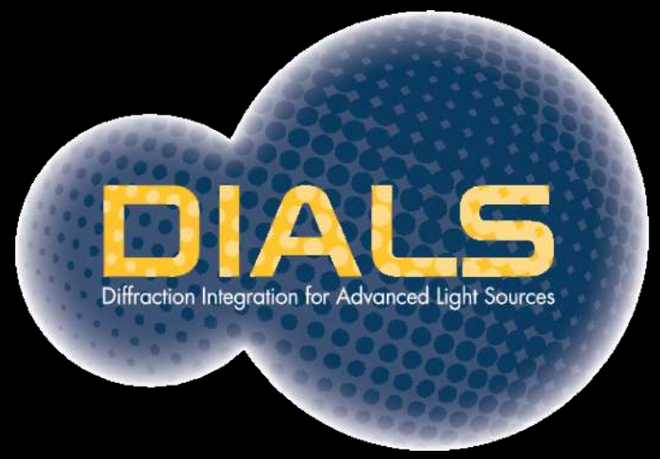
Data reduction: scaling and merging data

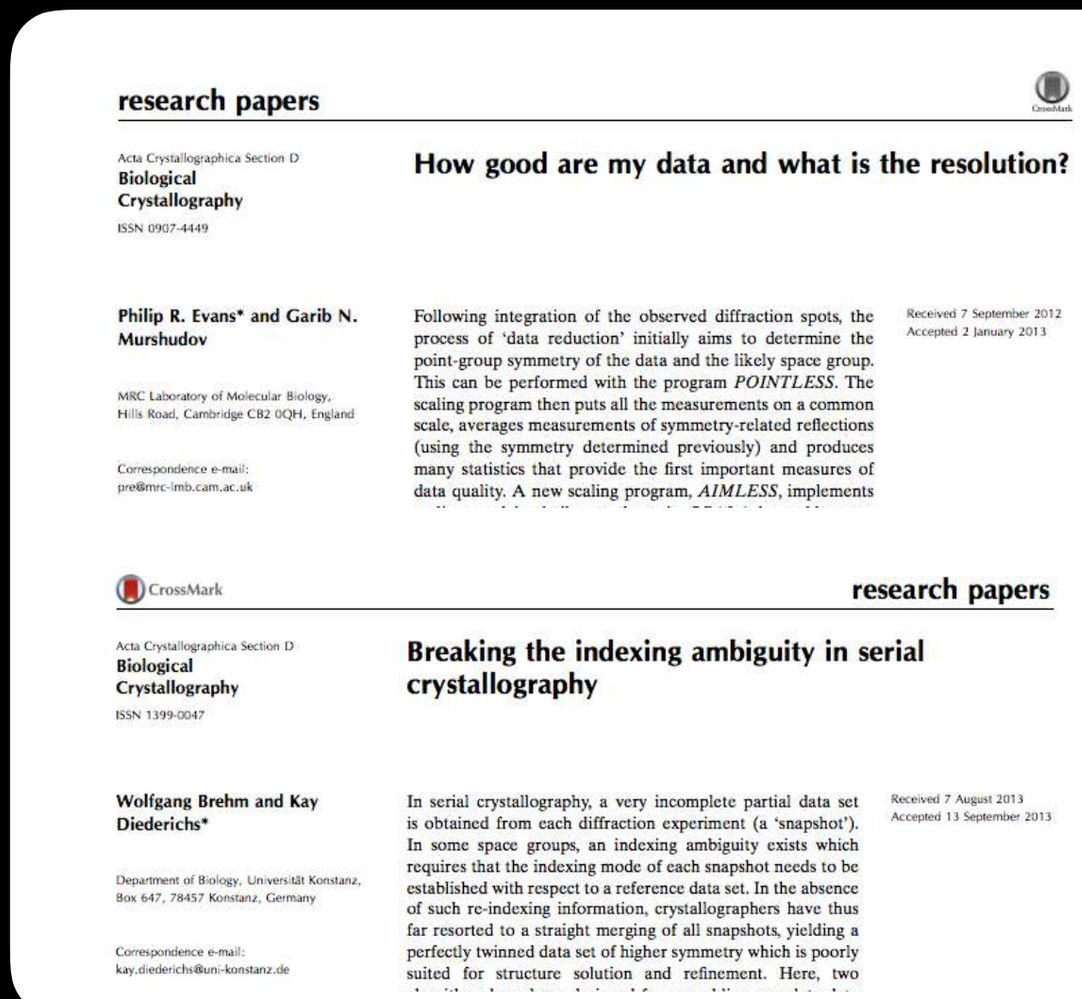
James Beilsten-Edmands
Diamond Light Source, U.K.



DATA REDUCTION

General principles shared across different data processing packages.

Grateful to other methods developers who share their methodologies in the literature.



research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

How good are my data and what is the resolution?

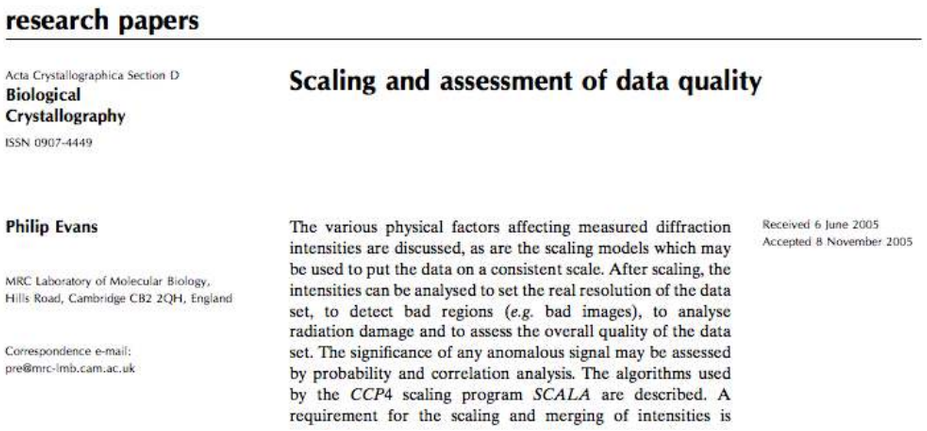
Philip R. Evans* and Garib N. Murshudov

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, England

Correspondence e-mail: pre@mrc-lmb.cam.ac.uk

Received 7 September 2012
Accepted 2 January 2013

Following integration of the observed diffraction spots, the process of 'data reduction' initially aims to determine the point-group symmetry of the data and the likely space group. This can be performed with the program *POINTLESS*. The scaling program then puts all the measurements on a common scale, averages measurements of symmetry-related reflections (using the symmetry determined previously) and produces many statistics that provide the first important measures of data quality. A new scaling program, *AIMLESS*, implements



research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Scaling and assessment of data quality

Philip Evans

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

Correspondence e-mail: pre@mrc-lmb.cam.ac.uk

Received 6 June 2005
Accepted 8 November 2005

The various physical factors affecting measured diffraction intensities are discussed, as are the scaling models which may be used to put the data on a consistent scale. After scaling, the intensities can be analysed to set the real resolution of the data set, to detect bad regions (*e.g.* bad images), to analyse radiation damage and to assess the overall quality of the data set. The significance of any anomalous signal may be assessed by probability and correlation analysis. The algorithms used by the *CCP*4 scaling program *SCALA* are described. A requirement for the scaling and merging of intensities is

CrossMark

research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Breaking the indexing ambiguity in serial crystallography

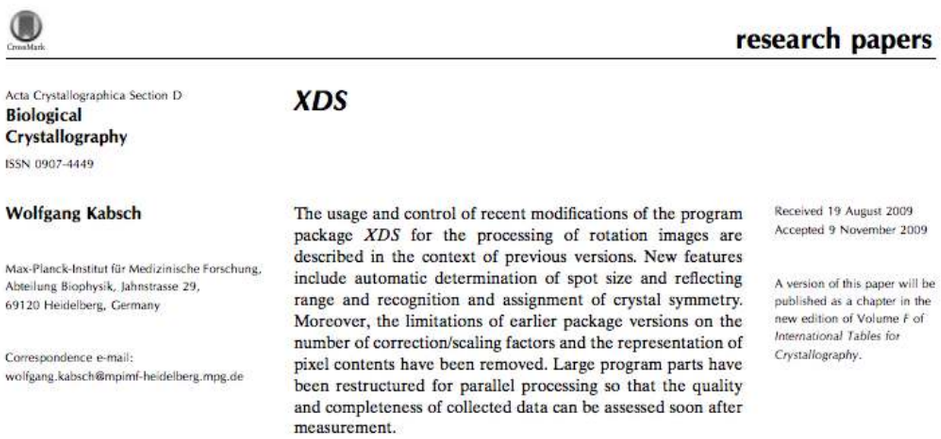
Wolfgang Brehm and Kay Diederichs*

Department of Biology, Universität Konstanz, Box 647, 78457 Konstanz, Germany

Correspondence e-mail: kay.diederichs@uni-konstanz.de

Received 7 August 2013
Accepted 13 September 2013

In serial crystallography, a very incomplete partial data set is obtained from each diffraction experiment (a 'snapshot'). In some space groups, an indexing ambiguity exists which requires that the indexing mode of each snapshot needs to be established with respect to a reference data set. In the absence of such re-indexing information, crystallographers have thus far resorted to a straight merging of all snapshots, yielding a perfectly twinned data set of higher symmetry which is poorly suited for structure solution and refinement. Here, two



research papers

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

XDS

Wolfgang Kabsch

Max-Planck-Institut für Medizinische Forschung, Abteilung Biophysik, Jahnstrasse 29, 69120 Heidelberg, Germany

Correspondence e-mail: wolfgang.kabsch@mpimf-heidelberg.mpg.de

Received 19 August 2009
Accepted 9 November 2009

A version of this paper will be published as a chapter in the new edition of Volume *F* of *International Tables for Crystallography*.

The usage and control of recent modifications of the program package *XDS* for the processing of rotation images are described in the context of previous versions. New features include automatic determination of spot size and reflecting range and recognition and assignment of crystal symmetry. Moreover, the limitations of earlier package versions on the number of correction/scaling factors and the representation of pixel contents have been removed. Large program parts have been restructured for parallel processing so that the quality and completeness of collected data can be assessed soon after measurement.

DATA REDUCTION

Definition: reducing our observations, a set of integrated intensities, to a set of symmetry-unique intensities (and sigmas).

$$I_1^{(1,1,1)}, \sigma_1^{(1,1,1)}, I_2^{(1,1,1)}, \sigma_2^{(1,1,1)}, \ldots \rightarrow \mathbf{I}^{(1,1,1)}, \boldsymbol{\sigma}^{(1,1,1)}$$

Crystallographic theory: the diffraction pattern symmetry is related to the crystal symmetry. There are set of reflections ('symmetry-equivalents') which must have the same scattering intensity (F^2). e.g. (+-1, 0, 0), (0,+-1, 0), (0, 0, +-1) in cubic crystal

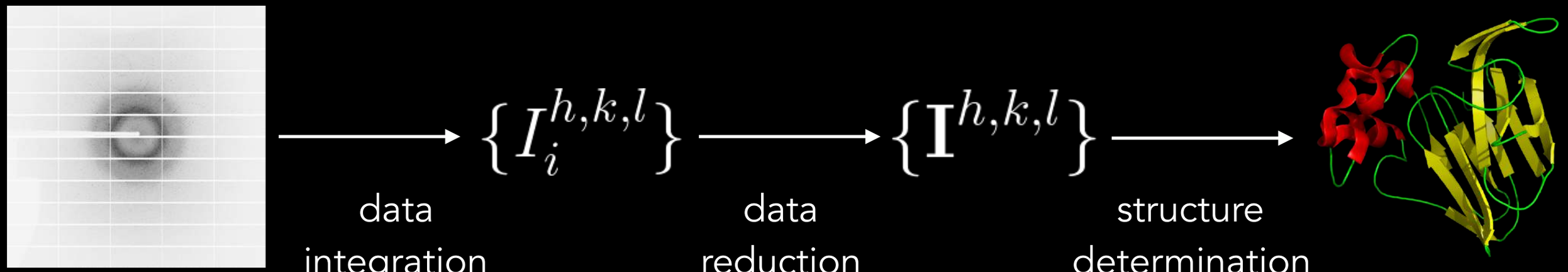
These intensities should then be proportional to F^2

$$F_{hkl} = \sum_{j}^{atoms} f_j e(2\pi i(hx_j + ky_j + lz_j))$$

position of atom in unit cell

which element

DATA REDUCTION



This talk: underlying principles behind data reduction programs.

You don't *need* to understand this to be able to use the programs (highly automated pipelines).

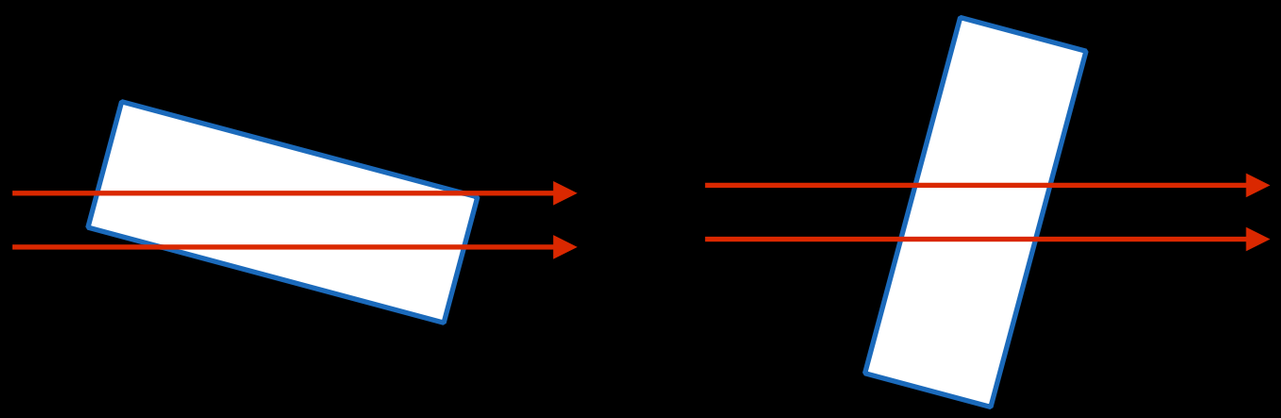
You might like to understand this to help you improve and understand your processed dataset. 🙂

EXPERIMENTAL EFFECTS

Integrated intensities are a combination of F^2, square of structure factor amplitudes, and experimental effects.

$$F_{hkl} = \sum_{j}^{atoms} f_j e(2\pi i(hx_j + ky_j + lz_j))$$

Why do symmetry-equivalent observations have different intensities?

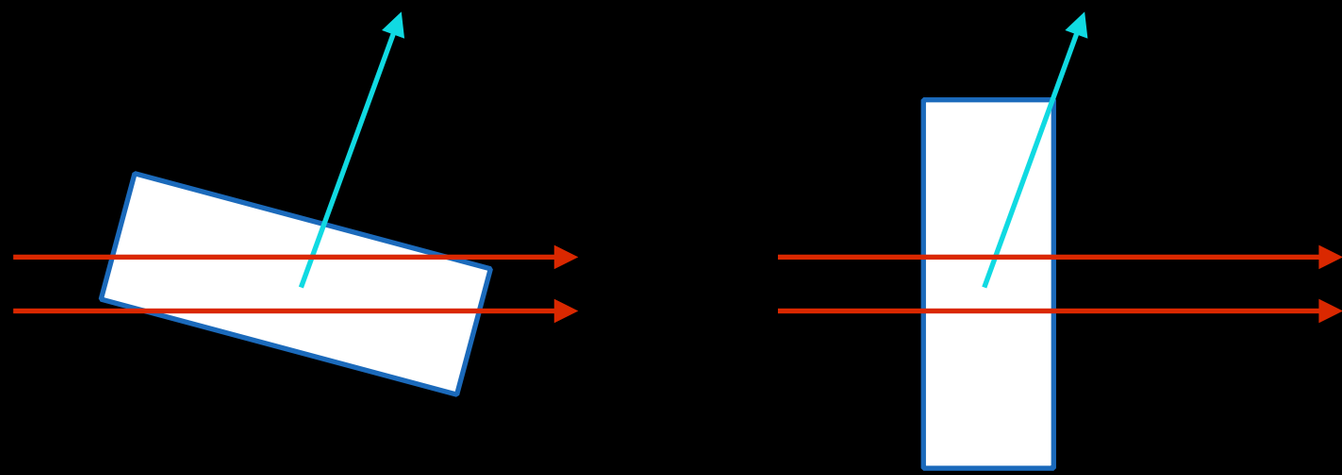


Beam effects - illuminated volume, changes in beam intensity (slow),

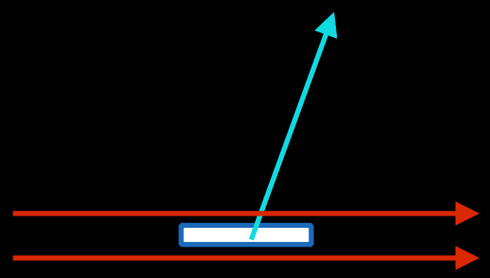
EXPERIMENTAL EFFECTS

Why do symmetry-equivalent observations have different intensities?

Crystal effects - absorption of the incident & scattered beam, average radiation damage.



Detector effects - gain & calibration issues, edge effects (CCDs).



We must also put measurements of multiple crystals on the same scale in order to merge them together!

SCALING AND MERGING

Use the redundancy within the dataset, i.e. repeated measurements of equivalent reflections, to determine a model to account for experimental effects.

Determine scale factors for the reflections, so that symmetry equivalent reflections have the best agreement - least deviation from a weighted average for the group. Then merge (average) the intensities in each group.

Experiment

?

Observations $I_{(001)} = 100$ $I_{(001)} = 50$ $R_{p.i.m.}=0.33$

Model scale factor=1 scale factor=0.5

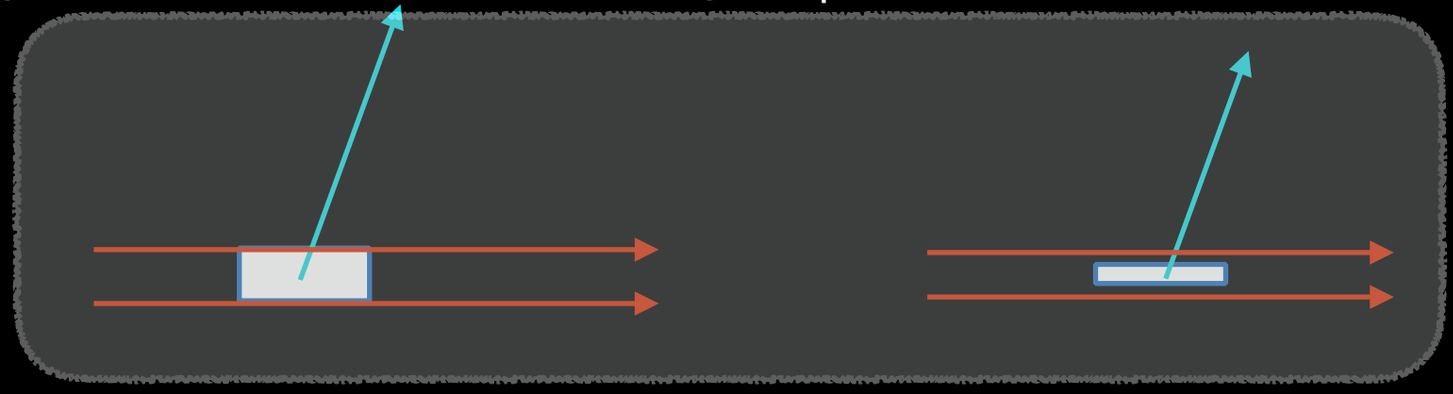
Reasoning (maths): $100/1 = 50/0.5$ $R_{p.i.m.}=0.0$

SCALING AND MERGING

Use the redundancy within the dataset, i.e. repeated measurements of equivalent reflections, to determine a model to account for experimental effects.

Determine scale factors for the reflections, so that symmetry equivalent reflections have the best agreement - least deviation from a weighted average for the group. Then merge (average) the intensities in each group.

Experiment



?

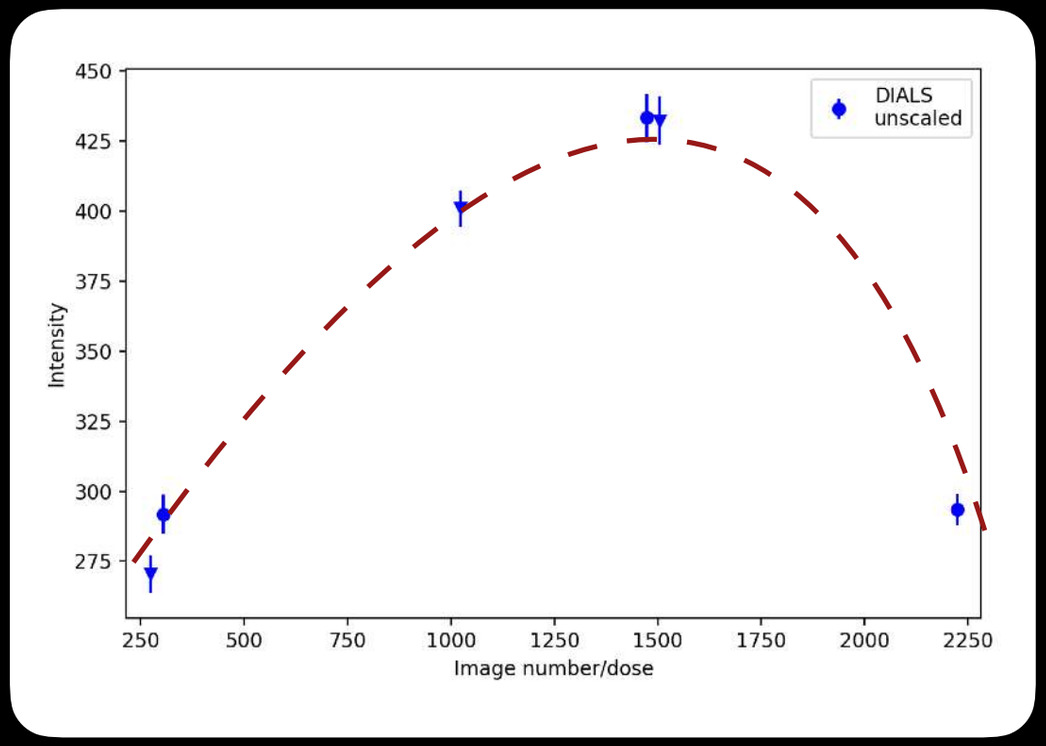
Observations $I_{(001)} = 100$ $I_{(001)} = 50$ $R_{p.i.m.}=0.33$

Model scale factor=1 scale factor=0.5

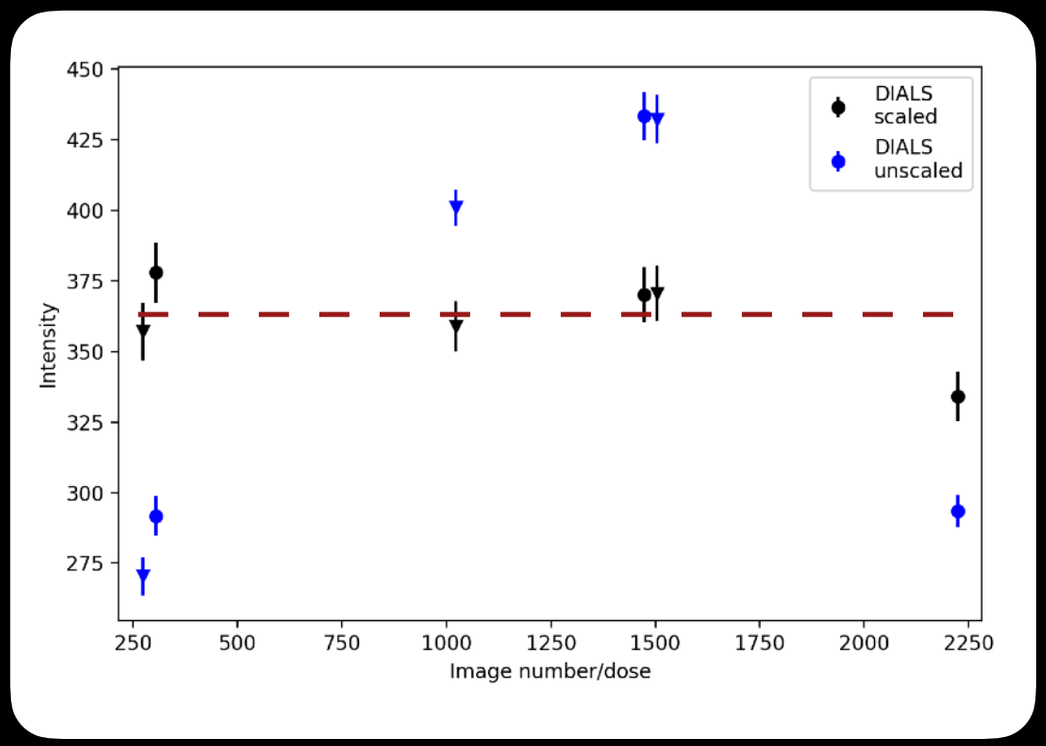
Reasoning (maths): 100/1 = 50/0.5 $R_{p.i.m.}=0.0$

SCALING AND MERGING

More realistic example, within a single sweep

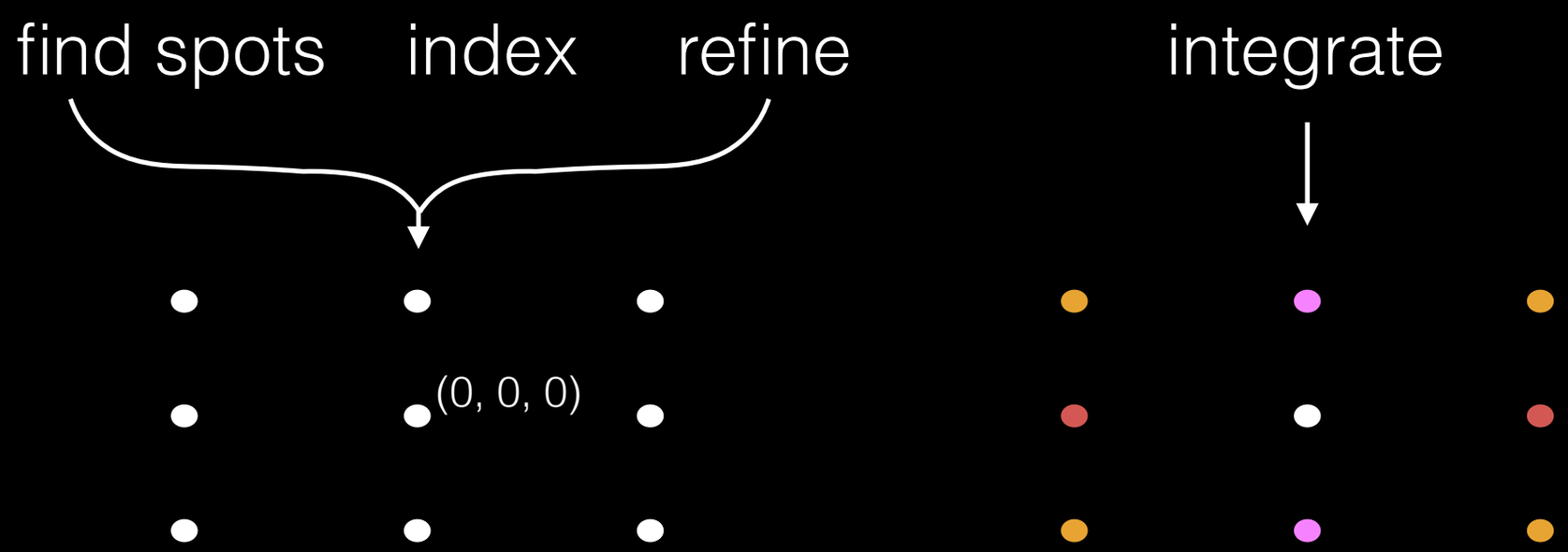


scale



PRECURSOR: SYMMETRY ANALYSIS

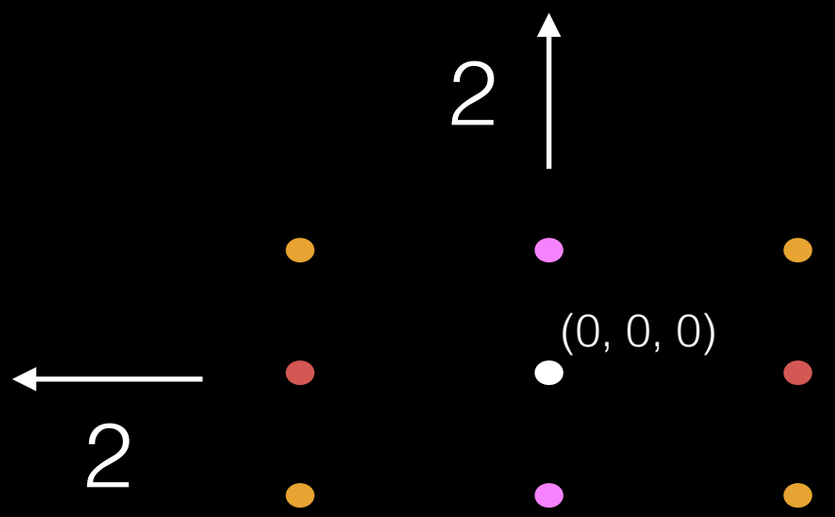
We need to know the symmetry of the diffraction pattern, so that observations are correctly grouped for scaling and merging.



Before integration, only know symmetry of the lattice (spot positions), not the diffraction pattern (intensities). Lattice symmetry != diffraction pattern symmetry

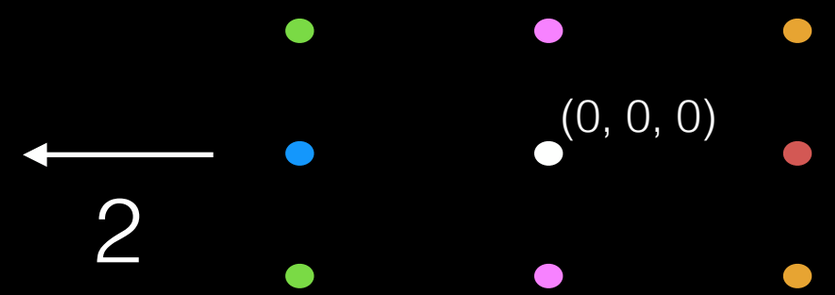
PRECURSOR: SYMMETRY ANALYSIS

Point group symmetry should be evaluated, so that intensities are correctly grouped in scaling.



P222

Can scale in lower symmetry point group, but then not using full equivalence of certain observations!

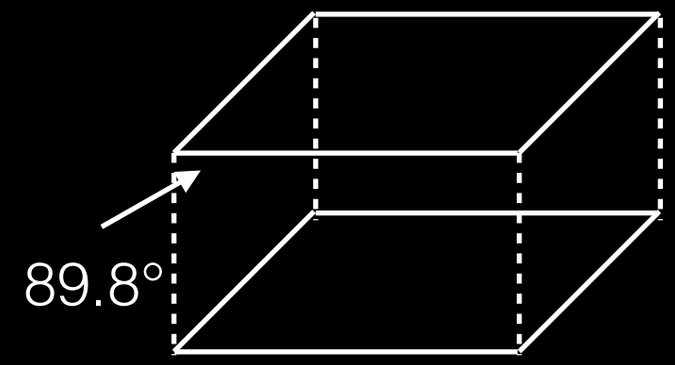


P2

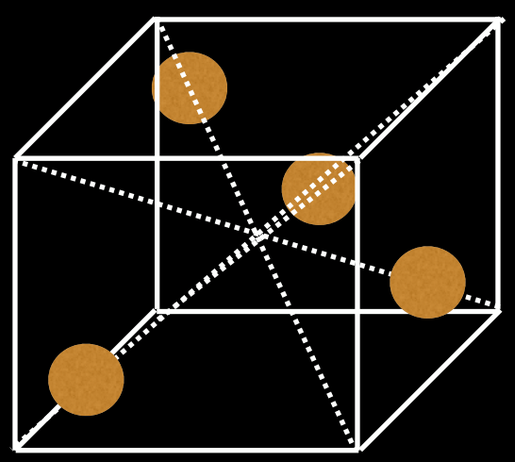
PRECURSOR: SYMMETRY ANALYSIS

Why is lattice symmetry != diffraction pattern symmetry?

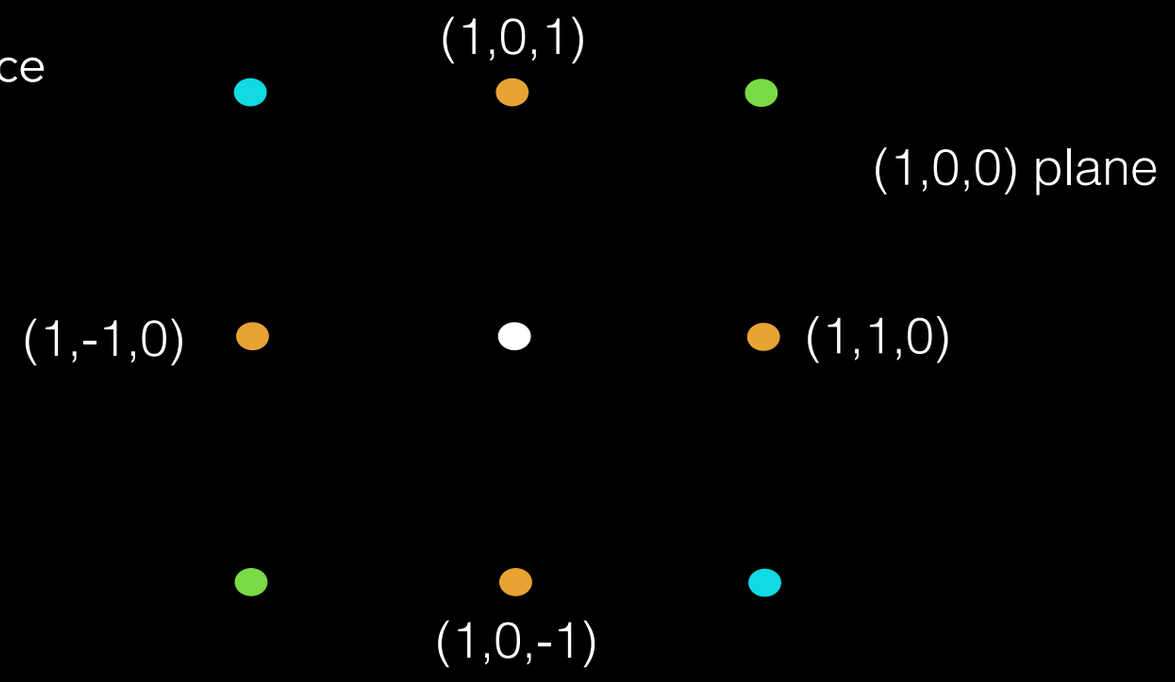
May have `accidental` symmetry of the lattice, e.g. P1 symmetry with angle close to 90° looks like P2.



Also able to have point group symmetry lower than the lattice symmetry.

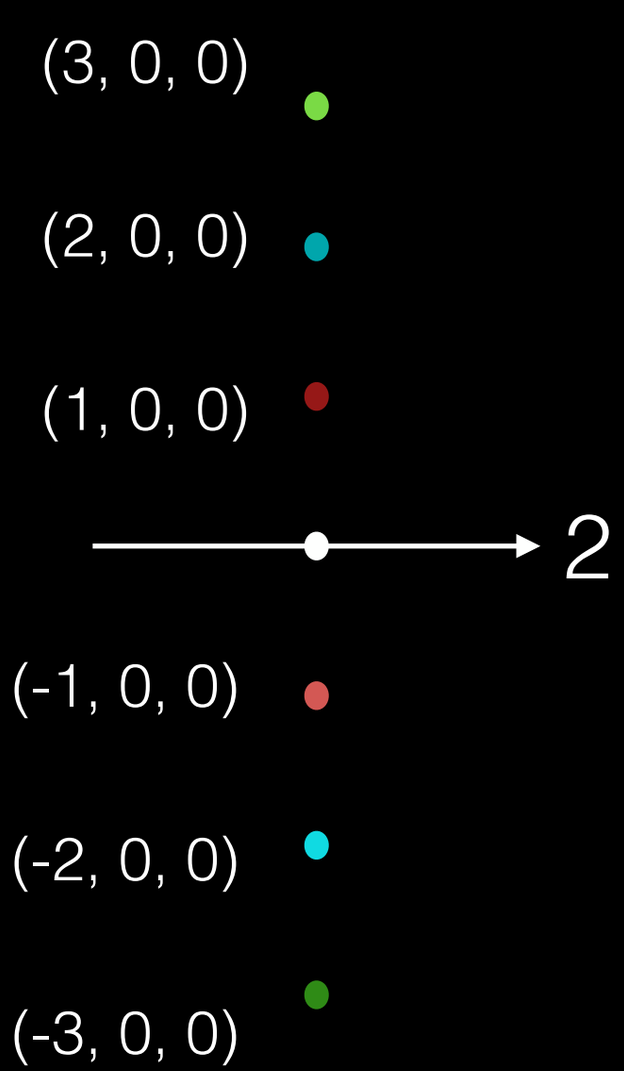


Point group 23, Lattice symmetry 432



Programs for symmetry assessment: *Pointless, dials.symmetry, XDS (CORRECT step)*

PRECURSOR: SYMMETRY ANALYSIS WITH *DIALS.SYMMETRY / POINTLESS / XDS*



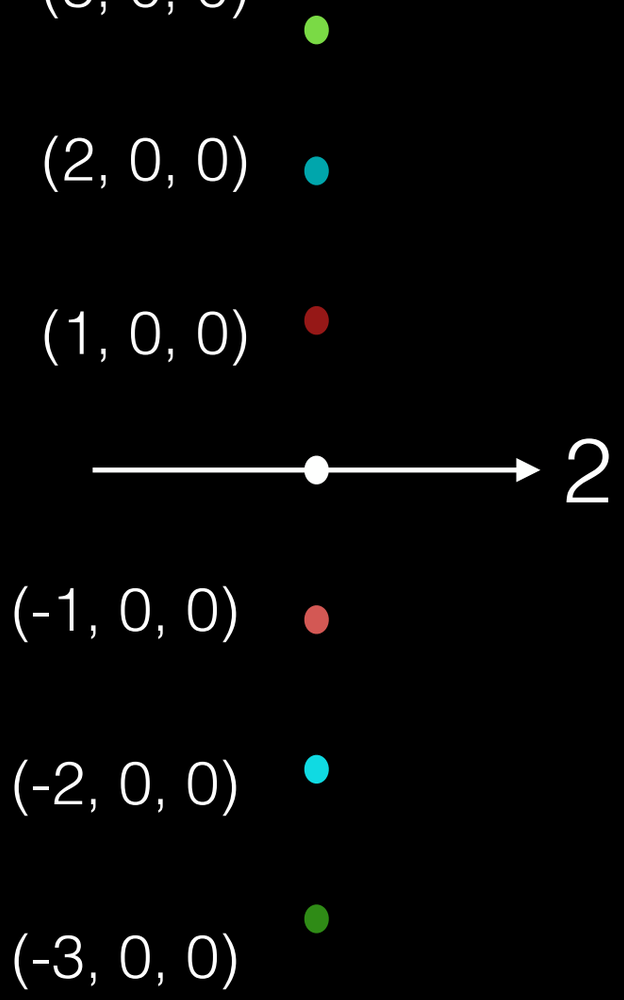
Score symmetry element based on correlation between intensities.

Cubic insulin

Scoring individual symmetry elements

likelihood	Z-CC	CC	N		Operator
0.943	9.90	0.99	905647	***	1 |(0, 0, 0)
0.196	5.02	0.50	1694845		4 |(1, 1, 0)
0.196	5.02	0.50	1693679		4 |(1, 0, 1)
0.196	5.03	0.50	1694455		4 |(0, 1, 1)
0.931	9.67	0.97	1713055	***	3 |(1, 0, 0)
0.939	9.80	0.98	1715415	***	3 |(0, 1, 0)
0.939	9.81	0.98	1715530	***	3 |(0, 0, 1)
0.932	9.68	0.97	1712444	***	3 |(1, 1, 1)
0.933	9.70	0.97	856921	***	2 |(1, 1, 0)
0.197	5.04	0.50	847218		2 |(-1, 1, 0)
0.938	9.78	0.98	857342	***	2 |(1, 0, 1)
0.197	5.05	0.50	851699		2 |(-1, 0, 1)
0.926	9.59	0.96	876485	***	2 |(0, 1, 1)
0.195	5.02	0.50	862772		2 |(0, -1, 1)
0.197	5.05	0.50	847298		2 |(1, 1, 2)
0.197	5.05	0.50	847688		2 |(1, 2, 1)
0.198	5.06	0.51	855659		2 |(2, 1, 1)

PRECURSOR: SYMMETRY ANALYSIS WITH *DIALS.SYMMETRY / POINTLESS / XDS*



Score different Patterson groups based on combination of symmetry elements of the lattice.

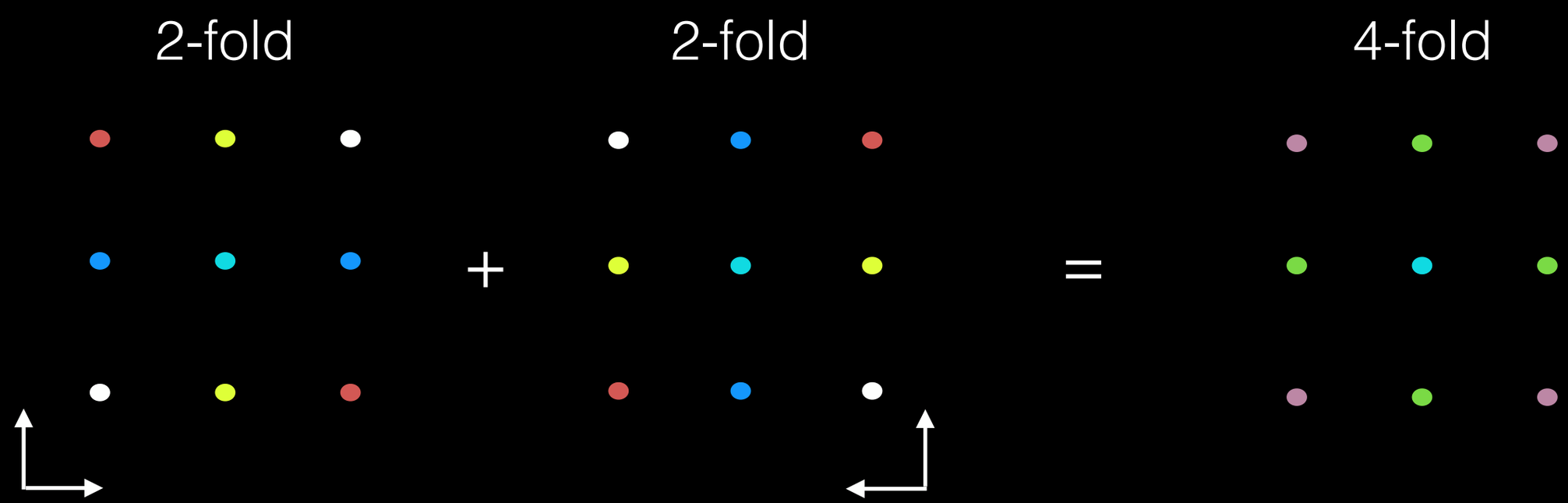
Scoring all possible sub-groups

Patterson group		Likelihood	NetZcc	Zcc+	Zcc-	CC	CC-	delta	Reindex operator
I m -3	***	1.000	4.70	9.74	5.04	0.97	0.50	0.0	-a,b,-c
I m m m		0.000	2.91	9.74	6.84	0.97	0.70	0.0	-a,b,-c
I m -3 m		0.000	7.62	7.62	0.00	0.74	0.00	0.0	-a,b,-c
I 4/m m m		0.000	0.78	8.07	7.29	0.74	0.74	0.0	-c,-a,b
I 4/m m m		0.000	0.78	8.07	7.29	0.74	0.74	0.0	-a,b,-c
I 4/m m m		0.000	0.78	8.07	7.29	0.74	0.74	0.0	b,-c,-a
R -3 :H		0.000	2.58	9.85	7.27	0.98	0.71	0.0	b-c,-a-b,-1/2*a+1
R -3 :H		0.000	2.57	9.85	7.27	0.98	0.71	0.0	-a+b,a-c,-1/2*a-1
I 1 2/m 1		0.000	2.56	9.84	7.28	0.98	0.72	0.0	a,-b,-c
I 1 2/m 1		0.000	2.51	9.80	7.28	0.98	0.72	0.0	-a,-c,-b
R -3 :H		0.000	2.50	9.79	7.28	0.98	0.71	0.0	-b-c,-a+c,-1/2*a+
R -3 :H		0.000	2.50	9.78	7.29	0.97	0.71	0.0	-a-c,b+c,1/2*a+1/
I 1 2/m 1		0.000	2.45	9.75	7.29	0.98	0.72	0.0	b,-a,c
I 4/m		0.000	1.13	8.54	7.41	0.75	0.74	0.0	-c,-a,b
I 4/m		0.000	1.09	8.51	7.42	0.75	0.74	0.0	-a,b,-c
I 4/m		0.000	1.04	8.47	7.43	0.74	0.74	0.0	b,-c,-a
P -1		0.000	2.44	9.90	7.46	0.99	0.73	0.0	x+y,x-z,y-z
F m m m		0.000	0.26	7.82	7.56	0.75	0.74	0.0	-b,-a+c,-a-c
F m m m		0.000	0.22	7.79	7.57	0.75	0.74	0.0	-c,-a-b,-a+b
F m m m		0.000	0.18	7.76	7.58	0.74	0.74	0.0	-a,b+c,b-c

Additionally, assess systematic absences to determine full space group symmetry (not needed for scaling).

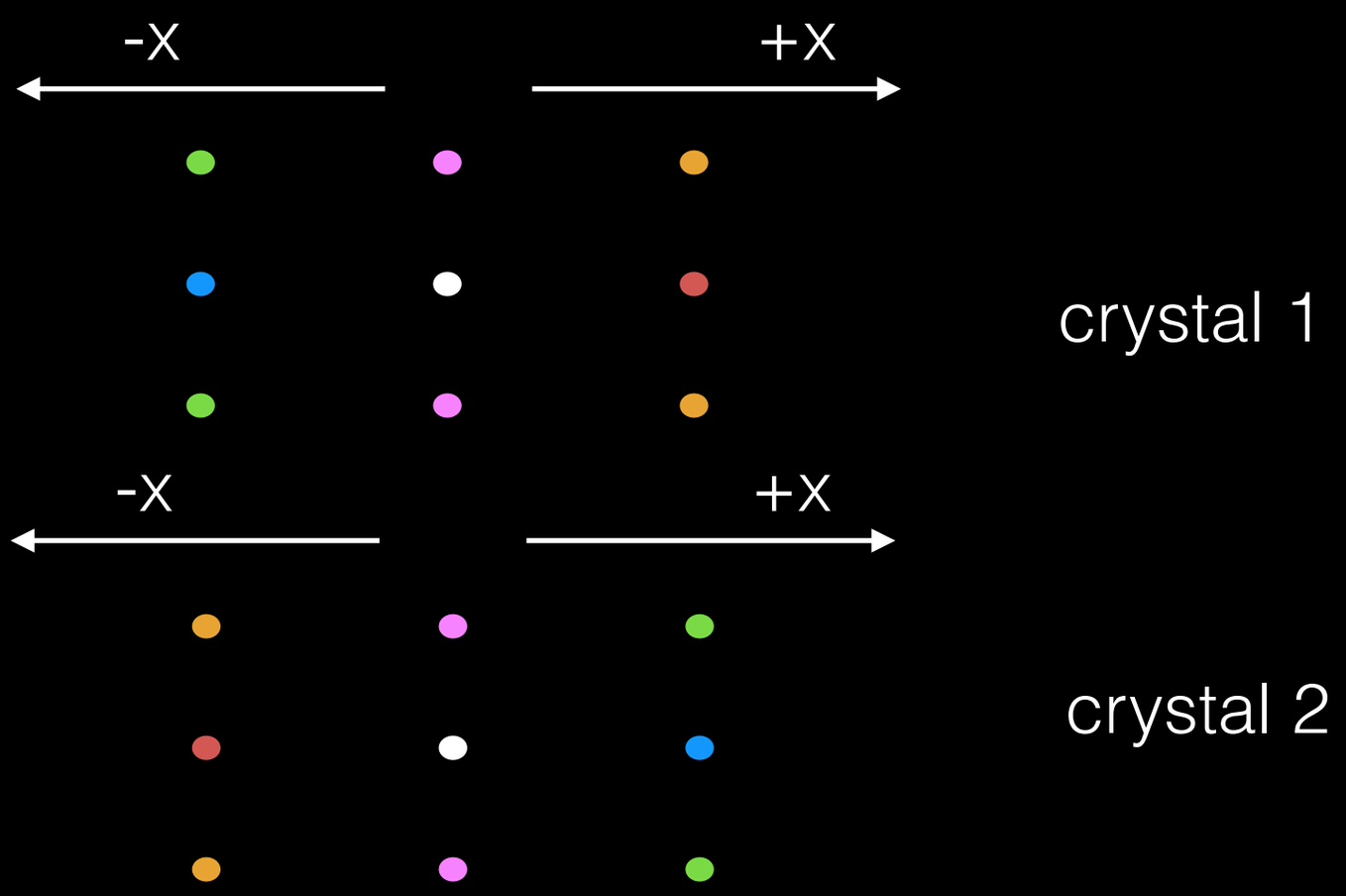
SYMMETRY ANALYSIS DIFFICULTIES

Effects of twinning & pseudo-symmetry, can give appearance of higher symmetry. Can not be certain that you know the symmetry until you have solved the structure!



SYMMETRY ANALYSIS DIFFICULTIES

Indexing ambiguities in multi-crystal experiments; for certain point groups, there are different ways to index.



Need consistent indexing in order to scale. A particular issue for sparse multi-crystal datasets - can use *dials.cosym* [1] to solve symmetry and reindex. Alternatives: xscale_isocluster, or use pointless/dials.reindex/XDS to reindex against reference.

[1] Determination of Patterson group symmetry from sparse multi-crystal data sets in the presence of an indexing ambiguity. Gildea, R. J. & Winter, G. (2018)

RECAP: DATA REDUCTION STEP 1

Assess point group (space group) symmetry*.
Beware of difficulties.

Integrated reflections

Integrated reflections, in correct point group

*Unless have prior knowledge for this system, or want to process in *P*1.

SCALING

Determine scale factors for the reflections, so that symmetry equivalent reflections have the best agreement. Makes the data *internally consistent.*

Refine a scaling model to give most internally consistent data.

Once data is scaled, can assess the internal consistency with a number of measures [1] - note that this assesses the *precision*, not the *accuracy,* need to be aware of possibility for systematic errors. See the next talk on assessing data quality for the details!

Programs can use: dials.scale, XDS/XSCALE, Aimless, scalepack, …..

[1] Assessing and maximising data quality in macromolecular crystallography, Karplus & Diederichs, 2015.

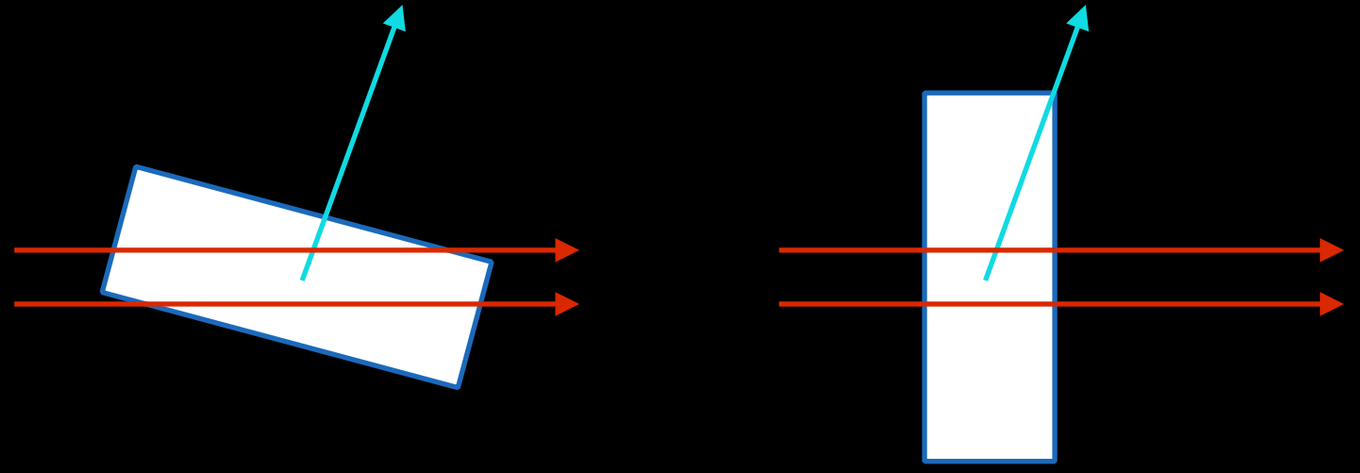
SCALING PROCESS COMPONENTS

- rounds of outlier rejection (right reflections)
- rounds of scaling model refinement (best scales)
- model the error distribution (sigmas) (better errors)
- combine profile and summation intensity estimates (best integrated intensities)

SCALING MODELS

Principle - experimental effects multiplicative in their effect on intensities - scale factor given by product of corrections.

$$g_{hl} = S_{hl}(p_1) \times T_{hl}(p_2) \times \ldots$$



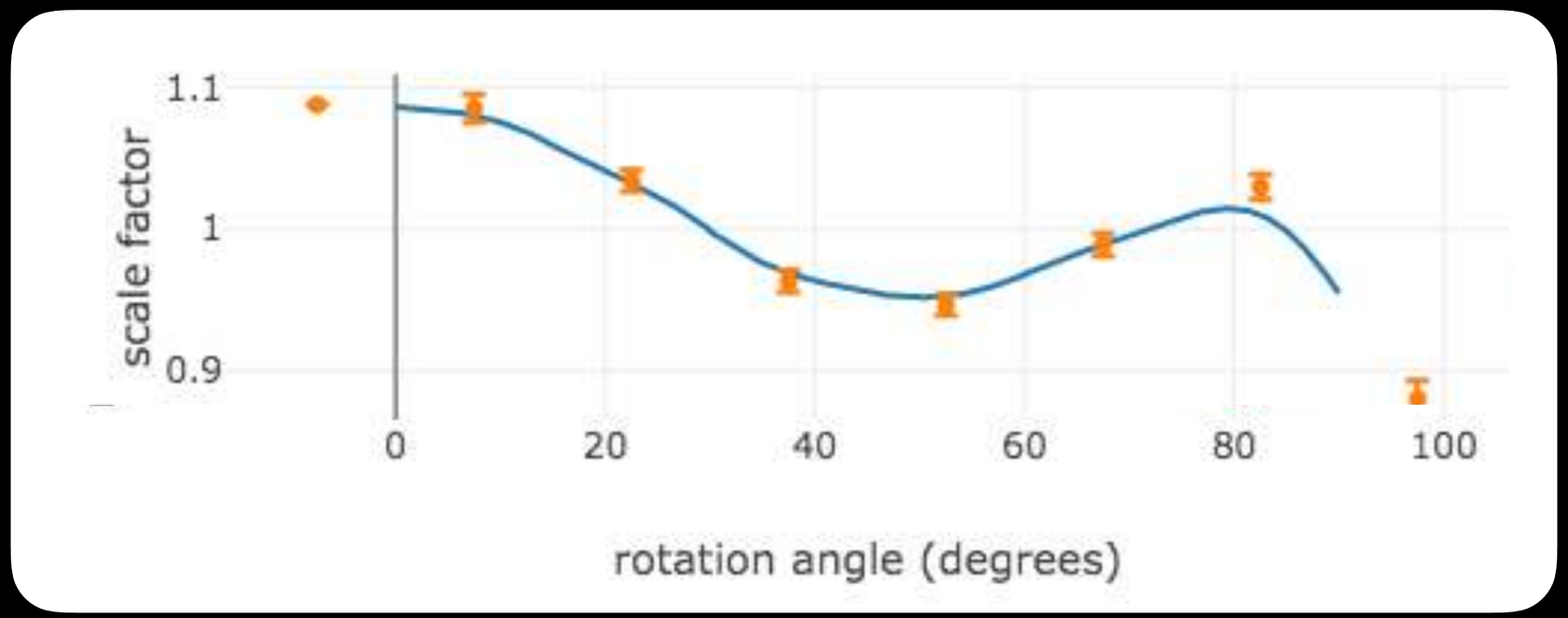
APPROACH 1: PHYSICAL SCALING MODEL

Three part model *per sweep/crystal*, approach used in Aimless & DIALS

Correction 1: Smoothly varying scale factor, dependent on crystal rotation angle.

Corrects for changes in illuminated volume, slow changes in beam intensity, primary absorption.

Gives the relative scale for multi-sweep/multi-crystal datasets.



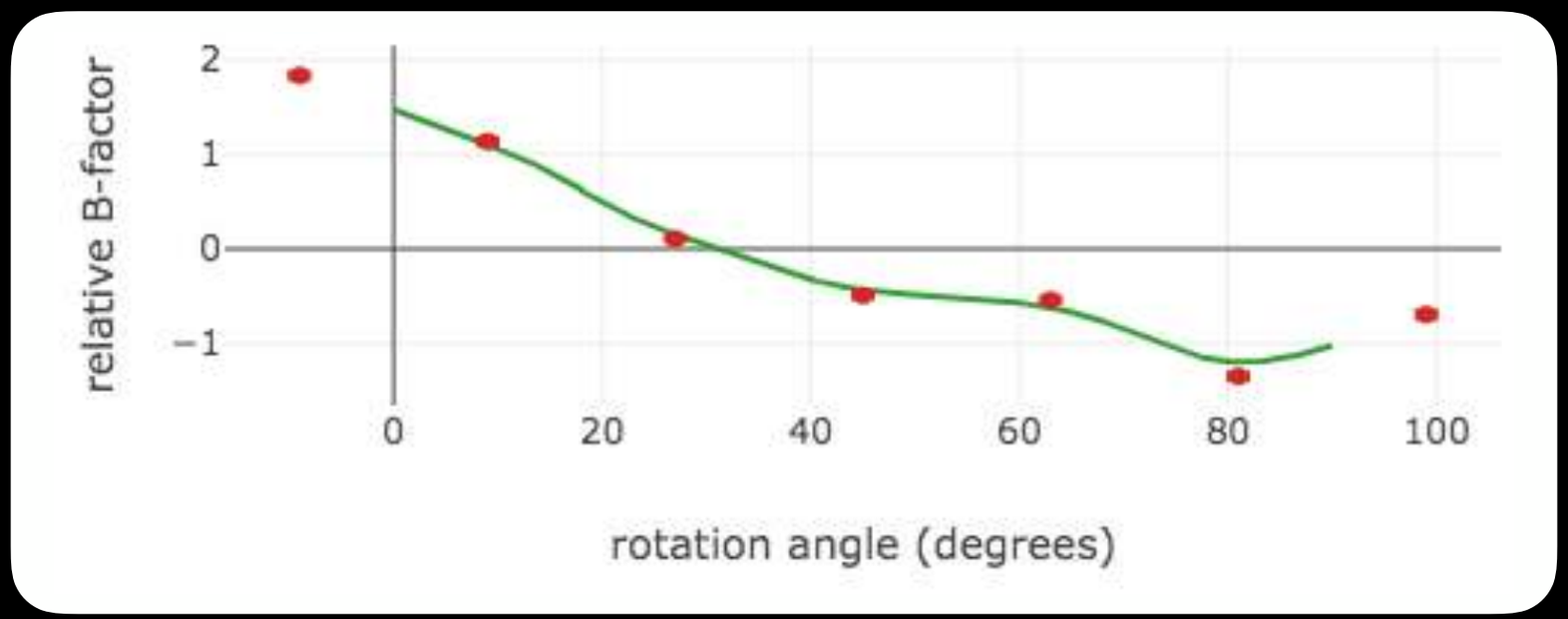
APPROACH 1: PHYSICAL SCALING MODEL

Three part model, approach used in Aimless & DIALS

Correction 2: Smoothly varying isotropic B-factor, dependent on crystal rotation angle (i.e. accumulated exposure time).

$$e(B(t)/2d^2)$$

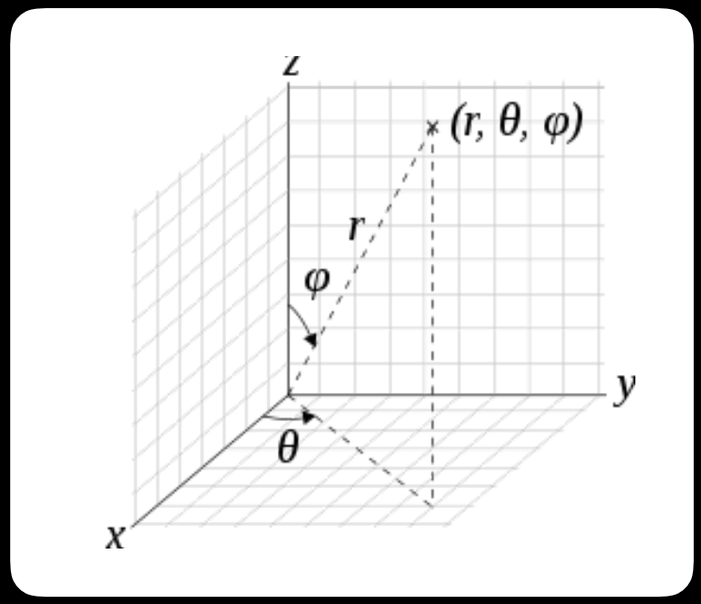
Corrects for global/average radiation damage (*not* site specific).



APPROACH 1: PHYSICAL SCALING MODEL

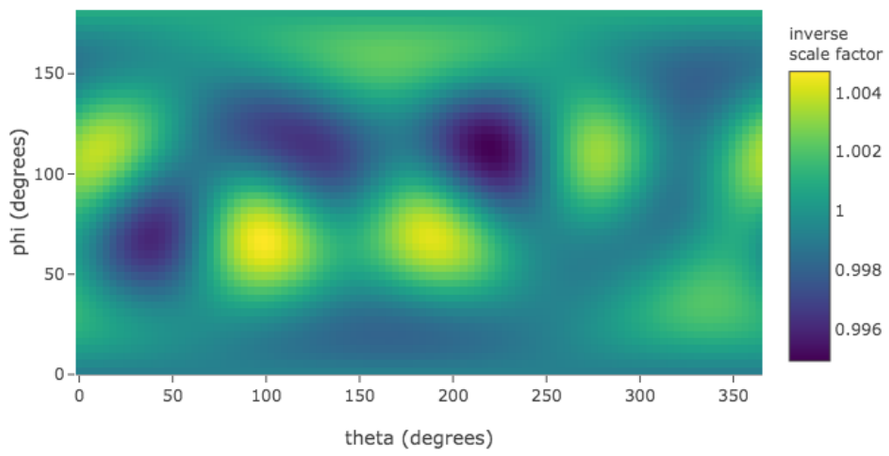
Three part model, approach used in Aimless & DIALS

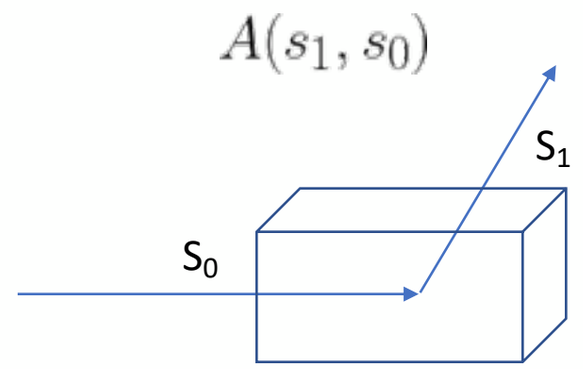
Correction 3: relative absorption correction for different directions in space, based on scattering vectors.



Define a correction dependent on direction relative to crystal frame. Use spherical harmonics as a basis to define a smoothly varying surface.

Typically small for MX (~1%), unless at long wavelength/absorption edge.





APPROACH 2: ARRAY SCALING MODEL

More generalised approach - just define what the corrections depend on. e.g d-value, rotation angle, position on the detector. XDS/XSCALE approach.

Components are arrays/grids of parameters e.g. 2D grid of parameters equally spaced in resolution and rotation angle.

Differences to physical scaling model: Less restrained by assumptions on form of corrections, more general/adaptable. Physical interpretation of the corrections may be less clear. Array model implementation typically requires many more parameters than physical model (100s vs ~60).

APPROACH 2: ARRAY SCALING MODEL

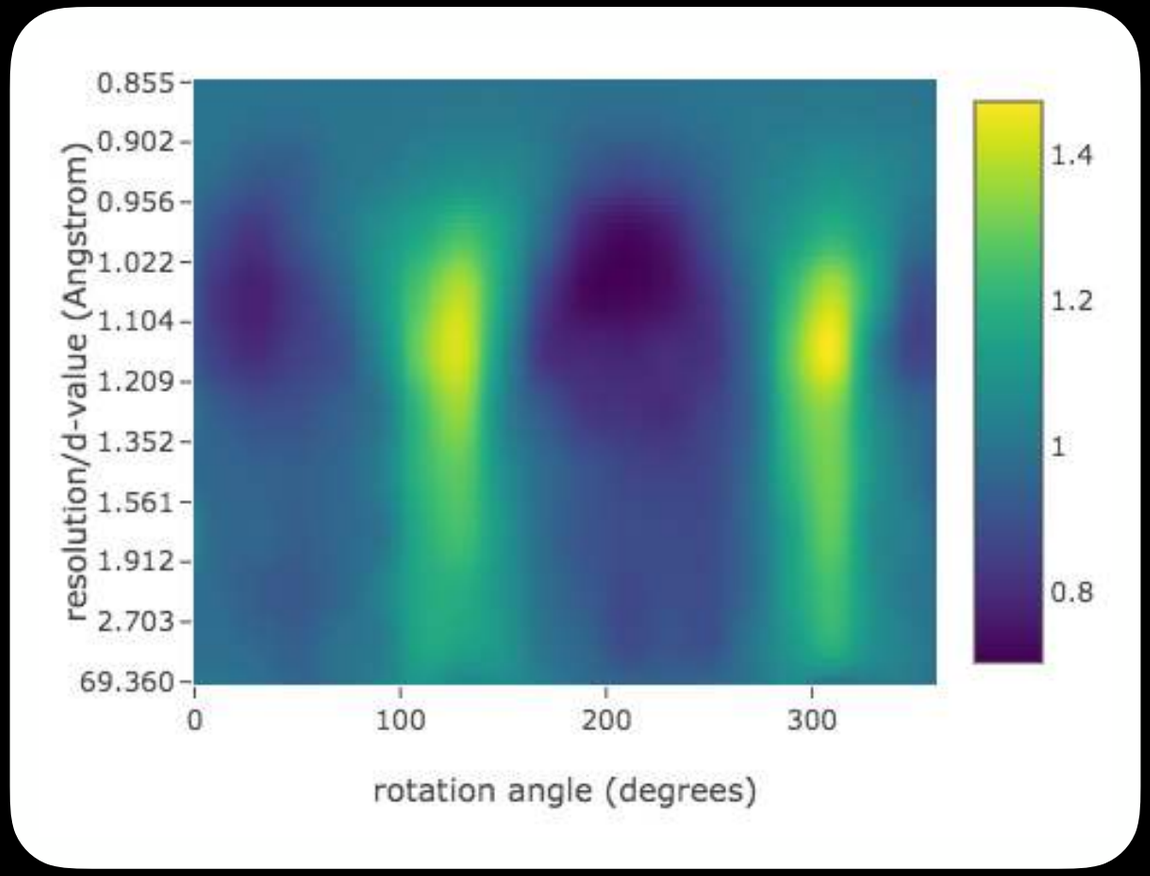
Have an implementation of this approach as an option within dials.scale.
Same principles as behind the model from XDS/Xscale.

model=array (default model=physical)

Smoothly weight a grid of parameters to get correction for a reflection.

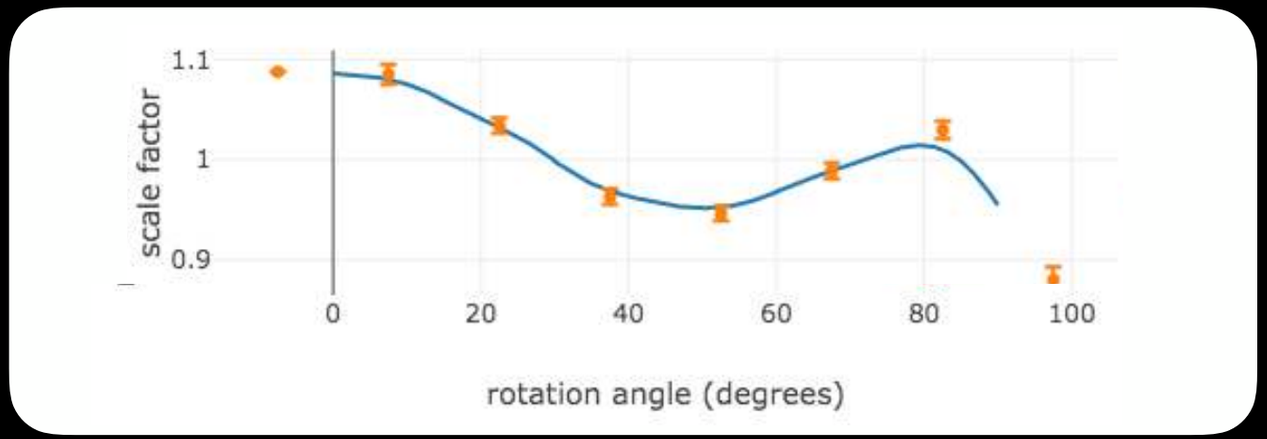
Decay correction - 2D grid of parameters; d-values & rotation angle.

Absorption correction - 3D grid of parameters; reflection x, y coordinate on detector, rotation angle.

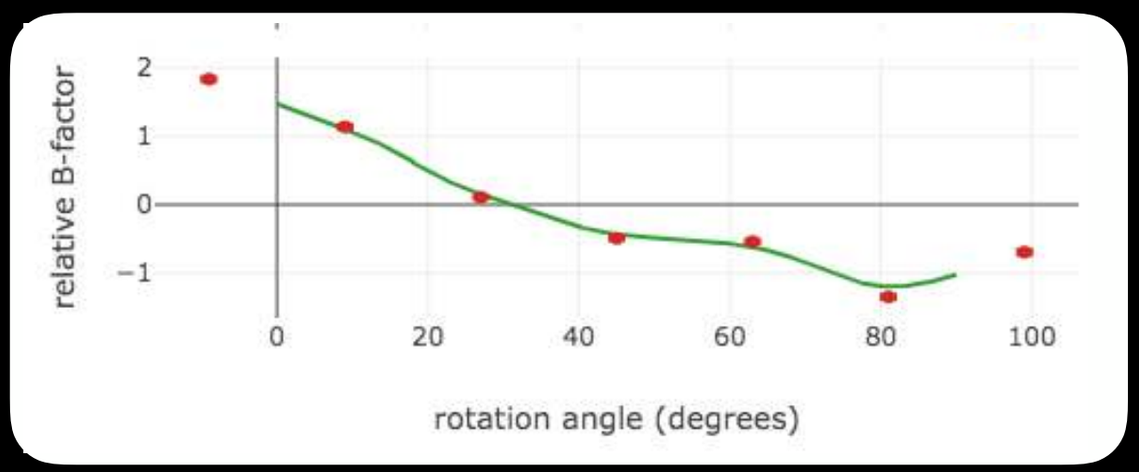


RECAP: DIALS SCALING MODEL

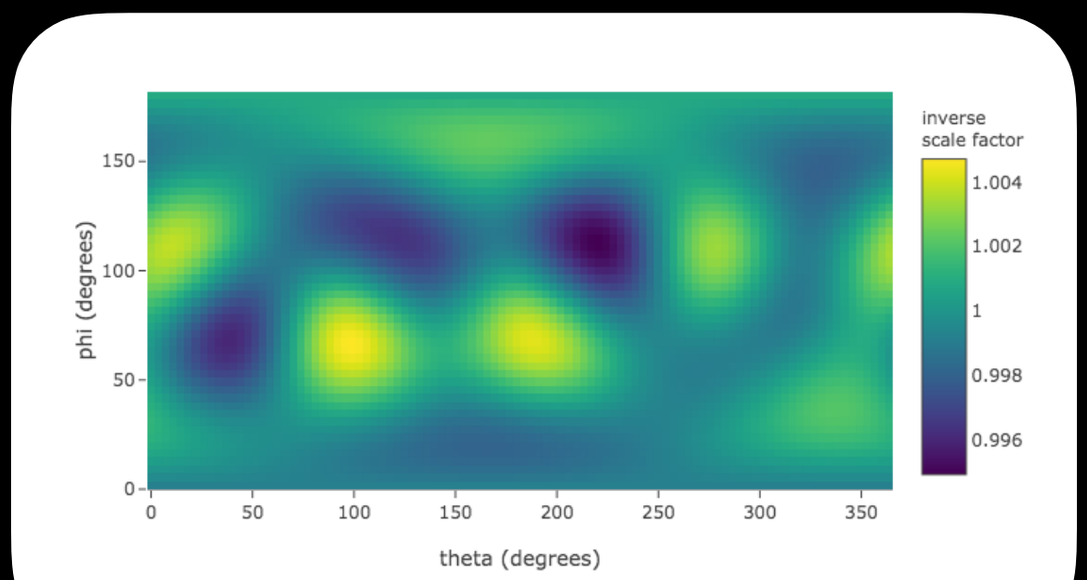
Scaling correction =



x



x

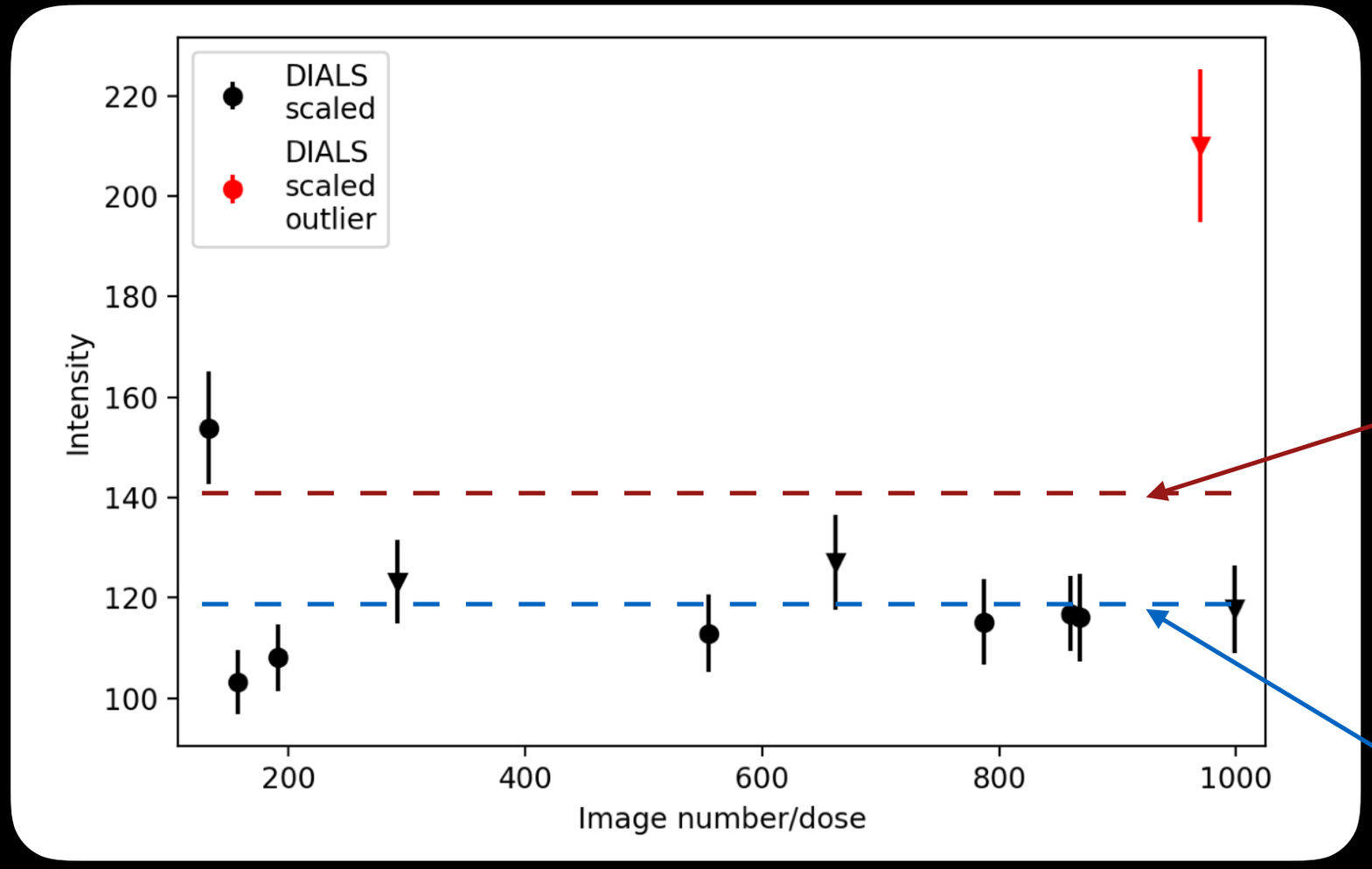


OUTLIER REJECTION

Outliers have a disproportionate effect - largest RMSD.

Test the normalised deviation from the symmetry group, reject those above a threshold. Several rounds of outlier rejection needed - at each point retest *all* reflections.

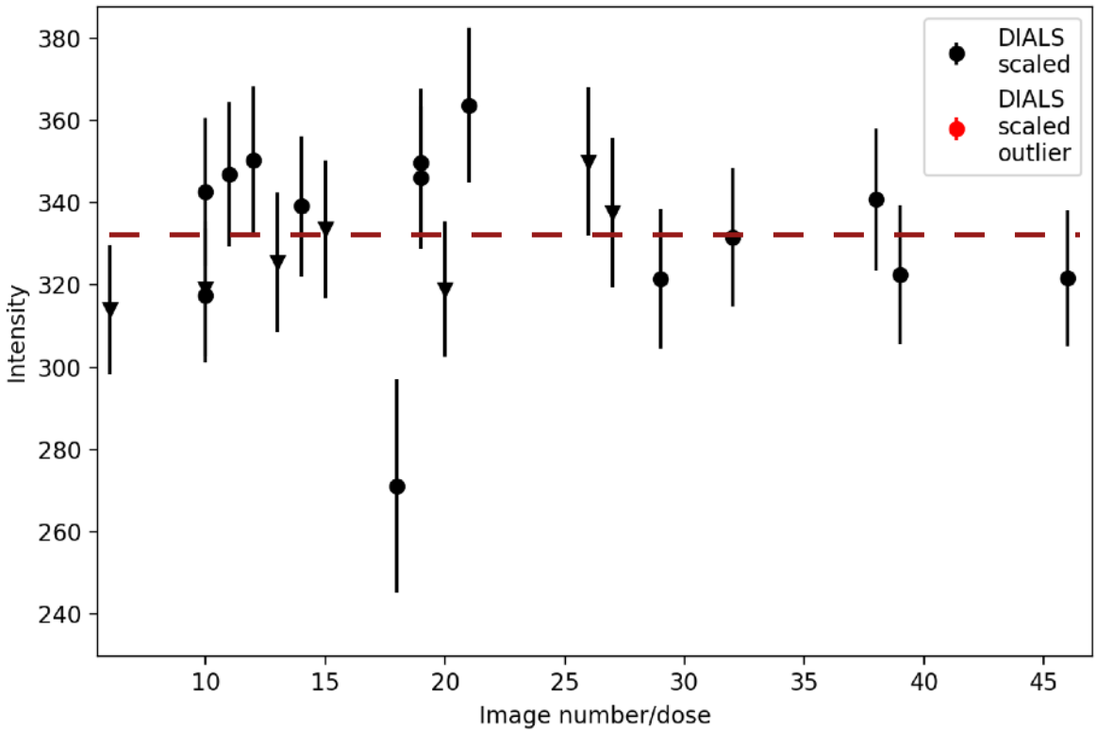
$$\delta = (I - g\langle I_h \rangle)/\sigma'$$

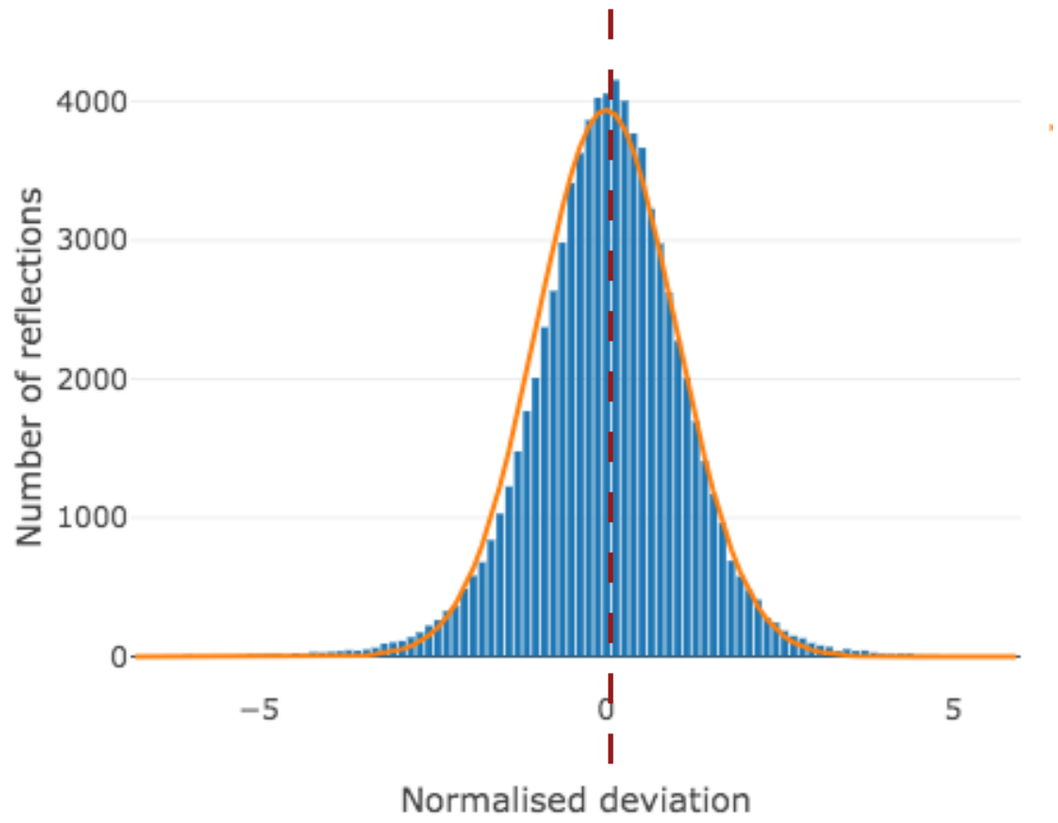


ERROR (SIGMA) DISTRIBUTION

At scaling step, able to investigate the distribution of intensities and errors (uncertainties) throughout the dataset. Chance to evaluate systematic errors.

Principle: for realistic errors, the overall distribution of intensities is consistent with the errors on individual reflections.

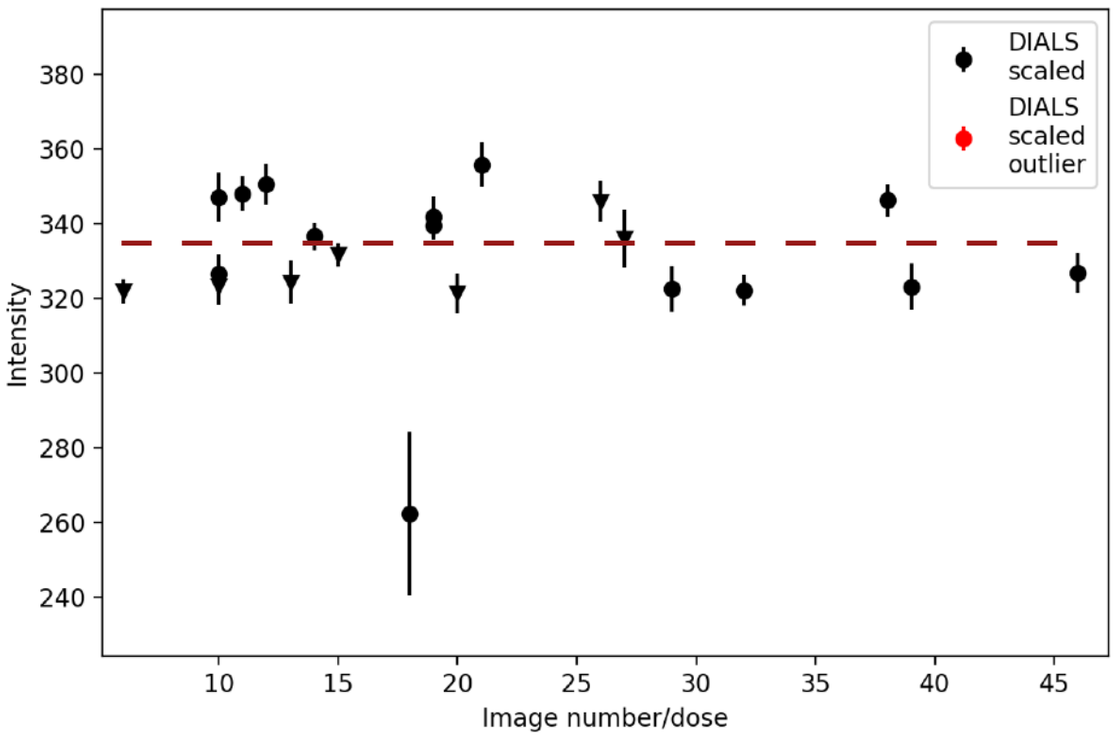




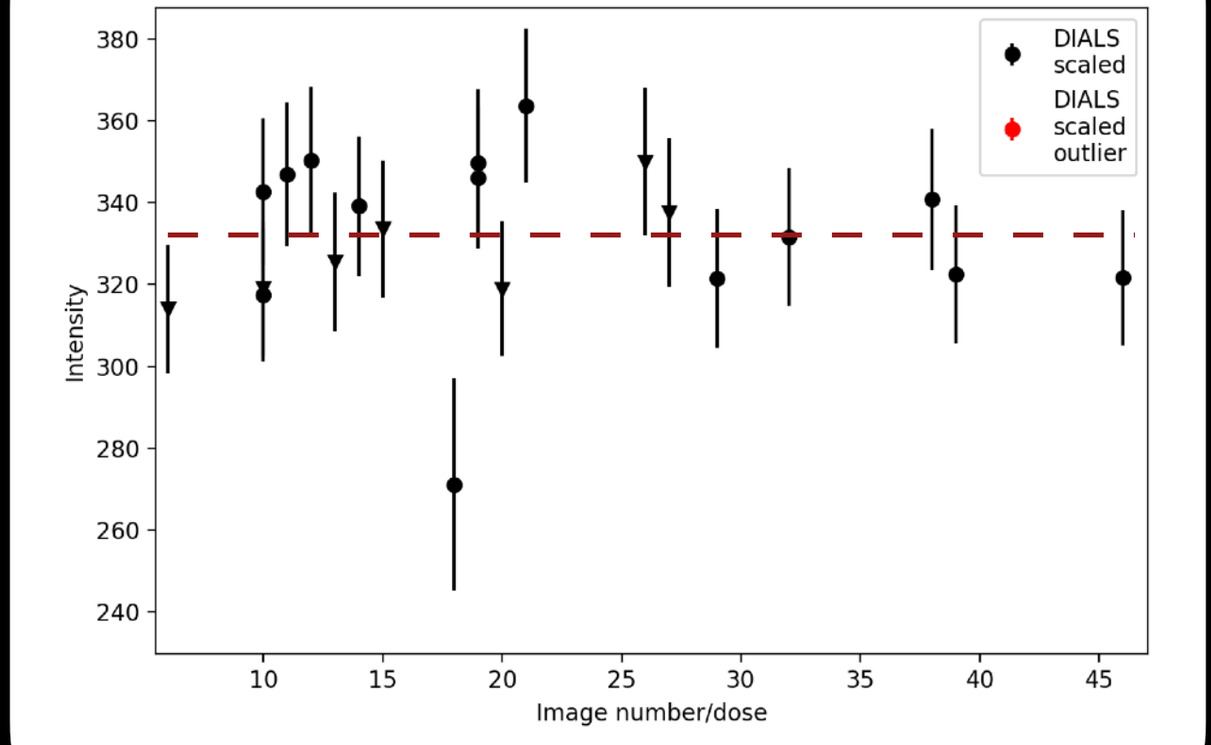
ERROR (SIGMA) DISTRIBUTION

In reality, the intensities deviate from the average by more than expected based upon the errors after integration.

The true error has been *underestimated*, and the effect is most pronounced for the most intense reflections.



Scaled, no error adjustment



Scaled, errors adjusted

Systematic experimental errors may not have been fully accounted for, alongside deficiencies in experimental models during data processing (including scaling).

ERROR (SIGMA) ADJUSTMENT

General approach is to use two parameters to describe the effect of systematic errors and optimise these by assessing the distribution of intensities and uncertainties.

$$\sigma'^2 = a^2(\sigma^2 + (bI)^2)$$

First parameter *a* - scales all sigmas by a constant factor - represents the effect of systematic errors that affect all reflections.

Second term *b* - intensity dependent factor - represents the effect of systematic errors that scale with intensity.

$$I/\sigma_{asymptotic} = 1/(a \times b)$$

a x b is measure of fraction of systematic error (2% very good)

ERROR (SIGMA) ADJUSTMENT

Similarities in form across different data processing programs

dials.scale $$\sigma'^2 = a^2(\sigma^2 + (bI)^2)$$

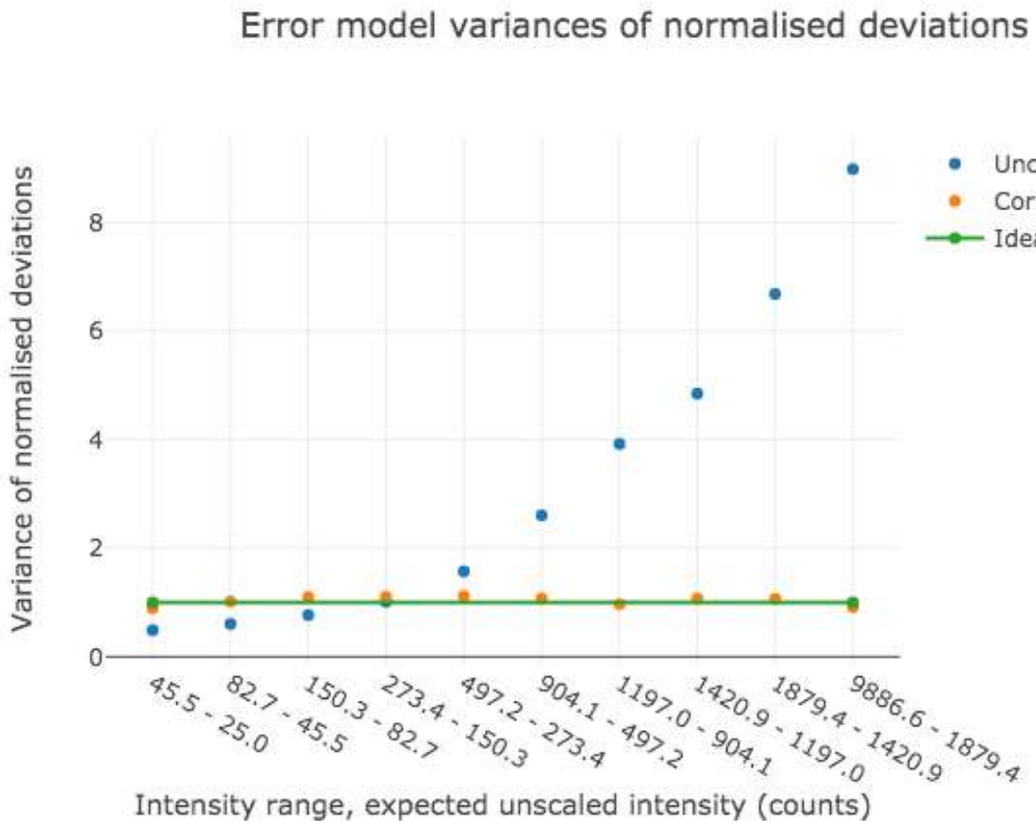
Aimless $$\sigma(I)_{\text{corr}} = SdFac * [\sigma(I)^2 + (SdB * I) + (SdAdd * I)^2]^{\frac{1}{2}}$$

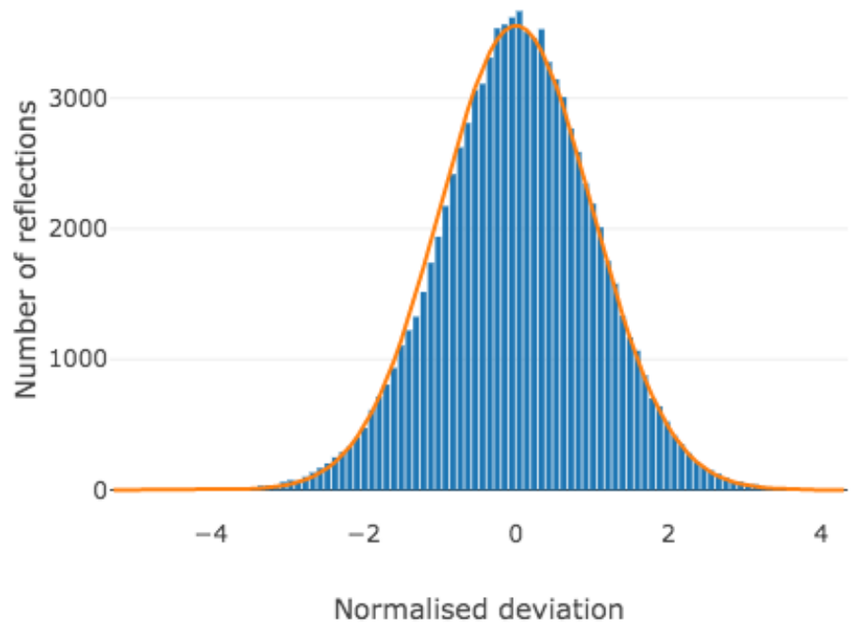
XDS $$\sigma(I)^2 = a * (\sigma(I)_0^2 + b * I^2)$$

Approaches to parameter fitting differ between programs, see papers for technical details.

DIALS ERROR (SIGMA) ADJUSTMENT

$$\sigma'^2 = a^2(\sigma^2 + (bI)^2)$$





dials.scale.html, ‘Error distribution plots’

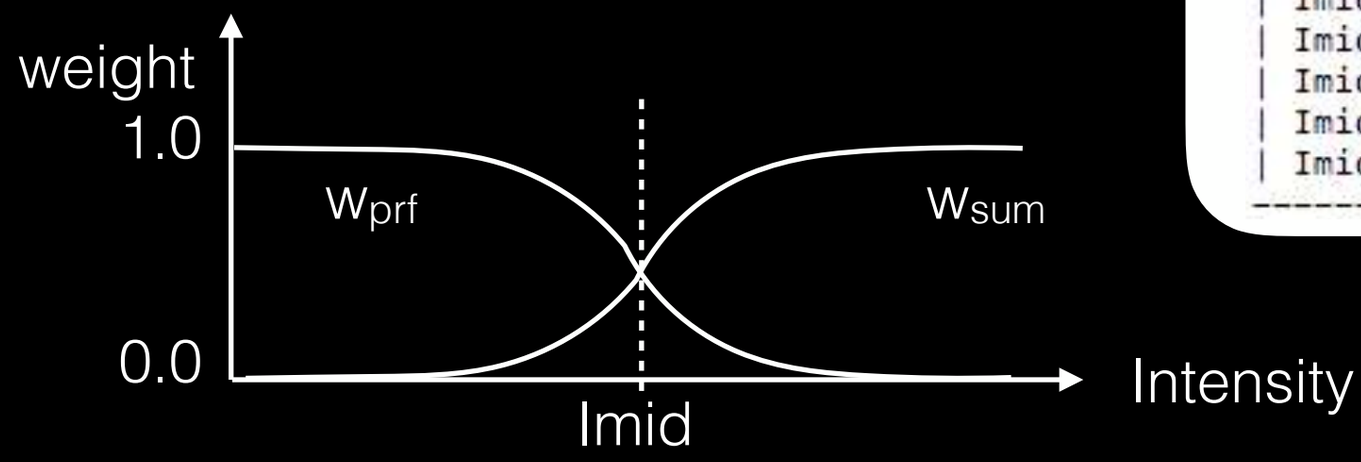
COMBINING SUMMATION/PROFILE INTENSITIES.

Integration programs give summation and profile fitted intensities. Profile intensity estimates are usually best, but summation estimates can be more accurate for the most intense observations.

Try to find best combination - smooth cross over between profile and summation, optimise crossover point. Score based on R_{meas} in dials.scale.

$$I_{scale} = w_{prf} I_{prf} + w_{sum} I_{sum}$$

$$w_{sum} = 1 - w_{prf}$$



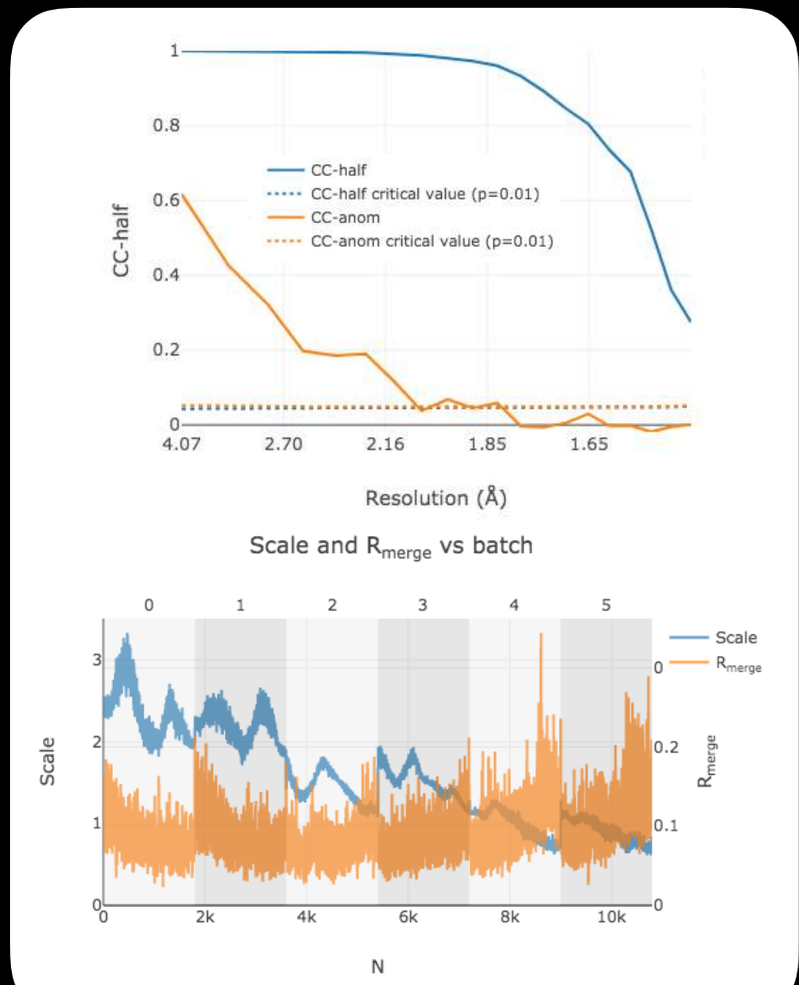

```
Performing profile/summation intensity optimisation.
-------------------------------------
| Combination    | CC1/2   | Rmeas   |
-------------------------------------
| prf only       | 0.99959 | 0.21789 |
| sum only       | 0.99933 | 0.33332 |
| Imid = 20.02   | 0.99956 | 0.22257 |
| Imid = 2003.81 | 0.99959 | 0.21789 |
| Imid = 200.38  | 0.9996  | 0.21773 |
| Imid = 20.04   | 0.99956 | 0.22256 |
| Imid = 2.0     | 0.99946 | 0.25967 |
-------------------------------------
```

ASSESSING DATA QUALITY

After scaling, can assess the overall data quality. Data reduction programs report key statistics. e.g. dials.scale/xia2, there is a html report with plots, text output too.

-------------Summary of merging statistics--------------

	Overall	Low	High
High resolution limit	1.80	4.88	1.80
Low resolution limit	58.76	58.79	1.83
Completeness	99.6	100.0	92.7
Multiplicity	17.6	29.4	6.6
I/sigma	26.1	112.2	1.3
Rmerge(I)	0.086	0.053	0.703
Rmerge(I+/-)	0.082	0.049	0.666
Rmeas(I)	0.088	0.054	0.766
Rmeas(I+/-)	0.086	0.051	0.784
Rpim(I)	0.018	0.010	0.291
Rpim(I+/-)	0.025	0.013	0.401
CC half	0.999	0.999	0.520
Anomalous completeness	98.3	100.0	85.9
Anomalous multiplicity	8.8	14.7	3.4
Anomalous correlation	0.415	0.445	-0.045
Anomalous slope	0.757		
dF/F	0.050		
dI/s(dI)	1.047		
Total observations	445648	37471	8003
Total unique	25368	1276	1207



If issues are identified, try rescaling the data, excluding bad images/datasets.

RECAP: DATA REDUCTION STEP 2

Scaling consists of a few key components:
Determining a scaling model to best correct the intensities,
rejecting outlier observations and improving the sigma estimates

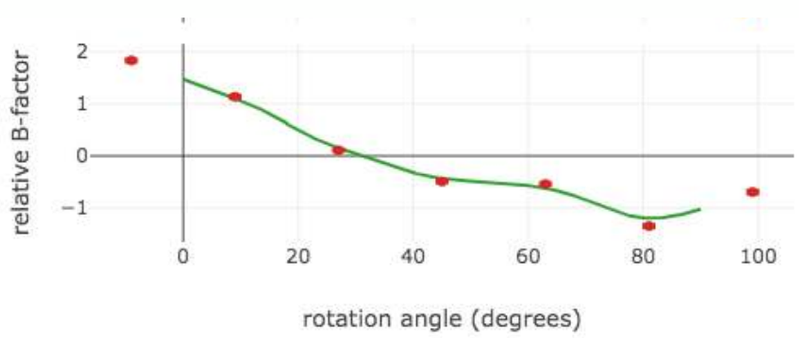
Integrated reflections

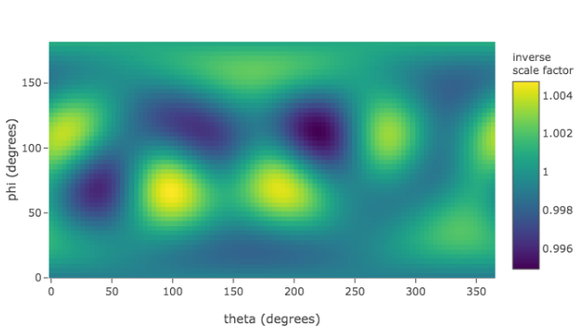
↓

Integrated reflections, in correct point group

↓

Scaled (corrected) intensities and sigmas





AFTER SCALING; MERGING AND TRUNCATING

Merge symmetry equivalent reflections to get the best estimate (optionally keeping anomalous pairs separate).

Some structure determination programs require structure factor amplitudes:

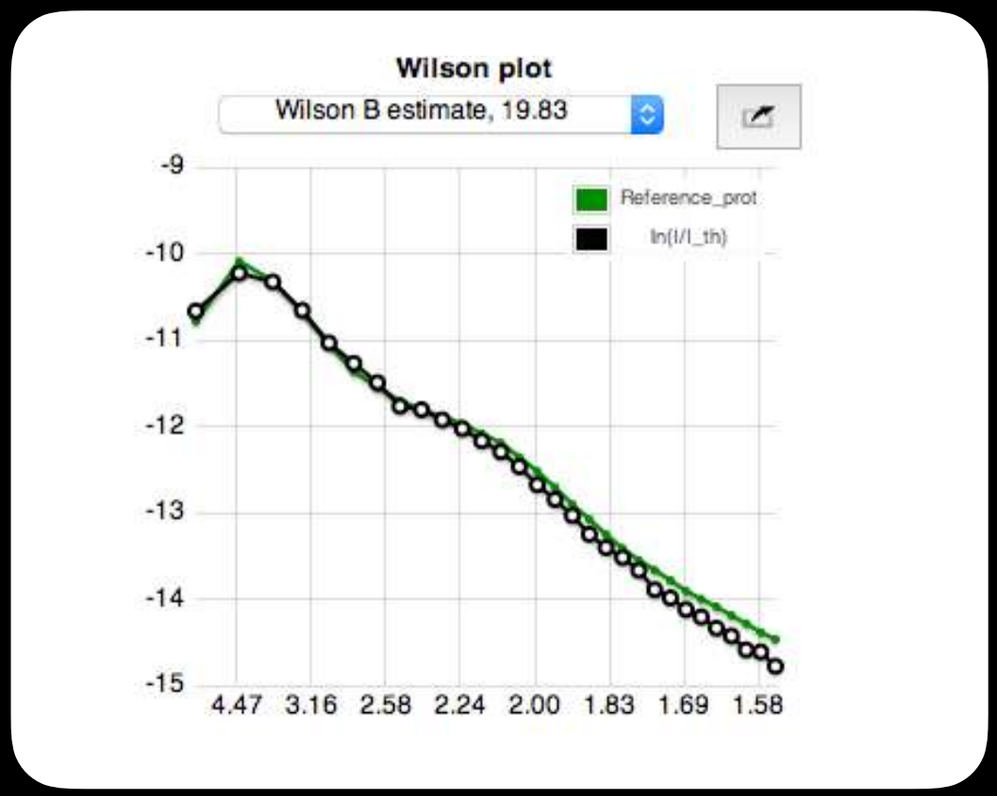
$$|F| = \sqrt{I}$$

Weak reflections can have negative intensities due to background subtraction!

French Wilson algorithm (1978), uses Bayesian statistics to calculate positive structure factors for all intensities (based on I and σ(I)).

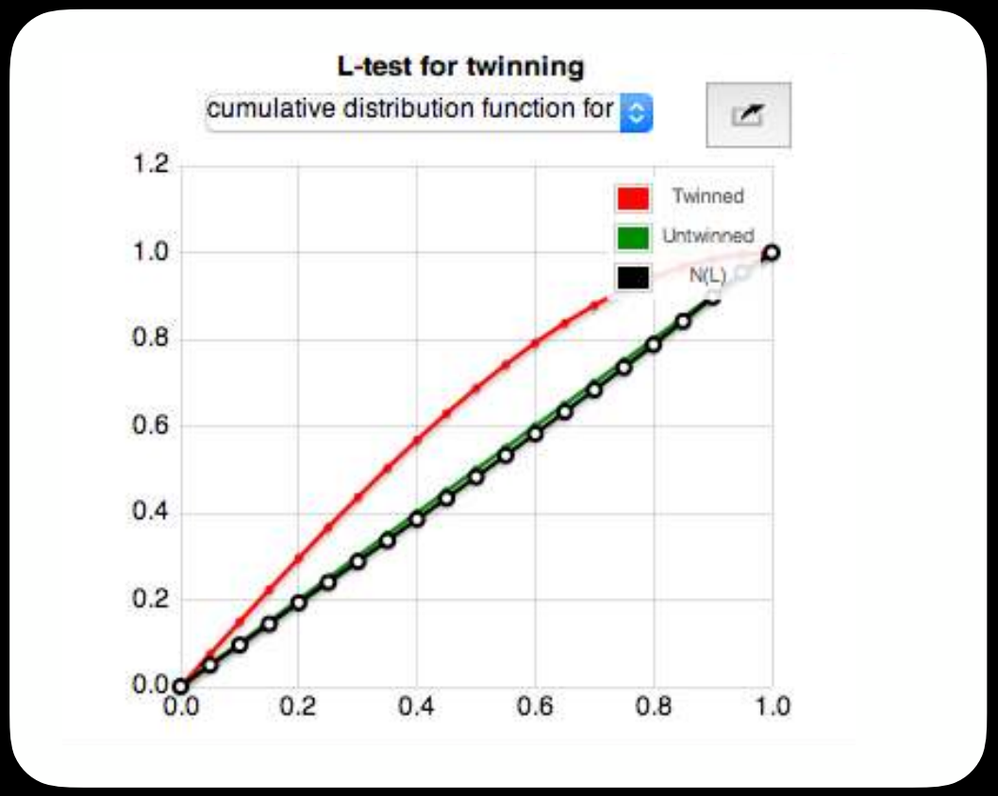
Program can use: dials.merge, ctruncate (ccp4),

INTENSITY STATISTICS



$$\langle I \rangle = C \exp(-\frac{B}{2}\frac{1}{d^2})$$

Average intensity falls with resolution due to atomic motions (B-factors).



Yeates–Padilla test (L-test)
Different expected cumulative intensity distribution depending on twin fraction.

RECAP: DATA REDUCTION STEP3

Merge equivalent reflections, convert to Fs*

Integrated reflections

↓

Integrated reflections, in correct point group

↓

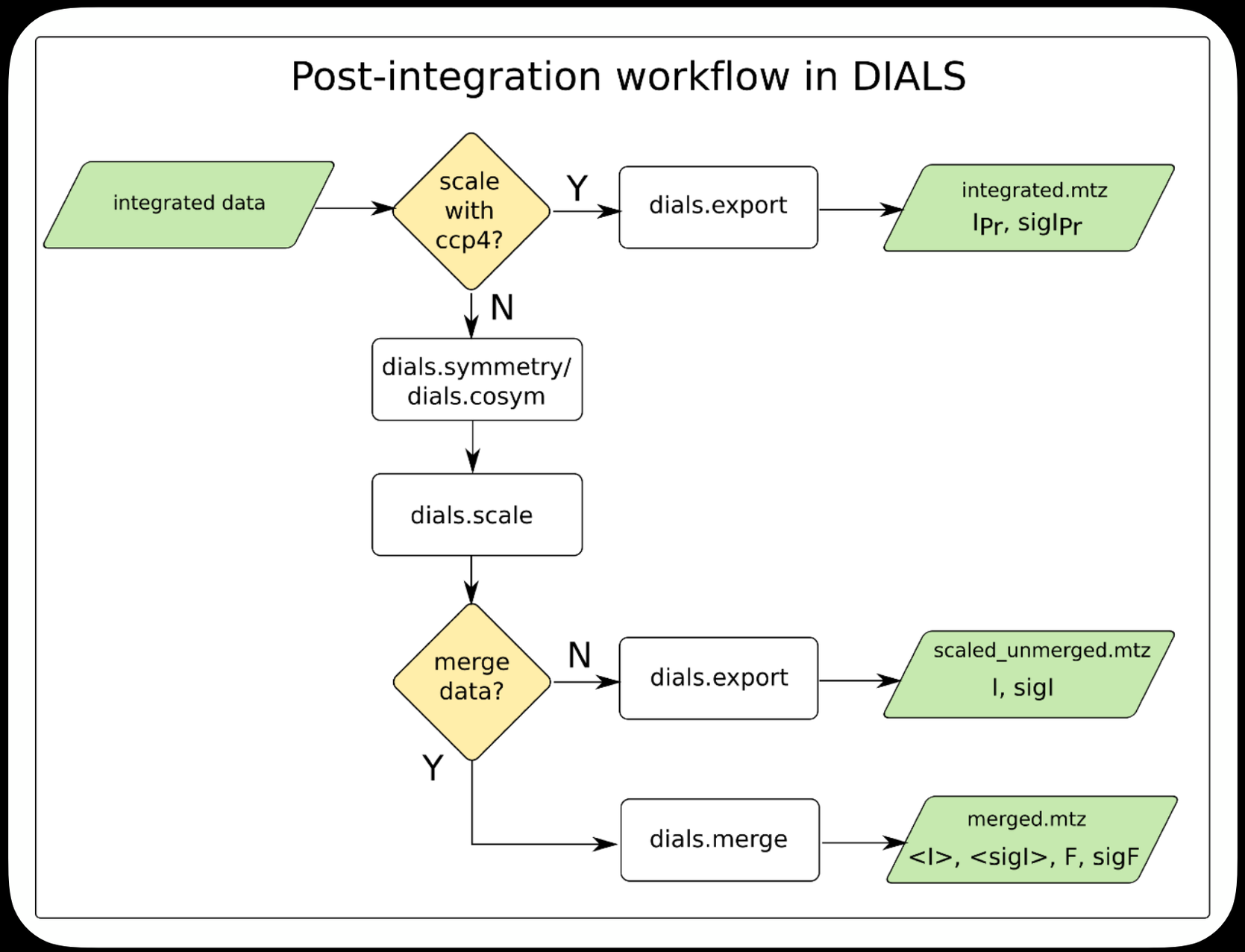
Scaled (corrected) intensities and sigmas

↓

Merged (symmetry-unique) set of I/F

*Some move away from converting I to F - e.g. Phaser can use I

DIALS DATA REDUCTION WORKFLOW



CCP4 equivalent: Pointless ⟶ Aimless ⟶ Ctruncate

ANOMALOUS DATA

Anomalous scattering: when Friedel's law is broken due to strong x-ray absorbance.

$I_{(hkl)}$!= $I_{(-h,-k,-l)}$ I+ != I-

Should I+, I- be treated separately in scaling? It depends…

My general rule of thumb: If anomalous signal is strong, and you have sufficient completeness/multiplicity, then probably better to treat anomalous pairs separately.

Anomalous metrics

```
-------------Summary of merging statistics--------------

                                Overall     Low      High
High resolution limit              1.80    4.88     1.80
Low resolution limit              58.76   58.79     1.83
Completeness                      99.6   100.0     92.7
Multiplicity                      17.6    29.4      6.6
I/sigma                           26.1   112.2      1.3
Rmerge(I)                        0.086   0.053    0.703
Rmerge(I+/-)                     0.082   0.049    0.666
Rmeas(I)                         0.088   0.054    0.766
Rmeas(I+/-)                      0.086   0.051    0.784
Rpim(I)                          0.018   0.010    0.291
Rpim(I+/-)                       0.025   0.013    0.401
CC half                          0.999   0.999    0.520
Anomalous completeness            98.3   100.0     85.9
Anomalous multiplicity             8.8    14.7      3.4
Anomalous correlation            0.415   0.445   -0.045
Anomalous slope                  0.757
dF/F                             0.050
dI/s(dI)                         1.047
Total observations              445648   37471     8003
Total unique                     25368    1276     1207
```

ANOMALOUS DATA

XDS: FRIEDEL'S_LAW=FALSE

Aimless: Friedel pairs always equivalent for scaling model determination

Important dials.scale options for strongly anomalous data (DIALS v3.5 onwards):

`anomalous=True` ← Separate Friedel pairs in scaling

`absorption_level=low/medium/high`

Suitable if high absorption expected
e.g. long wavelength or heavy atoms
in protein, I23 beamline at DLS

MULTI-CRYSTAL DATA REDUCTION

Single-sweep — One sweeps of large rotation width

Multi-sweep — Multiple sweeps of large rotation width

Multi-crystal — Many sweeps of small rotation width (~10 degrees)

Serial crystallography — Measurements on thousands of crystals, still measurements, XFELS, etc.

"Routine"*

"Exciting":
More active research, new developments, new things to be learnt

*overgeneralisation.
Still extremely important! 🙂

MULTI-CRYSTAL DATA REDUCTION

Multi-sweep processing

Considerations:

- Do we have isomorphism, is there some radiation damage within sweeps
- When do we have sufficient completeness, multiplicity, I/sigma

Can integrate sweeps individually, scale together, assess isomorphism and merge sweeps that agree.

Multi-crystal processing

Problems:

- Difficult to index/ambiguity in lattice symmetry
- Potential indexing ambiguities when combining sweeps
- Non-isomorphism, radiation damage
- Each sweep has low completeness, low multiplicity. Need to combine lots of 'lower quality' measurements.

↓

xia2.multiplex is the automated tool for doing multi-crystal data reduction, using DIALS data reduction programs. See the talk this afternoon by Selina Storm.

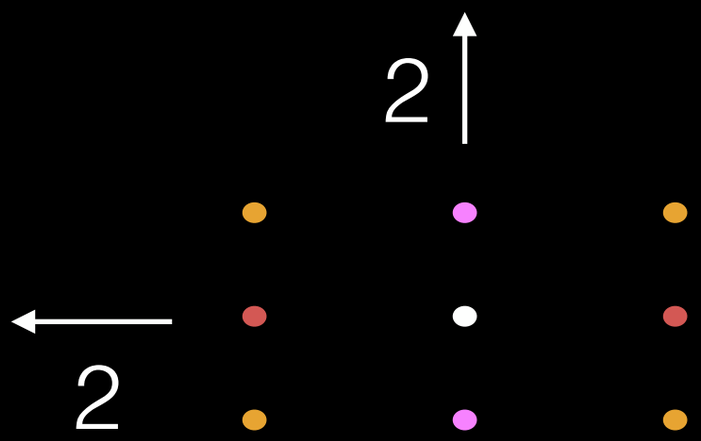
SUMMARY

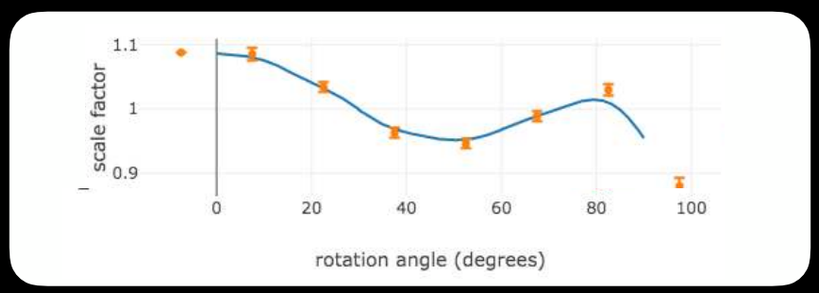
- Data reduction is the process of scaling and merging integrated intensities.

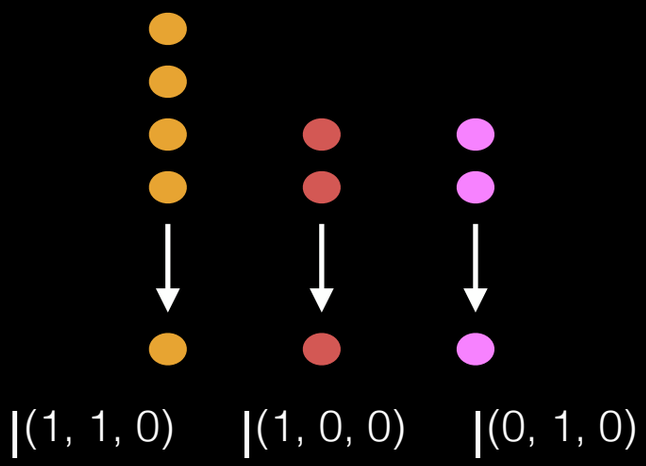
$$I_1^{(1,1,1)}, \sigma_1^{(1,1,1)}, I_2^{(1,1,1)}, \sigma_2^{(1,1,1)}, \ldots \rightarrow \mathbf{I}^{(1,1,1)}, \boldsymbol{\sigma}^{(1,1,1)}$$

- Scaling attempts to correct for experimental effects to make the data internally consistent.

- Critically evaluate the merging statistics and plots after scaling. Is the dataset complete, was radiation damage significant? Should the resolution be cut back? Weak data is not necessarily bad data. If many crystals are being merged, which data should be included?







ACKNOWLEDGEMENTS

DIALS at Diamond: Gwyndaf Evans, Graeme Winter, David Waterman, Richard Gildea, Nicholas Devenish, Luis Fuentes-Montero, Markus Gerstel, Ben Williams, James Parkhurst, Melanie Vollmar, Elena Pascal, David McDonagh, Noemi Frisina

Thanks to many DLS staff for datasets, feedback.

Thank you for your attention!