# Data reduction with POINTLESS and AIMLESS

**James Parkhurst**

CCP4 workshop, Faridabad

February 2016

# Acknowledgements

Phil Evans
(Developer of POINTLESS and AIMLESS)

# Acknowledgements


| | |
|---|---|
| Andrew Leslie | many discussions |
| Harry Powell | many discussions |
| Ralf Grosse-Kunstleve | cctbx |
| Kevin Cowtan | clipper, C++ advice |
| Airlie McCoy | C++ advice, code, useful suggestions, etc |
| Randy Read & co. | minimiser |
| Graeme Winter | testing & bug finding |
| Clemens Vonrhein | testing & bug finding |
| Eleanor Dodson | many discussions |
| Andrey Lebedev | intensity statistics & twinning |
| Norman Stein | ctruncate |
| Charles Ballard | ctruncate |
| George Sheldrick | discussions on symmetry detection |
| Garib Murshudov | intensity statistics |
| Martyn Winn & CCP4 gang | ccp4 libraries |
| Peter Briggs | ccp4i |
| Liz Potterton | ccp4i2 |
| Martin Noble | ccp4i2 |

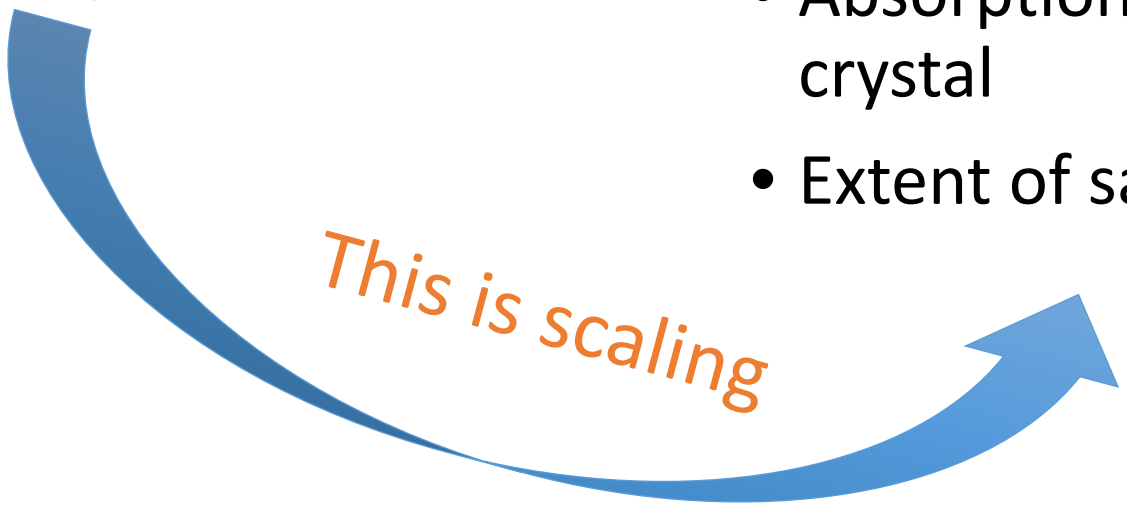
# Purpose

Things we know:

- I, sig(I), corrected for geometric effects
- Lots of observations
- Symmetry

Things we don't know:

- $|F|^2$
- Beam intensity
- Illuminated volume
- Absorption path through crystal
- Extent of sample decay



# Programs

## Pointless

- Determines likely point group
- Corrects space group *if* sufficient information
- Sorts reflections
- Detects screw axes & glide planes
- Re-indexes multiple datasets to a common setting

## Aimless

- Merges partial reflections together
- Puts data onto a common scale
- Merges each set of symmetry equivalent reflections into a single observation

## CTruncate

- Analyses scaled data according to an expected physical model
- Gives statistics on intensity distribution - *e.g.*
  - Wilson statistics
  - twinning analysis
- Outputs |F| values

# Symmetry determination (POINTLESS)

Data reduction with POINTLESS and AIMLESS

# What does POINTLESS do?

Indexing in eg MOSFLM or DIALS only gives the possible lattice symmetry, ie constraints of unit cell dimensions. Crystal classes: cubic, hexagonal/trigonal, tetragonal, orthorhombic, monoclinic, or triclinic, + lattice centring P, C, I, R, or F

**POINTLESS performs the following tasks:**

1. from the cell dimensions, determine the maximum possible lattice symmetry (ignoring any input symmetry)
2. for each possible rotation operator, score related observations pairs for agreement (correlation coefficients and R-factor)
3. score all possible combinations of operators to determine the point group (point groups from maximum down to P1)
4. score axial systematic absences to detect screw axes, hence space group (note that axial observations are sometimes unobserved)

# Score individual symmetry operators in the maximum lattice group

```
Analysing rotational symmetry in lattice group P m -3 m
-------------------------------------------
Scores for each symmetry element

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

  1   0.955   9.70   0.97   13557  0.073     identity
  2   0.062   2.66   0.27   12829  0.488     2-fold   ( 1 0 1)  {+l,-k,+h}
  3   0.065   2.85   0.29   10503  0.474     2-fold   ( 1 0-1)  {-l,-k,-h}
  4   0.056   0.06   0.01   16391  0.736     2-fold   ( 0 1-1)  {-h,-l,-k}
  5   0.057   0.05   0.00   17291  0.738     2-fold   ( 0 1 1)  {-h,+l,+k}
  6   0.049   0.55   0.06   13758  0.692     2-fold   ( 1-1 0)  {-k,-h,-l}
  7   0.950   9.59   0.96   12584  0.100 *** 2-fold k ( 0 1 0)  {-h,+k,-l}
  8   0.049   0.57   0.06   11912  0.695     2-fold   ( 1 1 0)  {+k,+h,-l}
  9   0.948   9.57   0.96   16928  0.136 *** 2-fold h ( 1 0 0)  {+h,-k,-l}
 10   0.944   9.50   0.95   12884  0.161 *** 2-fold l ( 0 0 1)  {-h,-k,+l}
 11   0.054   0.15   0.01   23843  0.812     3-fold   ( 1 1 1)  {+l,+h,+k} {+k,+l,+h}
 12   0.055   0.11   0.01   24859  0.825     3-fold   ( 1-1-1)  {-l,-h,+k} {-k,+l,-h}
 13   0.055   0.14   0.01   22467  0.788     3-fold   ( 1-1 1)  {+l,-h,-k} {-k,-l,+h}
 14   0.055   0.12   0.01   27122  0.817     3-fold   ( 1 1-1)  {-l,+h,-k} {+k,-l,-h}
 15   0.061  -0.10  -0.01   25905  0.726     4-fold h ( 1 0 0)  {+h,-l,+k} {+h,+l,-k}
 16   0.060   2.53   0.25   23689  0.449     4-fold k ( 0 1 0)  {+l,+k,-h} {-l,+k,+h}
 17   0.049   0.56   0.06   25549  0.653     4-fold l ( 0 0 1)  {-k,+h,+l} {+k,-h,+l}
```

Only orthorhombic symmetry operators are present

# Score possible point groups

All possible combinations of rotations are scored to determine the point group. Good scores in symmetry operations which are absent in the sub-group count against that group.

**Example: C-centred orthorhombic which might been hexagonal**

| | Laue Group | | Lklhd | NetZc | Zc+ | Zc- | CC | CC- | Rmeas | R- | Delta | ReindexOperator |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| = 1 | C m m m | *** | 0.989 | 9.45 | 9.62 | 0.17 | 0.96 | 0.02 | 0.08 | 0.76 | 0.0 | [h,k,l] |
| 2 | P 1 2/m 1 | | 0.004 | 7.22 | 9.68 | 2.46 | 0.97 | 0.25 | 0.06 | 0.56 | 0.0 | [-1/2h+1/2k,-l,-1/2h-1/2k] |
| 3 | C 1 2/m 1 | | 0.003 | 7.11 | 9.61 | 2.50 | 0.96 | 0.25 | 0.08 | 0.55 | 0.0 | [h,k,l] |
| 4 | C 1 2/m 1 | | 0.003 | 7.11 | 9.61 | 2.50 | 0.96 | 0.25 | 0.08 | 0.55 | 0.0 | [-k,-h,-l] |
| 5 | P -1 | | 0.000 | 6.40 | 9.67 | 3.27 | 0.97 | 0.33 | 0.06 | 0.49 | 0.0 | [1/2h+1/2k,1/2h-1/2k,-l] |
| 6 | C m m m | | 0.000 | 1.91 | 5.11 | 3.20 | 0.51 | 0.32 | 0.34 | 0.51 | 2.5 | [1/2h-1/2k,-3/2h-1/2k,-l] |
| 7 | P 6/m | | 0.000 | 1.16 | 4.59 | 3.43 | 0.46 | 0.34 | 0.41 | 0.46 | 2.5 | [-1/2h-1/2k,-1/2h+1/2k,-l] |
| 8 | C 1 2/m 1 | | 0.000 | 1.51 | 5.15 | 3.64 | 0.52 | 0.36 | 0.33 | 0.47 | 2.5 | [1/2h-1/2k,-3/2h-1/2k,-l] |
| 9 | C 1 2/m 1 | | 0.000 | 1.51 | 5.15 | 3.64 | 0.51 | 0.36 | 0.33 | 0.47 | 2.5 | [-3/2h-1/2k,-1/2h+1/2k,-l] |
| 10 | P -3 | | 0.000 | 1.04 | 4.75 | 3.71 | 0.48 | 0.37 | 0.40 | 0.45 | 2.5 | [-1/2h-1/2k,-1/2h+1/2k,-l] |
| 11 | C m m m | | 0.000 | 2.13 | 5.23 | 3.10 | 0.52 | 0.31 | 0.32 | 0.52 | 2.5 | [-1/2h-1/2k,-3/2h+1/2k,-l] |
| 12 | C 1 2/m 1 | | 0.000 | 1.64 | 5.25 | 3.61 | 0.53 | 0.36 | 0.32 | 0.47 | 2.5 | [-1/2h-1/2k,-3/2h+1/2k,-l] |
| 13 | C 1 2/m 1 | | 0.000 | 1.67 | 5.27 | 3.60 | 0.53 | 0.36 | 0.32 | 0.47 | 2.5 | [-3/2h+1/2k,1/2h+1/2k,-l] |
| 14 | P -3 1 m | | 0.000 | 0.12 | 4.00 | 3.87 | 0.40 | 0.39 | 0.44 | 0.44 | 2.5 | [-1/2h-1/2k,-1/2h+1/2k,-l] |
| 15 | P -3 m 1 | | 0.000 | 0.14 | 4.00 | 3.86 | 0.40 | 0.39 | 0.44 | 0.44 | 2.5 | [-1/2h-1/2k,-1/2h+1/2k,-l] |
| 16 | P 6/m m m | | 0.000 | 3.93 | 3.93 | 0.00 | 0.39 | 0.00 | 0.44 | 0.00 | 2.5 | [-1/2h-1/2k,-1/2h+1/2k,-l] |

```
Possible spacegroups:
---------------------
Indistinguishable space groups are grouped together on successive lines

'Reindex' is the operator to convert from the input hklin frame to the standard spacegroup frame.

'TotProb' is a total probability estimate (unnormalised)

'SysAbsProb' is an estimate of the probability of the space group based on
the observed systematic absences.

'Conditions' are the reflection conditions (absences)

   Spacegroup          TotProb SysAbsProb     Reindex          Conditions
  <P 21 21 21> ( 19)     0.838   0.851                        h00: h=2n, 0k0: k=2n, 001: l=2n (zones 1,2,3)
   <P 2 21 21> ( 18)     0.104   0.106                        0k0: k=2n, 001: l=2n (zones 2,3)
   <P 21 2 21> ( 18)     0.025   0.026                        h00: h=2n, 001: l=2n (zones 1,3)
   <P 21 21 2> ( 18)     0.012   0.012                        h00: h=2n, 0k0: k=2n (zones 1,2)
```

```
Best Solution      space group P 21 21 21

   Reindex operator:                          [h,k,l]
   Laue group probability:                    0.985
   Systematic absence probability:            0.851
   Total probability:                         0.838
   Space group confidence:                    0.784
   Laue group confidence                      0.982

   Unit cell:    34.16    54.8      68          90      90      90

   17.00 to    1.78    - Resolution range used for Laue group search

   17.00 to    1.78    - Resolution range in file, used for systematic absence check

   Number of batches in file:     100
```

**Note high confidence in Laue group, but lower confidence in space group**

# What can go wrong?

- Pseudo-symmetry or twinning (often connected) can suggest a point group symmetry which is too high. Careful examination of the scores for individual symmetry operators may indicate the truth (the program is not foolproof!)
- POINTLESS works (usually) with unscaled data (hence use of correlation coefficients), so data with a large range of scales, including a dead crystal, may give a too-low symmetry.
- In bad cases either just use the first part of the data, or scale in P1 and run POINTLESS on the scaled unmerged data
- Potential axial systematic absences may be absent or few, so it may not be possible to determine the space group. In that case the output file is labelled with the “space group” with no screw axes, eg P2, P222, P622 etc, and the space group will have to be determined later

**NOTE that the space group is only a hypothesis until the structure has been determined and satisfactorily refined**

# Scaling (AIMLESS)

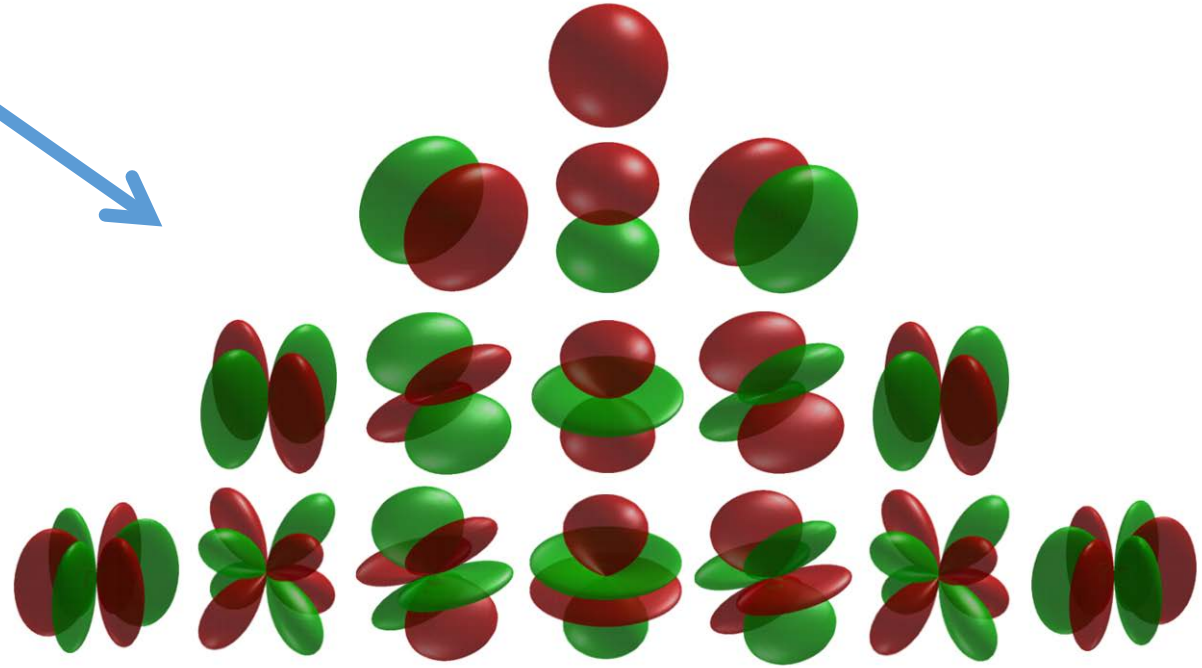
Data reduction with POINTLESS and AIMLESS

# Scaling

- Corrections for some of the things we don’t know can be determined experimentally
- In most cases however empirical corrections are determined
- Have a model for: overall scale (beam intensity + illuminated volume) sample decay and absorption
- Refine model against data, to minimise differences between symmetry related intensities

# Scaling models

- Time or frame # dependent – overall scale
- Time and resolution dependent – decay
- Direction dependent – absorption – for example as spherical harmonics
- All depends on multiplicity



# Objective of scaling

- To model all of the unknown contributions to the measured intensity
- To recover $I=k|F|^2$ for each observation
- Achieved by minimizing the differences between observations – internally consistent not necessarily correct!
- Final result of scaling is average $I=k|F|^2$ for each unique Miller index
- May want to keep I+ and I- separate

# Factors related to incident X-ray beam

- 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 beam smaller than crystal
- absorption in primary beam by crystal: indistinguishable from (b)
- variations in rotation speed and shutter synchronisation. These errors are disastrous, difficult to detect, and (almost) 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 which may be positive, unlike the usual negative bias)*
- Data collection with open shutter (eg with Pilatus detector) avoids synchronisation errors (though variation in rotation speed could still cause trouble, and there is a dead time during readout)

# Factors related to crystal and diffracted beam

- Absorption in secondary beam - serious at long wavelength (including CuKα)
- radiation damage - serious on high brilliance sources. Not easily correctable unless small as the structure is changing
- *Maybe extrapolate back to zero time? (but this needs high multiplicity)*
- *The relative B-factor is largely a correction for the average radiation damage*

# 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 diffraction data. There are known problems in the tile corners of CCD detectors (corrected for in XDS)
- The useful area of the detector should be calibrated or told to the integration program
  - Calibration should flag defective pixels (hot or cold) and dead regions eg between tiles
  - The user should tell the integration program about shadows from the beamstop, beamstop support or cryo-cooler (define bad areas by circles, rectangles, arcs etc)

# Data Quality

Data reduction with POINTLESS and AIMLESS

# Judging data quality

- Are there bad batches?
- Was the radiation damage such that you should exclude the later parts?
- Is the outlier detection working well?
- What is the real resolution? Should you cut the high-resolution data?
- Is there any apparent anomalous signal?
- What is the overall quality of the dataset?
- Are the data twinned?

# AIMLESS summary statistics

| | Overall | InnerShell | OuterShell |
|---|---|---|---|
| Low resolution limit | 150.01 | 150.01 | 1.19 |
| High resolution limit | 1.17 | 6.41 | 1.17 |
| | | | |
| Rmerge (within I+/I-) | 0.063 | 0.024 | 0.000 |
| Rmerge (all I+ and I-) | 0.071 | 0.027 | 0.149 |
| Rmeas (within I+/I-) | 0.077 | 0.029 | 0.000 |
| Rmeas (all I+ & I-) | 0.079 | 0.030 | 0.210 |
| Rpim (within I+/I-) | 0.044 | 0.016 | 0.000 |
| Rpim (all I+ & I-) | 0.034 | 0.013 | 0.149 |
| Rmerge in top intensity bin | 0.030 | - | - |
| Total number of observations | 324157 | 3150 | 300 |
| Total number unique | 71073 | 662 | 286 |
| Mean((I)/sd(I)) | 10.8 | 36.6 | 2.1 |
| Mn(I) half-set correlation CC(1/2) | 0.999 | 0.999 | 0.775 |
| Completeness | 82.0 | 99.9 | 6.9 |
| Multiplicity | 4.6 | 4.8 | 1.0 |
| | | | |
| Anomalous completeness | 71.3 | 100.0 | 0.4 |
| Anomalous multiplicity | 2.2 | 3.1 | 1.0 |
| DelAnom correlation between half-sets | 0.004 | 0.149 | 0.000 |
| Mid-Slope of Anom Normal Probability | 0.997 | - | - |

# R-factors

$$R_{merge} = \frac{\sum_{hkl} \sum_j \left| I_{hkl,j} - < I_{hkl} > \right|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

The traditional overall measures of quality, but increases with multiplicity although the data improves

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{j=1}^{N} \left| I_{hkl,j} - < I_{hkl} > \right|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

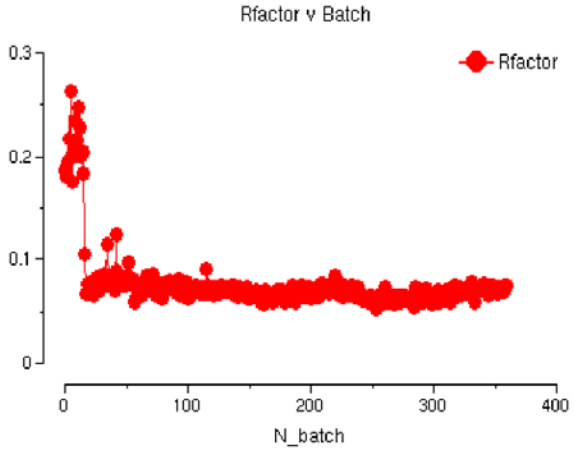
Multiplicity-weighted, better (but larger)

$$R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{j=1}^{N} \left| I_{hkl,j} - < I_{hkl} > \right|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

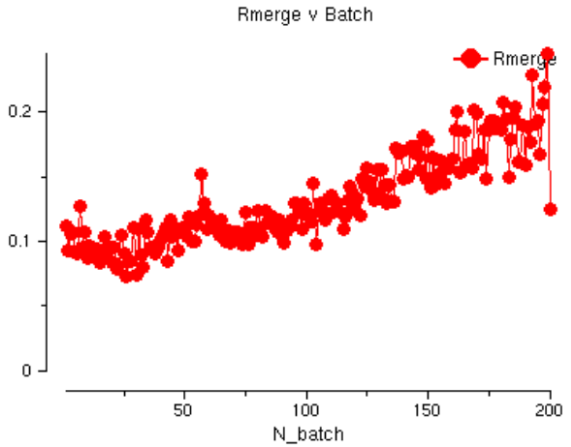
“Precision-indicating R-factor” gets better (smaller) with increasing multiplicity, ie it estimates the precision of the merged <I>

# Rmerge: finding bad batches

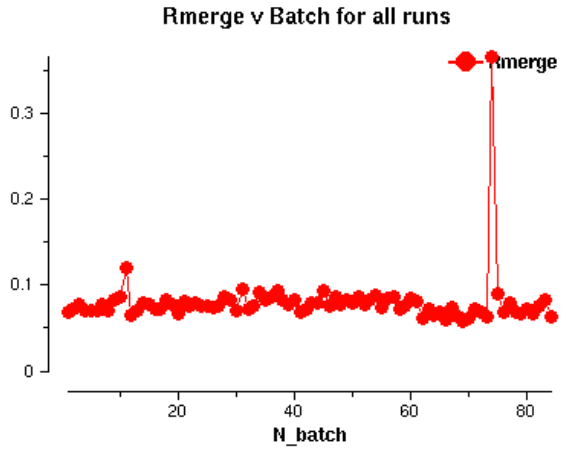
Horribly wrong at beginning



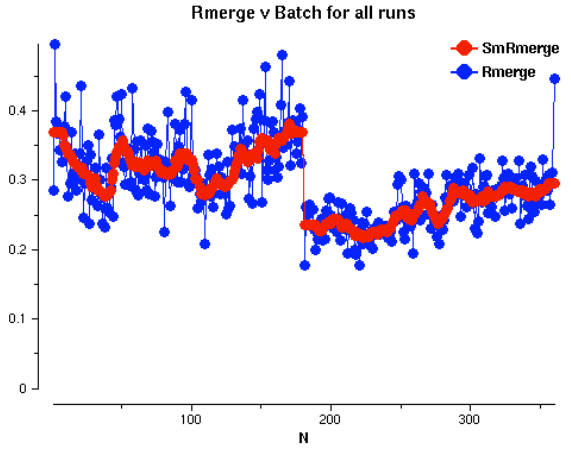

Steady decline in quality




One bad batch



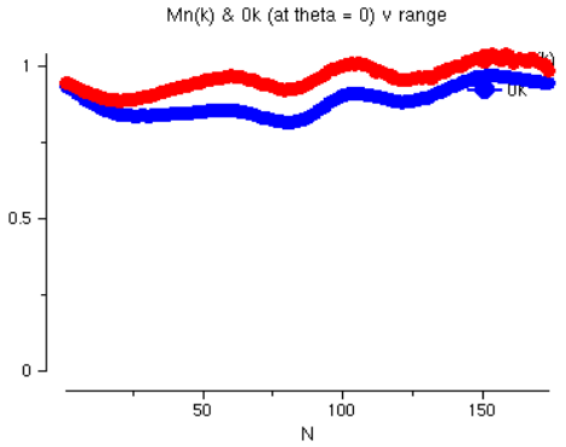

Batches for 2 crystals



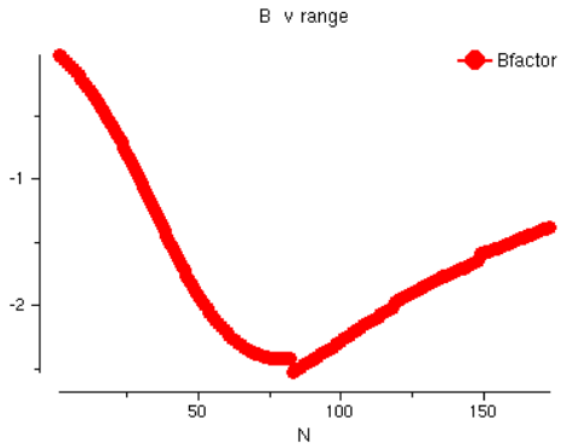

**Would like to have relatively stable Rmerge across all batches**

# Scales and B-factors: radiation damage

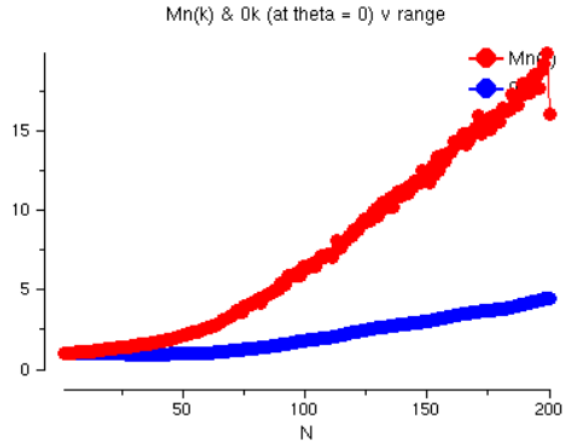
Good: scales uniform



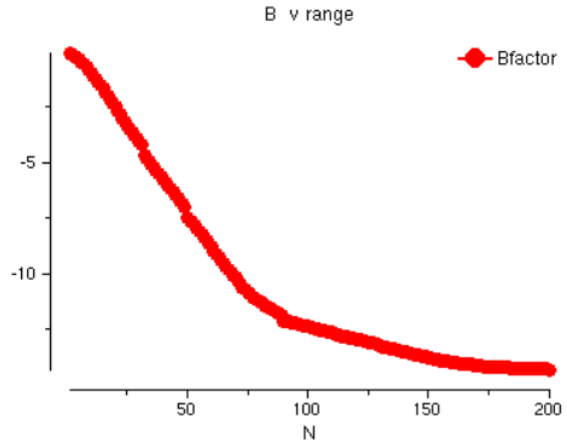

Good: small B-factors




Bad: scales increase sharply




Bad: B-factors large and negative

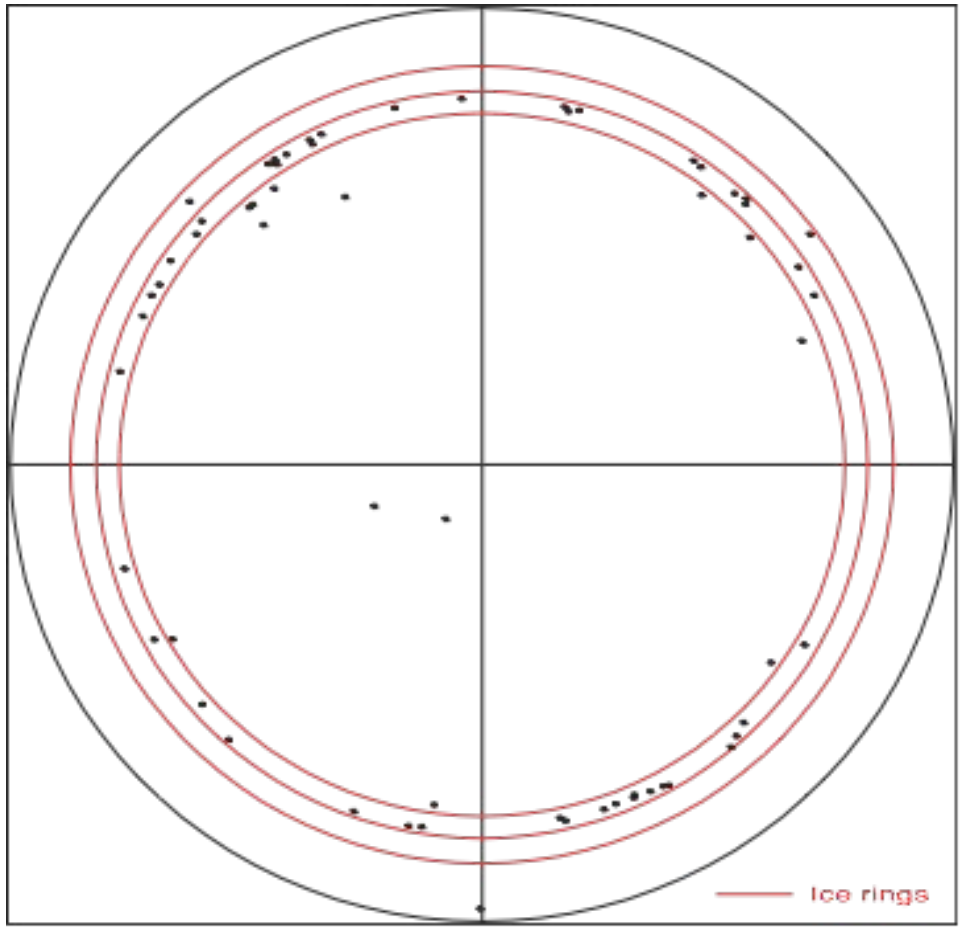



**Ideally have constant scaling factor of 1; except if crystals have an irregular shape.**
**Drop in B factor below -10 indicates radiation damage**

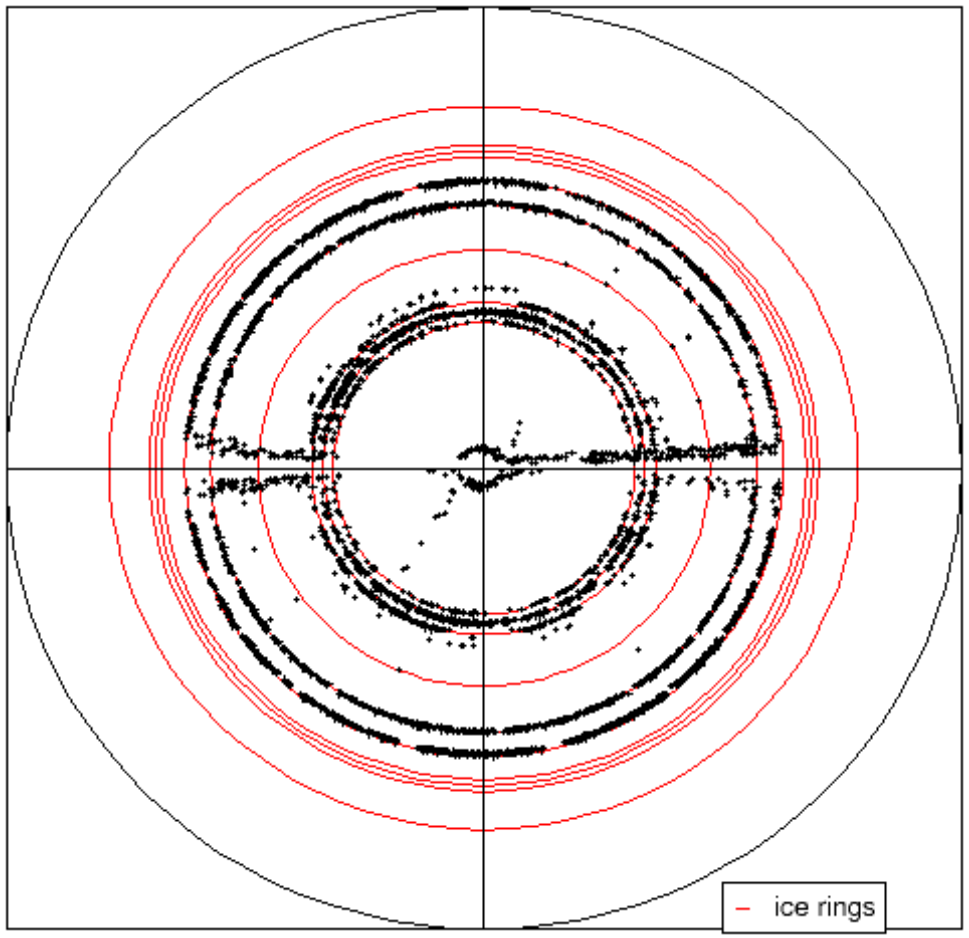
# Outliers: why do we get them?

- outside reliable area of detector (eg behind shadow)
  - specify backstop shadow, calibrate detector
- ice spots
  - do not get ice on your crystal!
- multiple lattices
  - find single crystal
- zingers
- bad prediction (spot not there)
  - improve prediction
- spot overlap
  - lower mosaicity, collect finer sliced data, move detector back, deconvolute overlaps

# Outliers: ROGUEPLOT




A few outliers on ice rings




Lots of reflections on ice rings

# Outliers: number of rejections per image

| N | Run.Rot | MidPhi | Batch | Bfactor | Mn(k) | 0k | Number | NumReject |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.1 | -49.50 | 1 | -0.694 | 1.0651 | 0.9940 | 1703 | 0 |
| 2 | 1.2 | -48.50 | 2 | -0.688 | 1.0622 | 0.9905 | 2193 | 0 |
| 3 | 1.3 | -47.50 | 3 | -0.677 | 1.0564 | 0.9851 | 2219 | 0 |
| 4 | 1.4 | -46.50 | 4 | -0.668 | 1.0453 | 0.9774 | 2202 | 0 |
| 5 | 1.5 | -45.50 | 5 | -0.656 | 1.0339 | 0.9671 | 2198 | 0 |
| 6 | 1.6 | -44.50 | 6 | -0.641 | 1.0180 | 0.9542 | 2217 | 1 |
| 7 | 1.7 | -43.50 | 7 | -0.629 | 1.0017 | 0.9395 | 2208 | 0 |
| 8 | 1.8 | -42.50 | 8 | -0.614 | 0.9811 | 0.9185 | 2217 | 0 |

**Want low number of rejected reflections per image; a maximum of around 5**

# Resolution

**What do we mean by the “resolution” of the data?**

We want to determine the point at which adding another shell of data does not add any “significant” information.

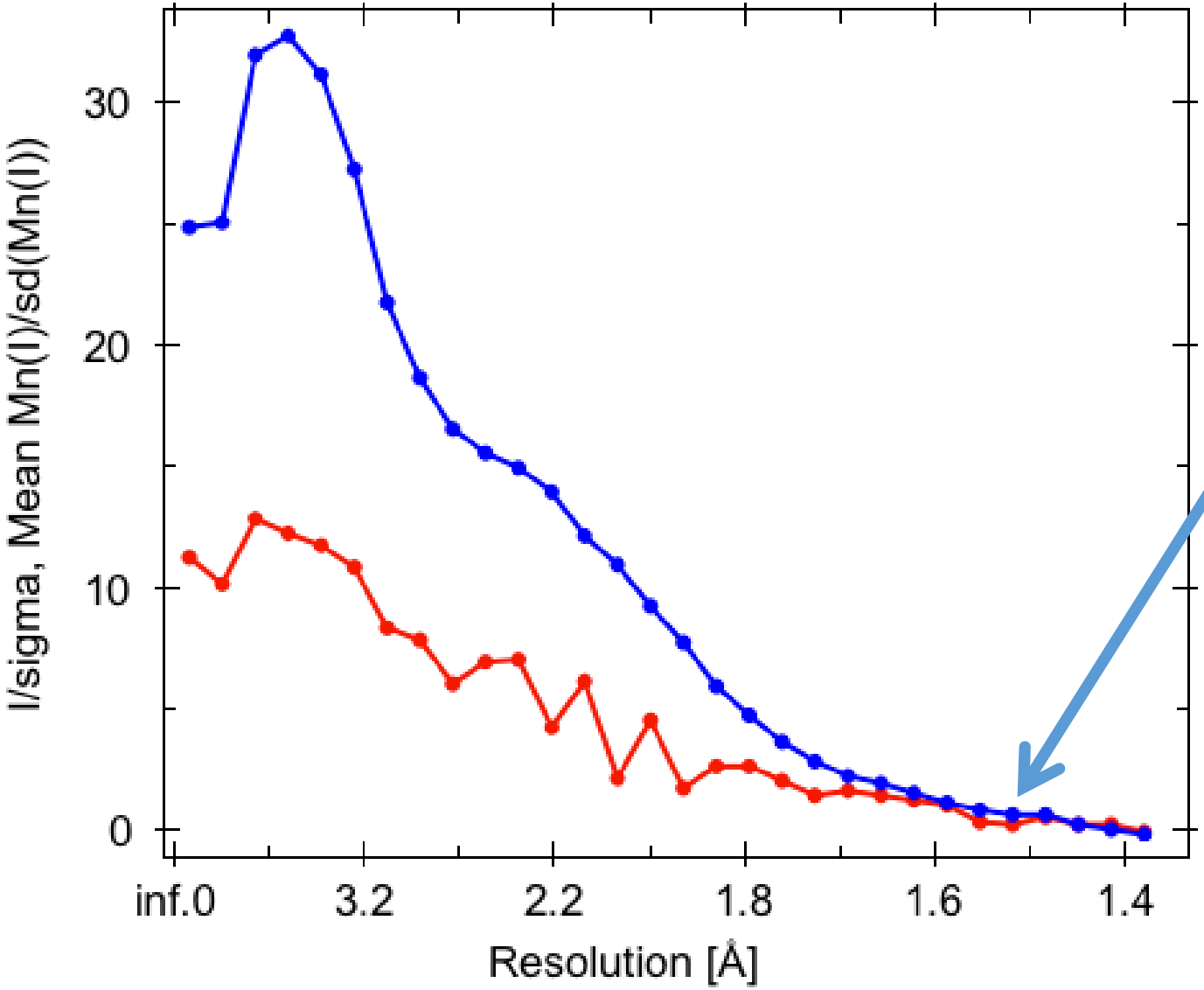
# Resolution

**“Best” resolution is different for different purposes, so don’t cut it too soon**

- Experimental phasing: substructure location is generally unweighted, so cut back conservatively to data with high signal/noise ratio. For phasing, use all “reasonable” data
- Molecular replacement: Phaser uses likelihood weighting, but there is probably no gain in using the very weak high resolution data
- Model building and refinement: if everything is perfectly weighted (perfect error models!), then extending the data should do no harm and may do good

**There is no reason to suppose that cutting back the resolution to satisfy referees will improve your model!**

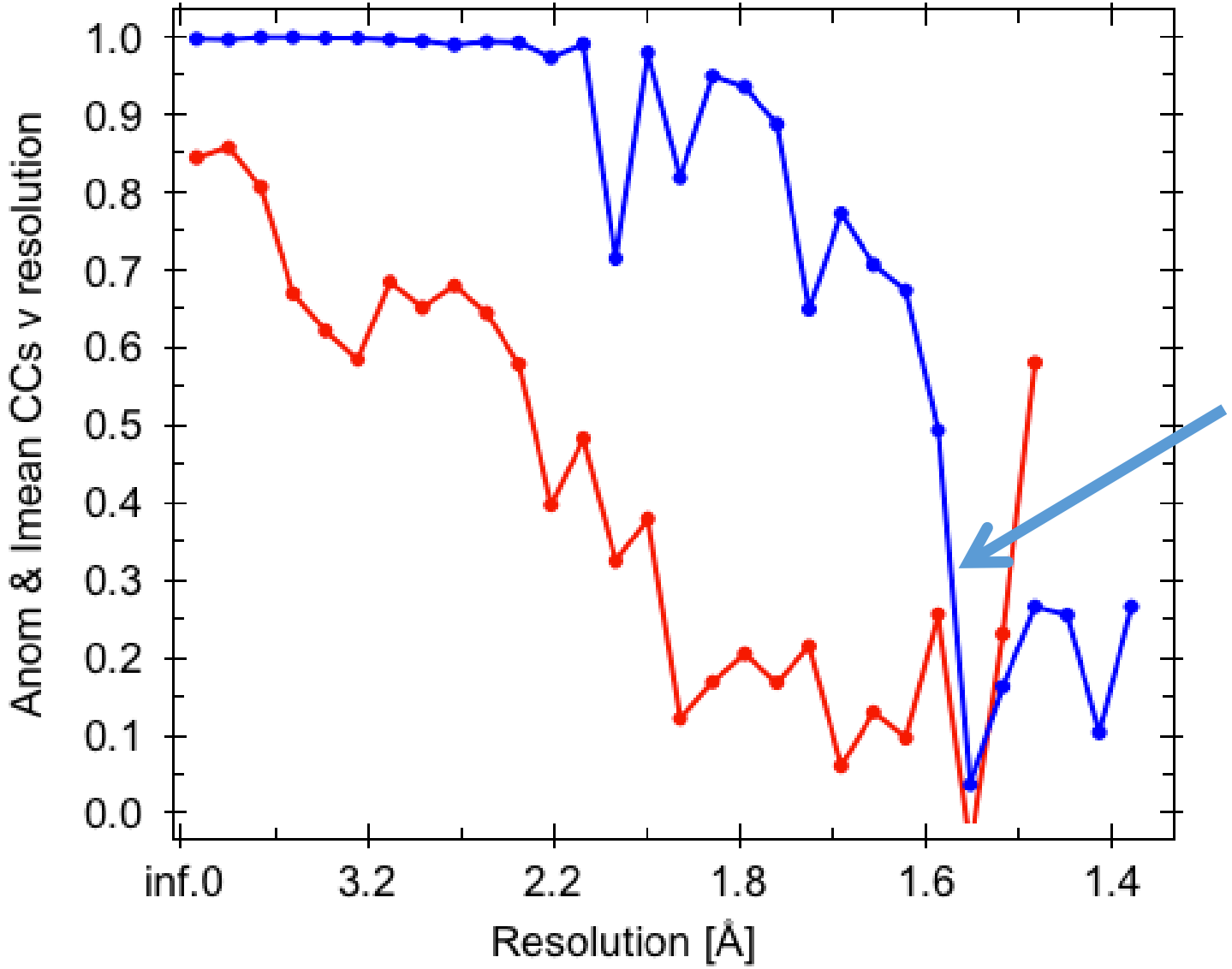
# Resolution: I/sig(I)




**I/sig(I) around 1.5**

A reasonably good criterion, but it relies on σ(I), which is not entirely reliable

# Resolution: CC 1/2



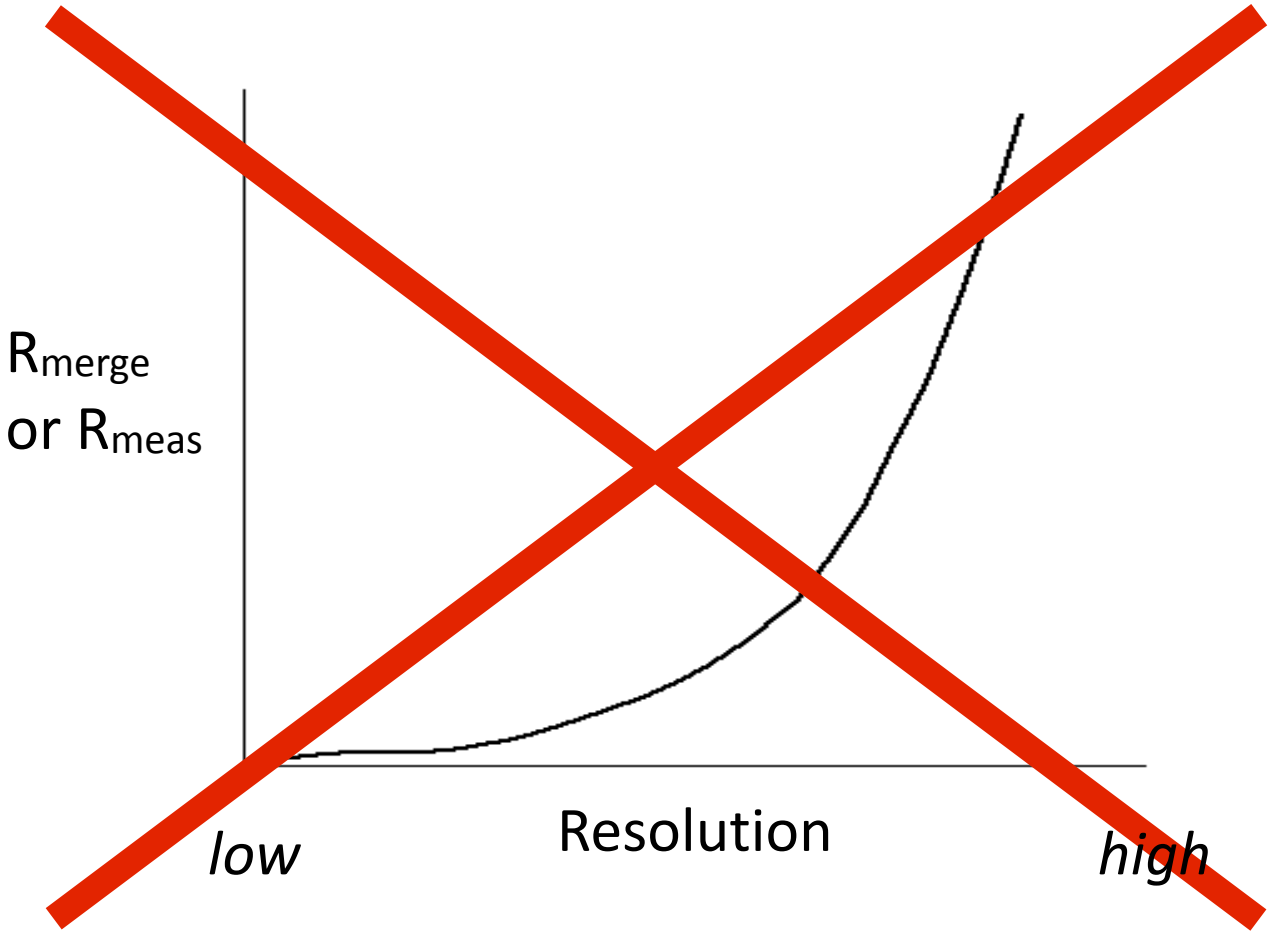

**CC ½ around 0.3**

Split observations for each reflection randomly into 2 halves, and calculate the correlation coefficient between them

Advantages:

- Clear meaning to values (1.0 is perfect, 0 is no correlation) , known statistical properties
- Independent of σ(I)

# Resolution: Rmerge/Rmeas

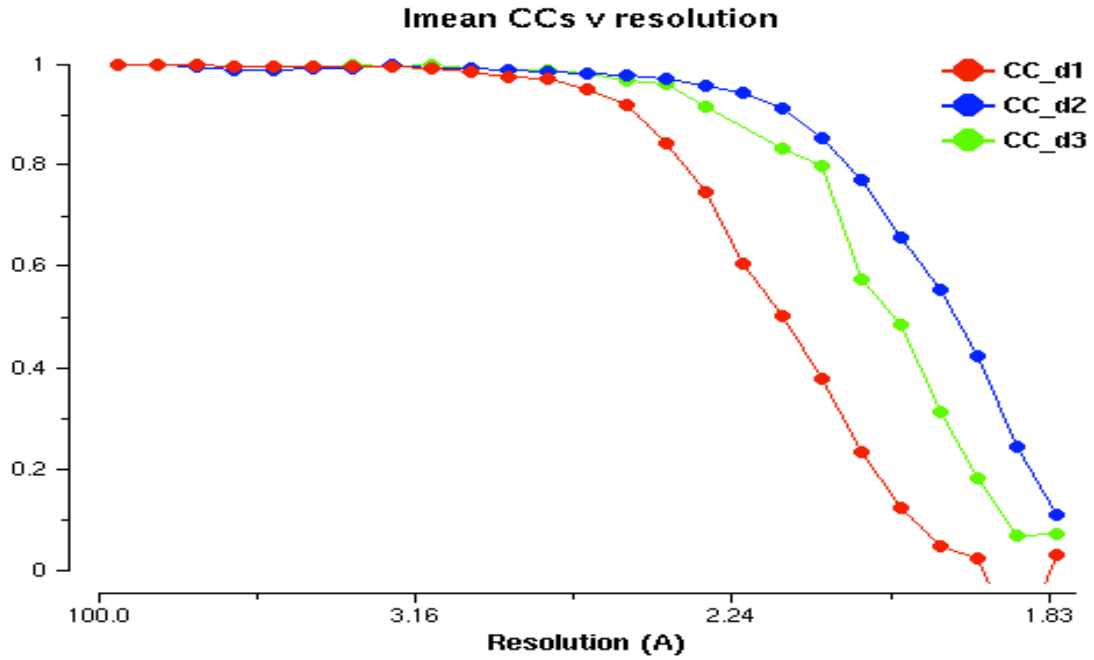


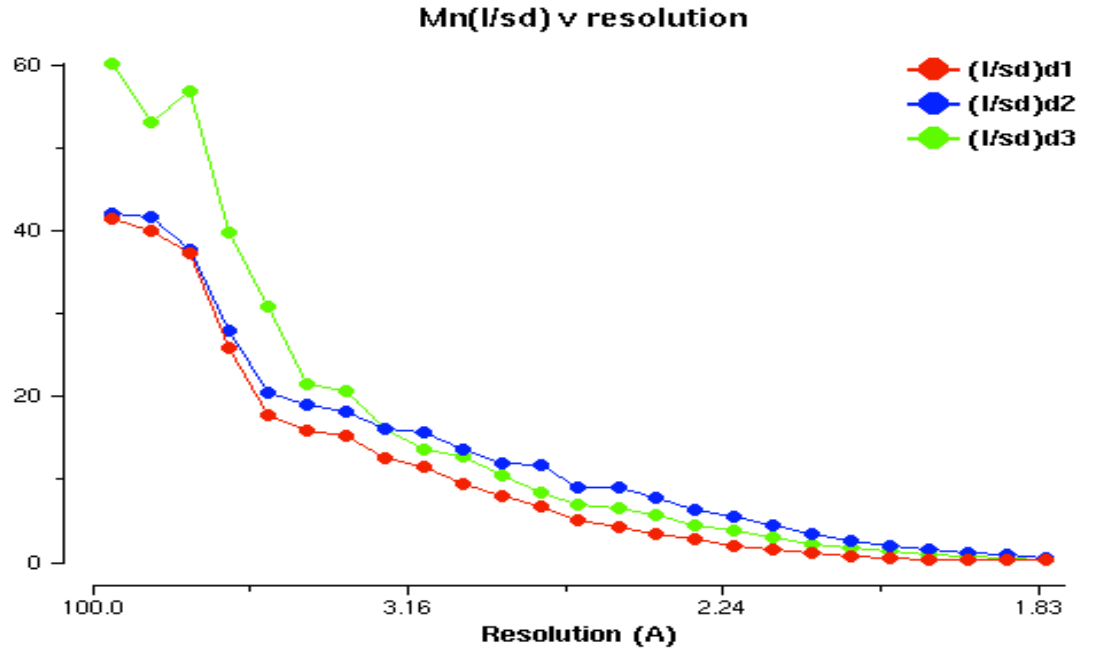

Note that $R_{merge}$ and $R_{meas}$ are useful for other purposes, but not for deciding the resolution cutoff

Note that the crystallographic R-factor behaves quite differently: at higher resolution as the data become noisier, $R_{cryst}$ tends to a constant value, not to infinity

# Resolution: anisotropy

- Many (perhaps most) datasets are *anisotropic*
- The principal directions of anisotropy are defined by symmetry (axes or planes), except in the monoclinic and triclinic systems, in which we can calculate the orthogonal principle directions
- We can then analyse half-dataset CCs or <I/σ(I)> in cones around the principle axes, or as projections on to the axes
- Anisotropic cutoffs are probably a Bad Thing, since it leads to strange series termination errors and problem with intensity statistics



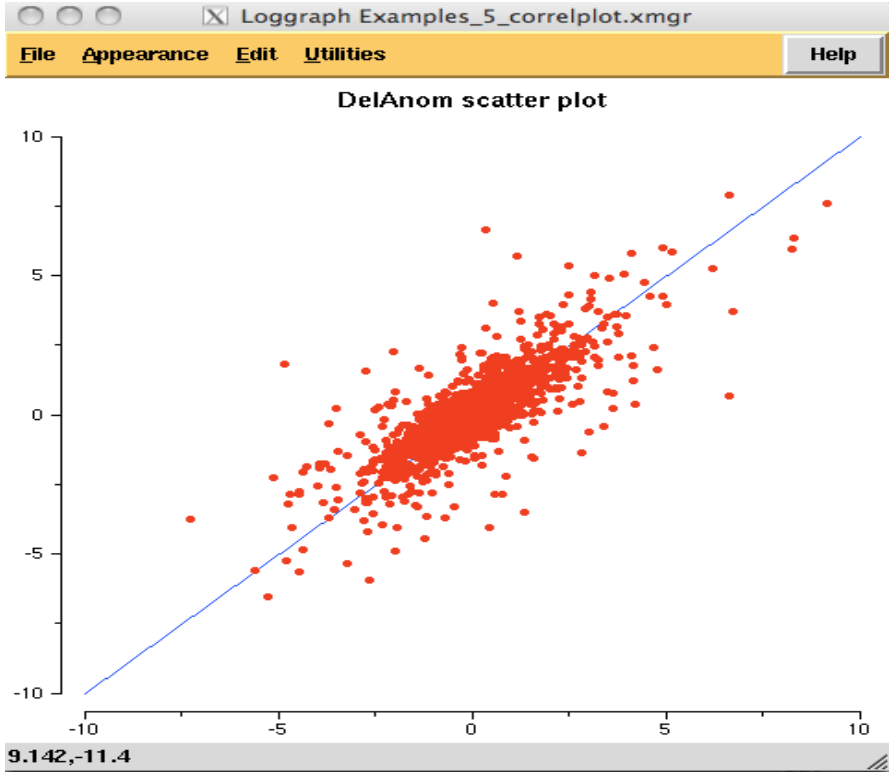
# Resolution: aimless log file

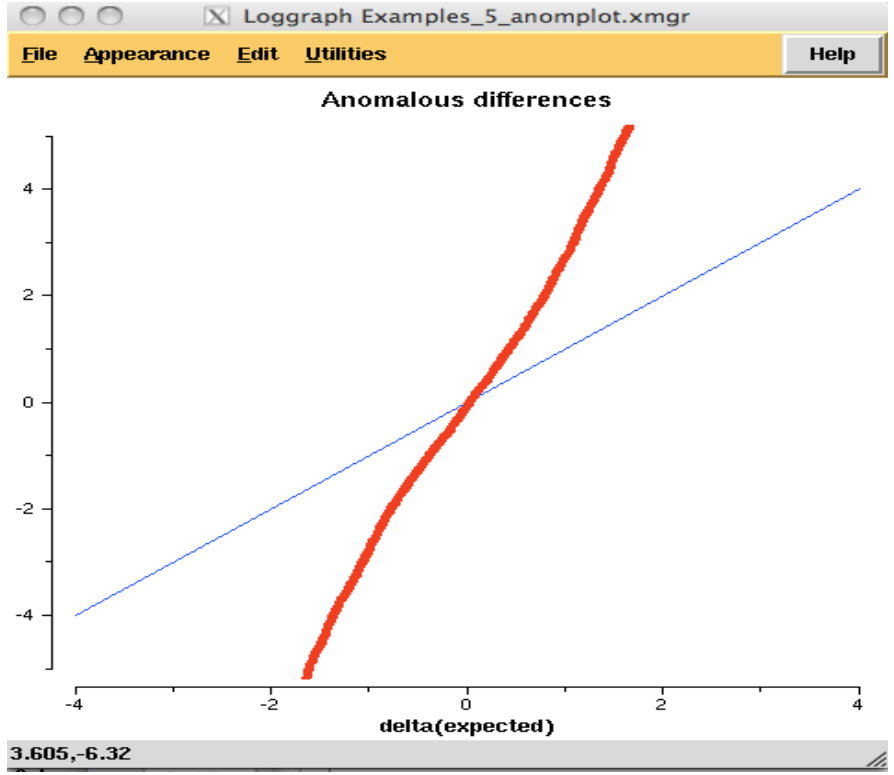
```
Estimates of resolution limits: overall
   from half-dataset correlation CC(1/2) >  0.30: limit =  3.15A
   from Mn(I/sd) >  1.50:                        limit =  3.17A
   from Mn(I/sd) >  2.00:                        limit =  3.30A

Estimates of resolution limits in reciprocal lattice directions:
  Along h k plane
   from half-dataset correlation CC(1/2) >  0.30: limit =  3.42A
   from Mn(I/sd) >  1.50:                        limit =  3.31A
  Along l axis
   from half-dataset correlation CC(1/2) >  0.30: limit =  3.00A  == maximum resolution
   from Mn(I/sd) >  1.50:                        limit =  3.00A  == maximum resolution
```

# Anomalous signal

- The data contains both I+ (hkl) and I- (-h-k-l) observations and we can detect whether there is a significant difference between them.
  - Split one dataset randomly into two halves, calculate correlation between the two halves or
  - compare different wavelengths (MAD)

# Anomalous signal: strong

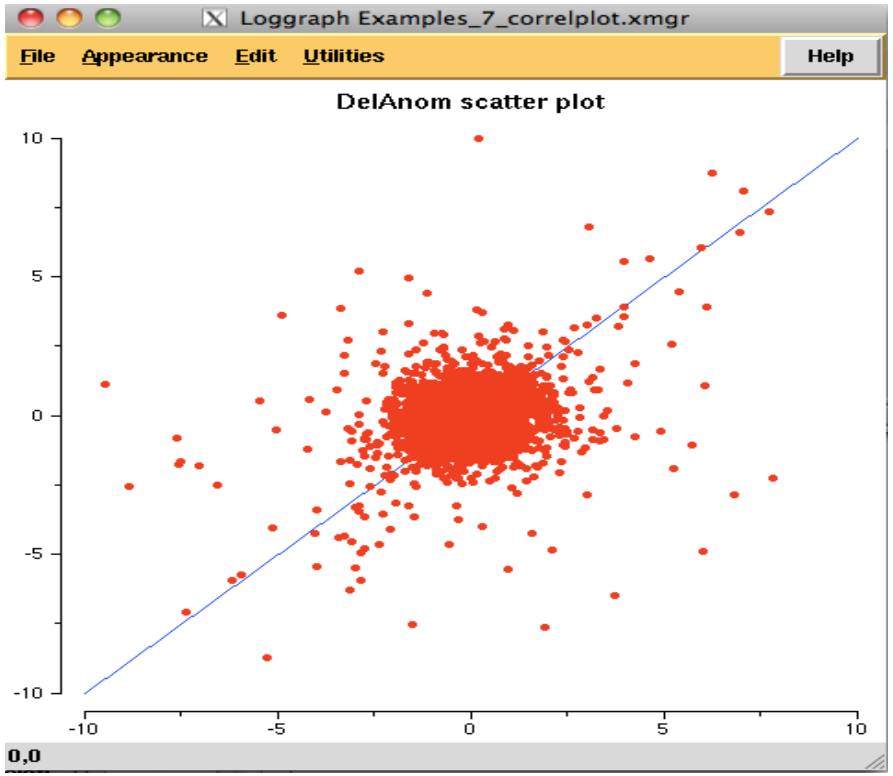


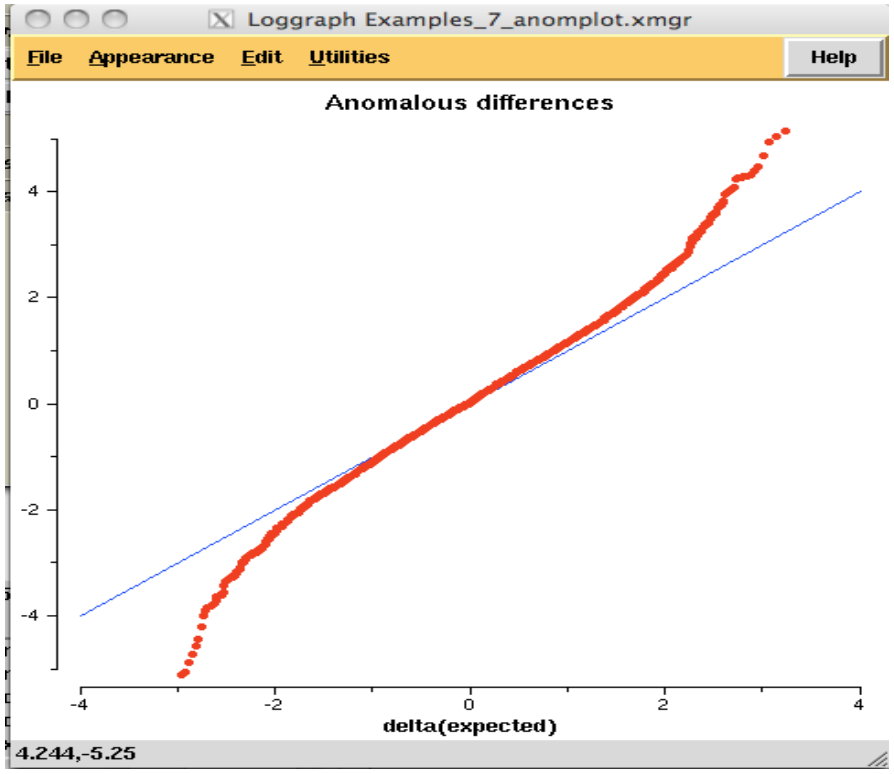




Plot $\Delta I_1$ against $\Delta I_2$ should be elongated along diagonal

Slope > 1.0 means that $\Delta I > \sigma$

# Anomalous signal: weak but useful

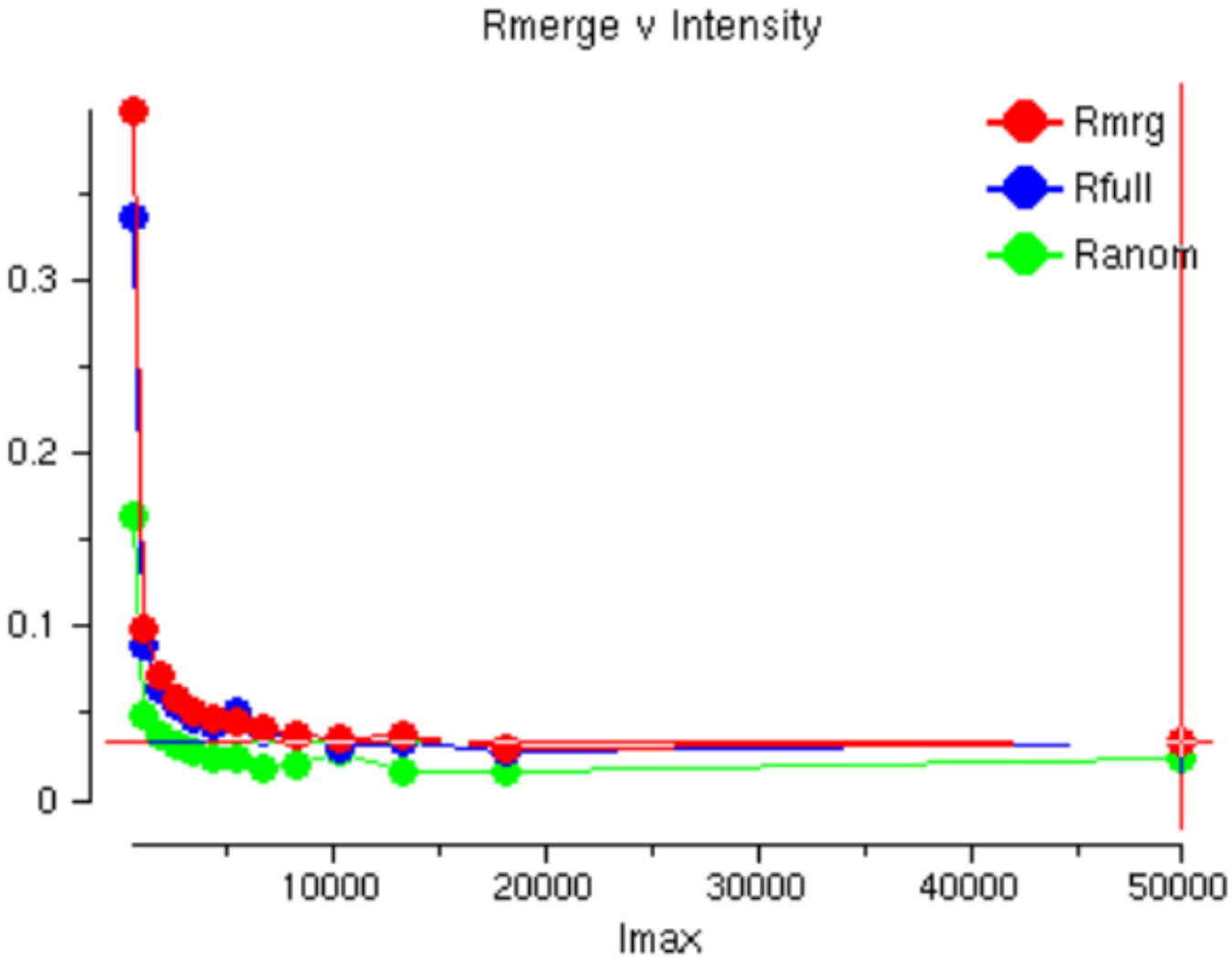



Plot $\Delta I_1$ against $\Delta I_2$ should be elongated along diagonal

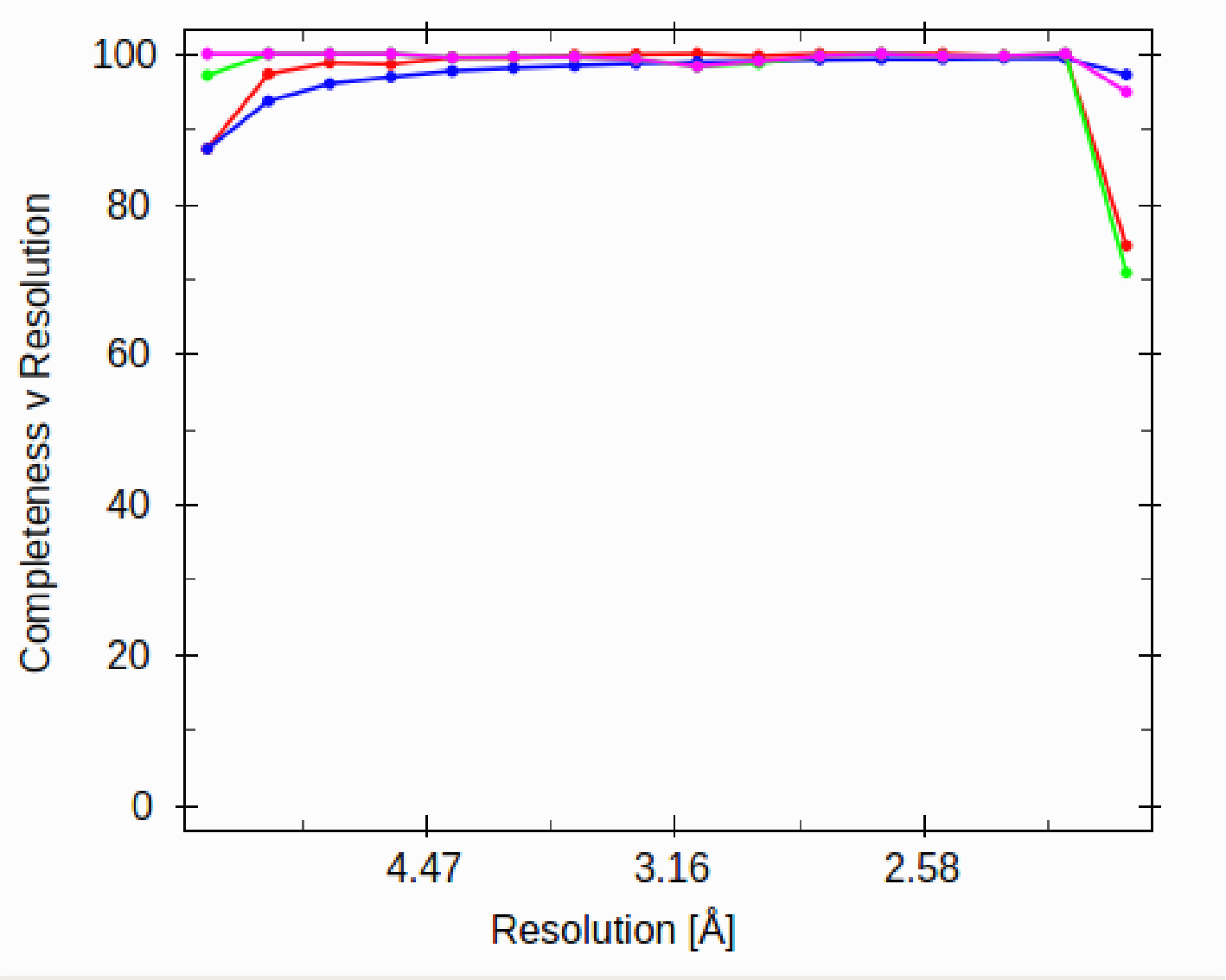
Slope > 1.0 means that $\Delta I > \sigma$

# Data Quality: Rmerge vs intensity




Rmerge is always large for small intensities. For large intensities it should be in the range 0.01 to 0.04 for good data. Larger values suggest that there are systematic errors.

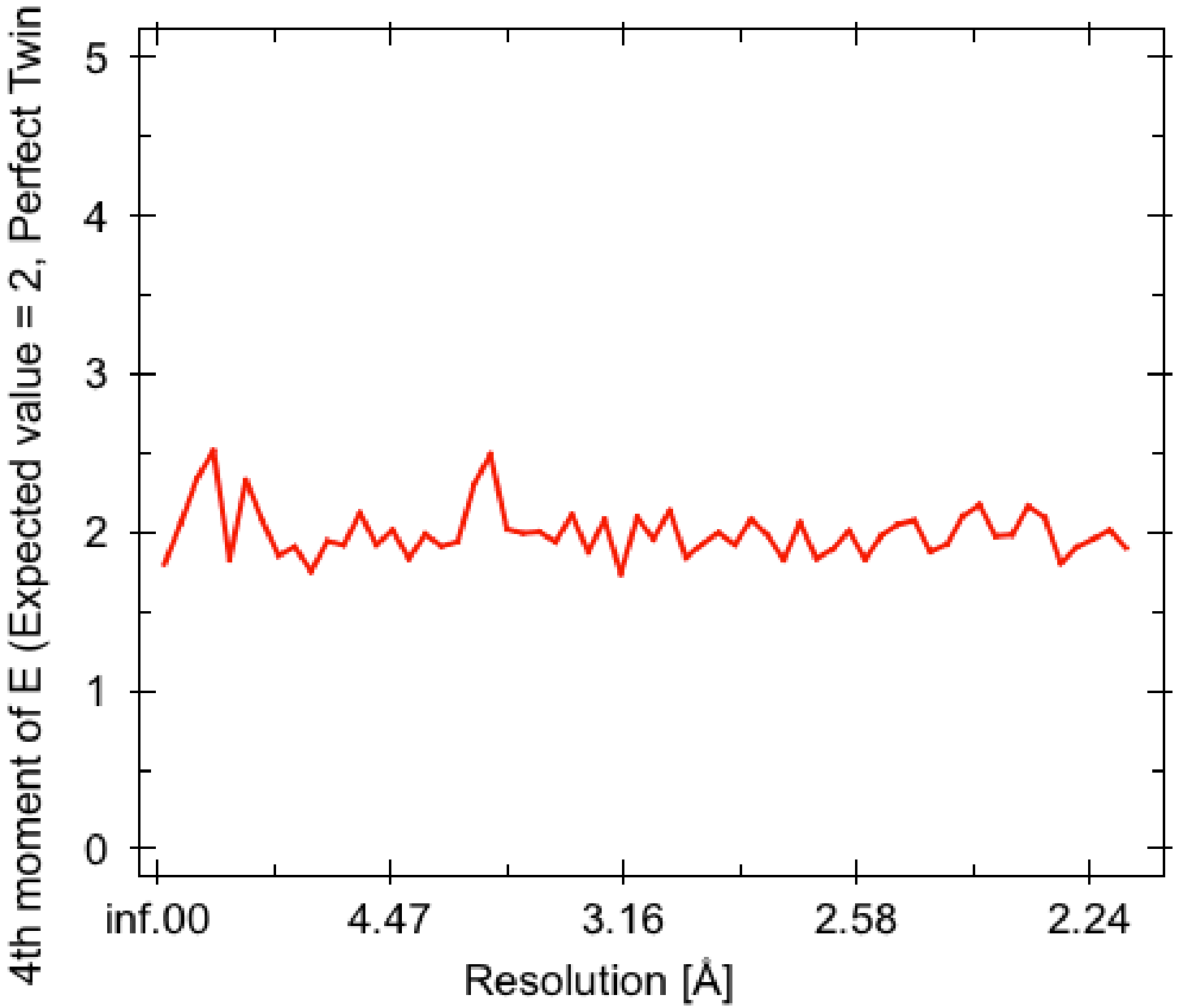
# Data Quality: completeness




Completeness of data should be as close to 100% as possible.
Watch out for data with < 95% completeness.
Some loss of completeness can be tolerated in the outermost resolution bins.
If you integrate to the corners of the detector, you may have low completeness at high resolution.

# Detecting twinning

- Depends on moments of intensity distributions
- Acentric $E^4$ is useful: if 2 probably not twinned, if 1.5 probably twinned
- Measures the spread of the merged intensity distribution
- Look at ctruncate output
- More twinning tests are performed, check ctruncate log


4th moment of E (Expected value = 2, Perfect Twin
5
4
3
2
1
0
inf.00
4.47
3.16
2.58
2.24
Resolution [Å]

# Things that might look like twinning but are not

Translational non-crystallographic symmetry:

- A whole classes of reflections may be weak eg h odd with a NCS translation of ~1/2, 0 0. <I> over all reflections is misleading, so Z values are inappropriate. The reflection classes should be separated (not yet done)

Anisotropy: <I> is misleading so Z values are wrong

- ctruncate applies an anisotropic scaling before analysis

Weak data: the ideal statistics are based on perfect data.

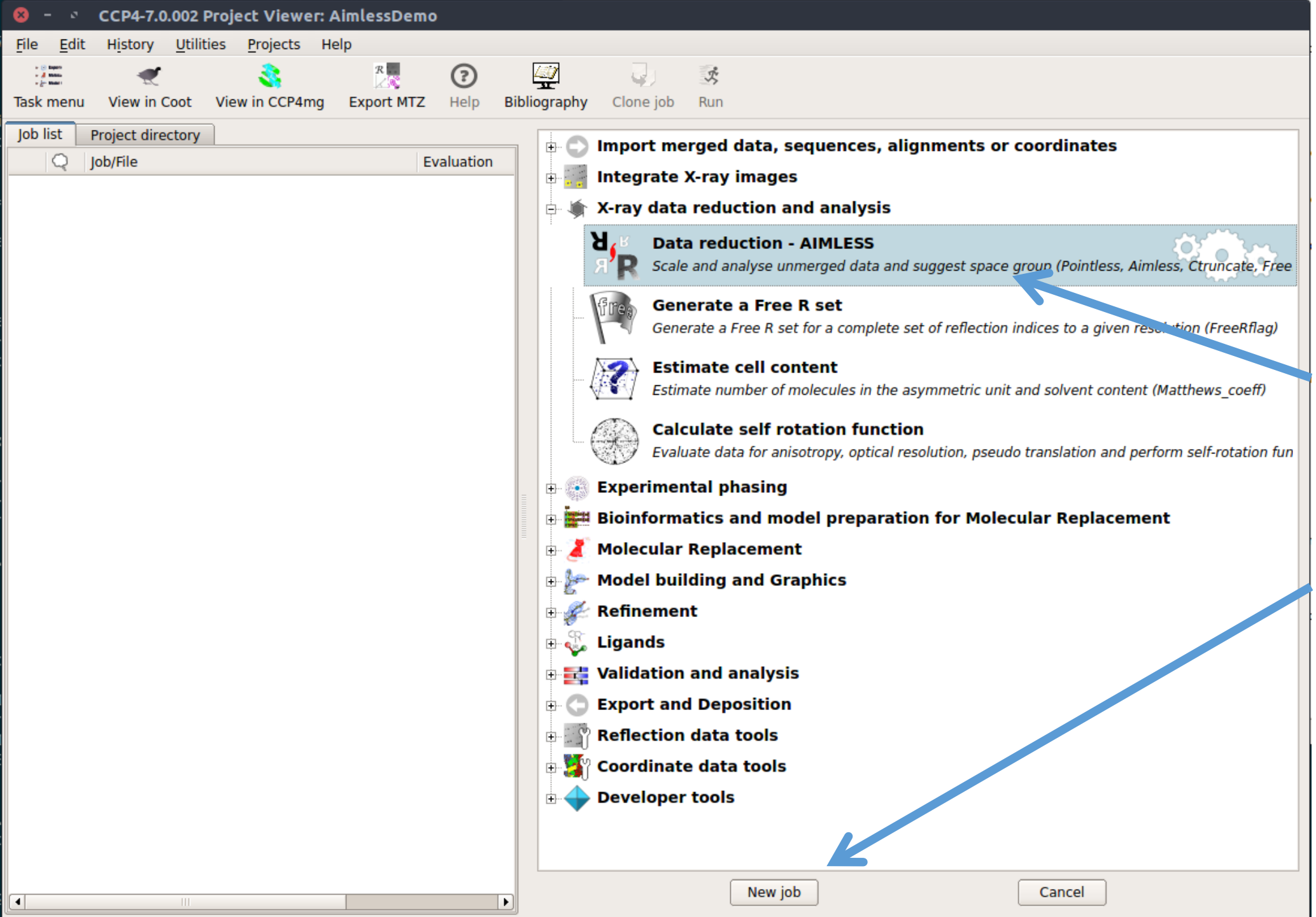
- If the signal/noise ratio is small, then the statistics may falsely suggest twinning

Systematic over-estimation of reflection intensities

- With overlapping spots, strong reflections can inflate the value of weak neighbours, leading to too few weak reflections
- Bad outlier rejection for background determination. If background is systematically underestimated, reflections are systematically overestimated (mostly occurs in very weak data).

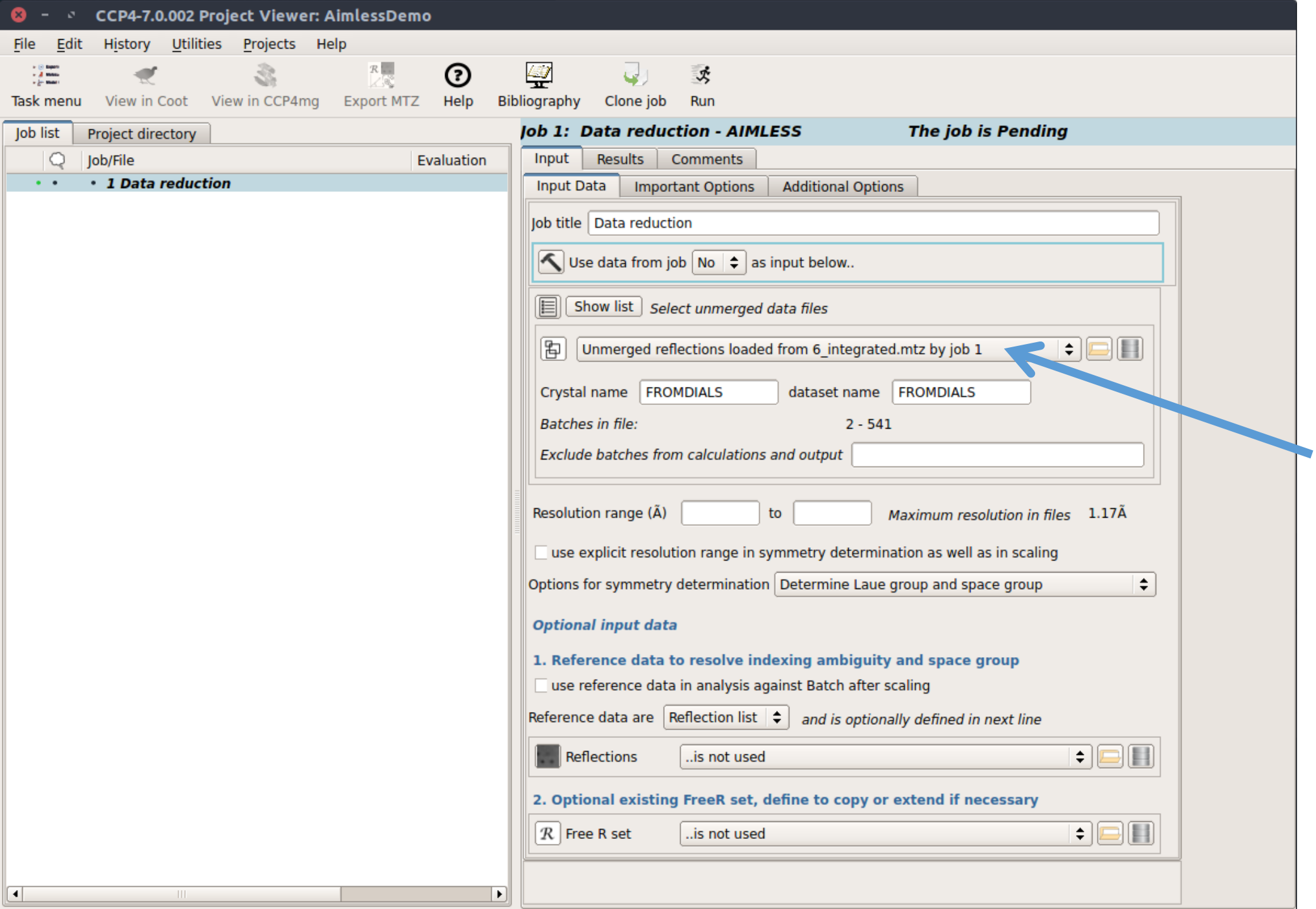
# Data reduction using CCP4 I2

Data reduction with POINTLESS and AIMLESS

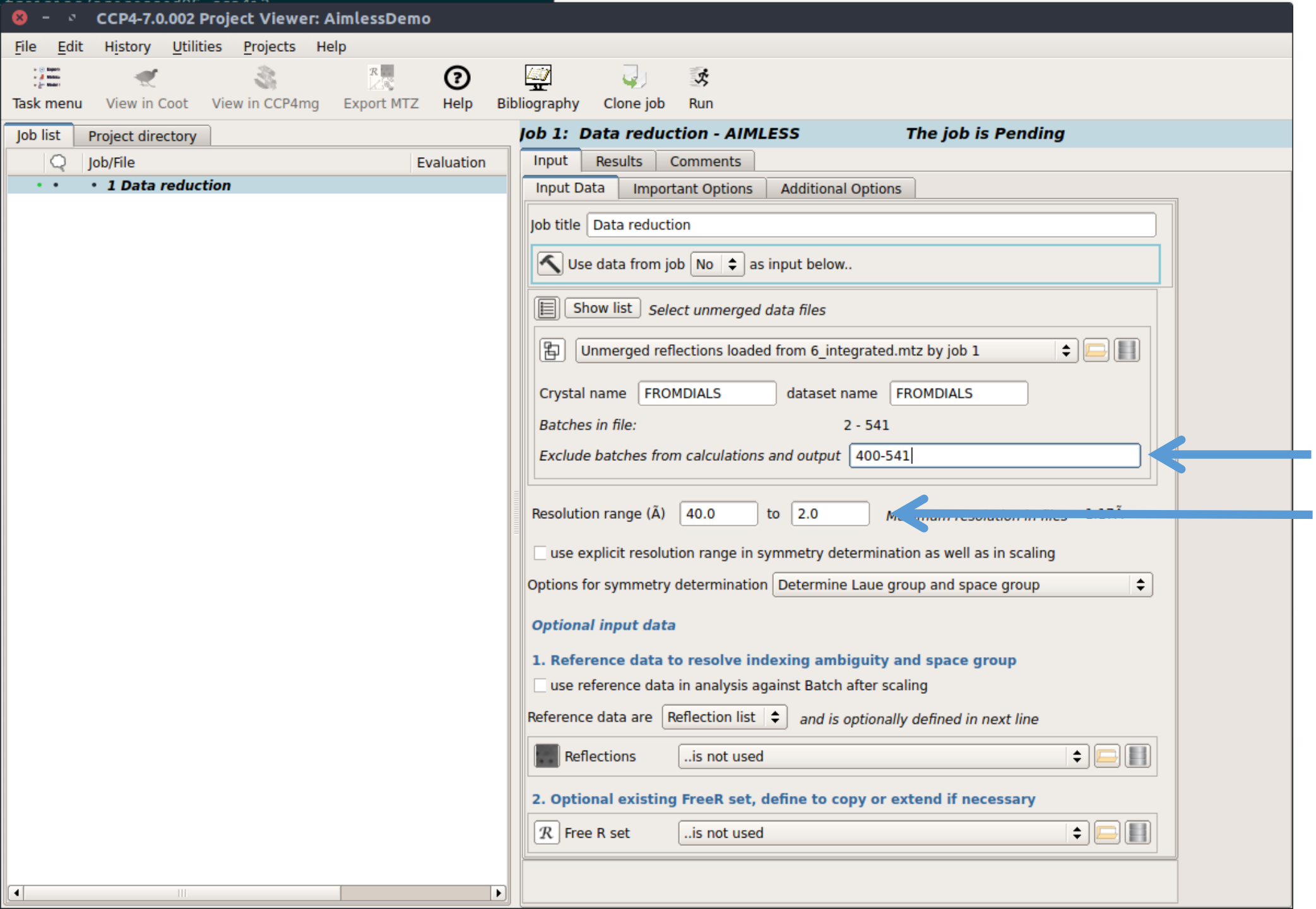



Click the aimless data reduction job item.

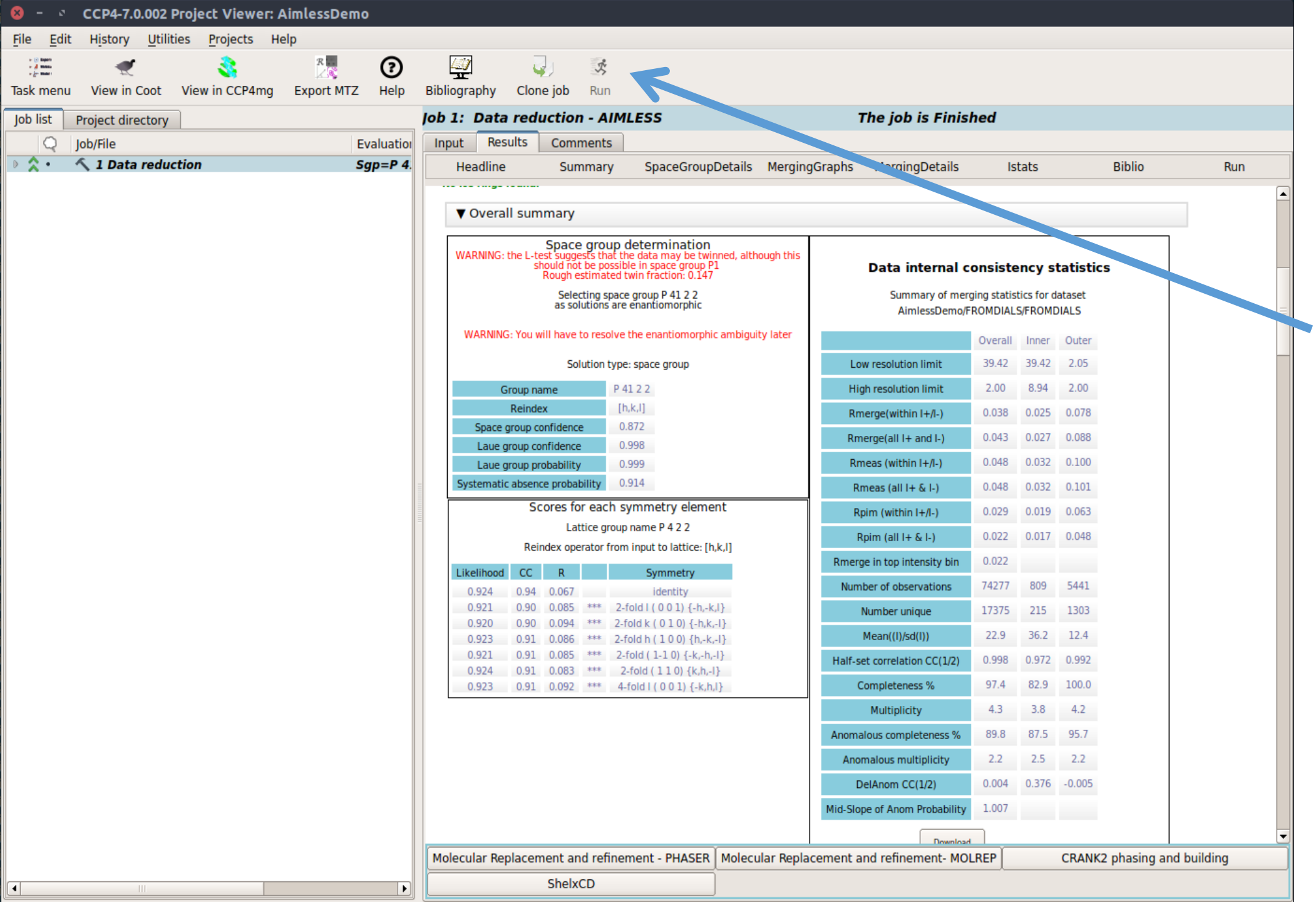
Click “new job” to open the aimless job window.



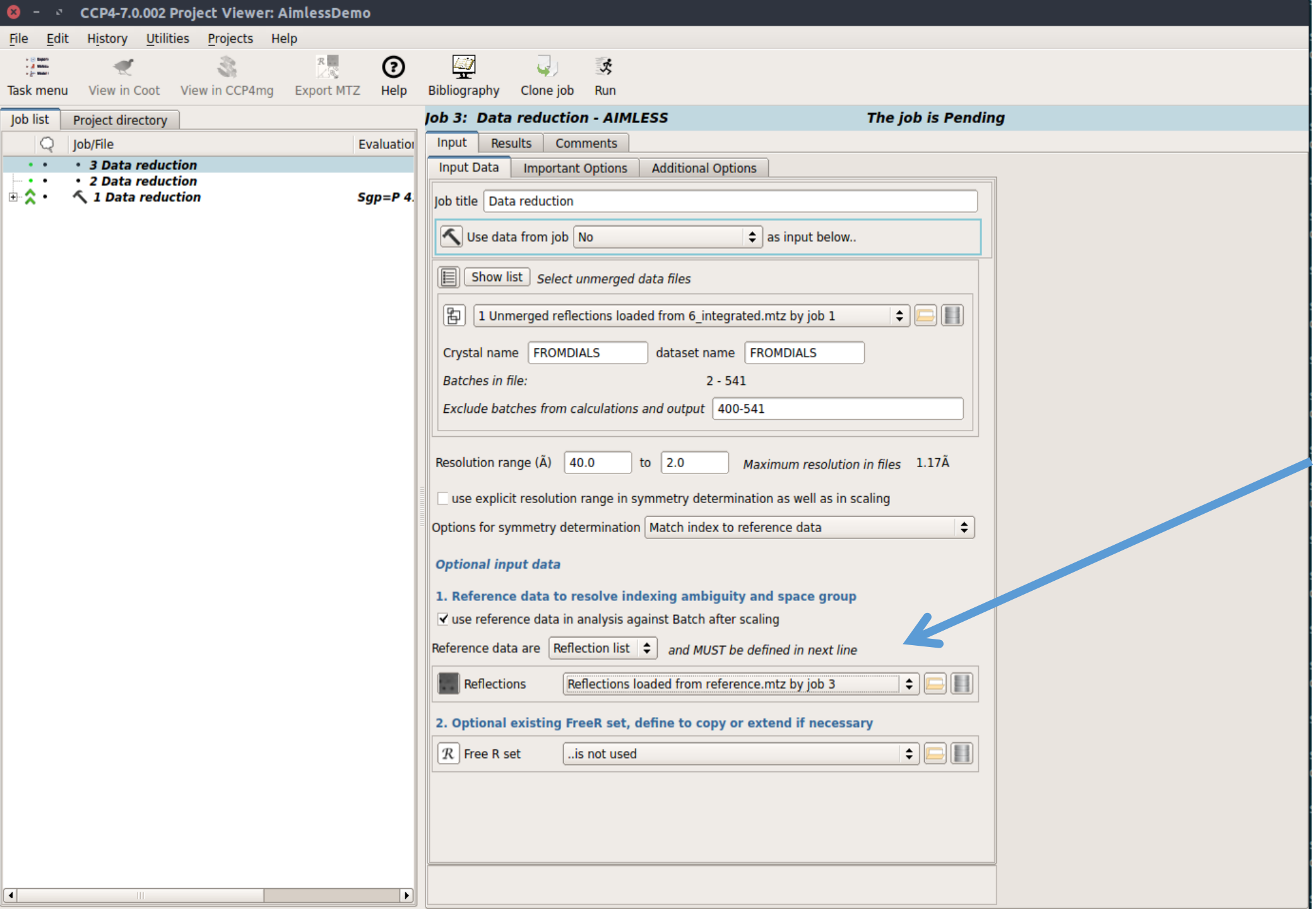

Select an MTZ file containing integrated reflections from MOSFLM, DIALS or XDS etc




If necessary, exclude batches or set a resolution range for scaling.



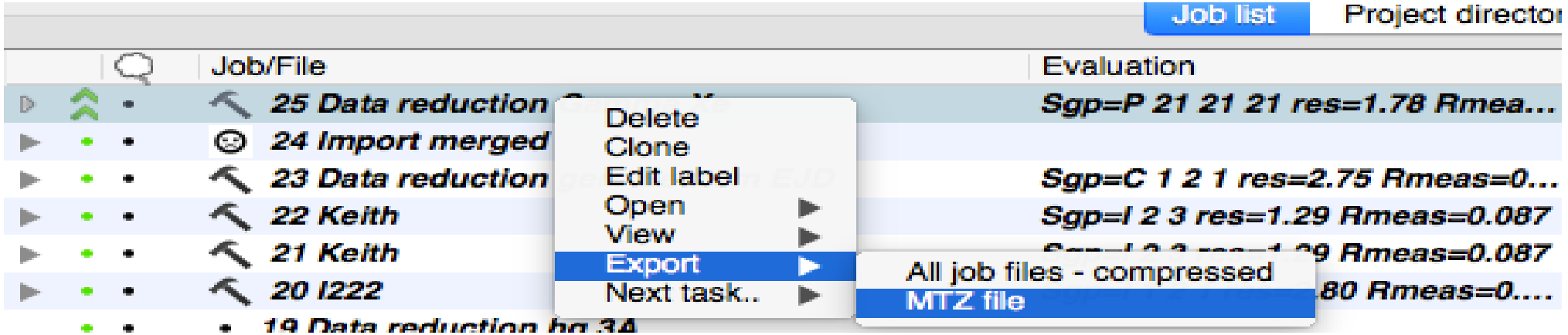

To execute the job, click “Run”. When the job has finished, results will be presented.




To select a reference MTZ file to resolve indexing ambiguity, select “Reflection list” and specify the reference reflection file.

# Exporting from I2

Right-click the finished job in the Job list and choose Export -> MTZ file



# Using the command line

```
$ pointless < pointless.dat | tee pointless.log

--- contents of pointless.dat ---
HKLIN integrated.mtz
HKLOUT unscaled.mtz
HKLREF reference.mtz                    # optional

$ aimless < aimless.dat | tee aimless.log

--- contents of aimless.dat ---
HKLIN unscaled.mtz
HKLOUT scaled.mtz
RESOLUTION HIGH 2.0                     # optional
EXCLUDE BATCH 450 TO 500                # optional
```

# Summary

- *Do look critically at the data processing statistics*

- What is the point group (Laue group)?
- What is the space group?
- Was the crystal dead at the end?
- Is the dataset complete?
- Do you want to cut back the resolution?
- Is this the best dataset so far for this project?
- Should you merge data from multiple crystals?
- Is there anomalous signal (if you expect one)?
- Are the data twinned?

**Try alternative processing strategies: different choices of cutoffs, merging crystals, etc.**
**Data processing is not necessarily something you just do once.**

# Thank you for listening!

**http://www.ccp4.ac.uk**