (N(z))


Cumulative intensity distribution (Acentric and centric)

Twinned


Cumulative intensity distribution (Acentric and centric)

Untwinned