Refinements of High Pressure Data

Karen Friese[1], Jose Maria Posse[2] & Andrzej Grzechnik[3]

[1]Jülich Centre for Neutron Science, Research Centre Jülich
[2]Condensed Matter Physics, University of the Basque Country
[3]Institute of Crystallography, RWTH University Aachen

5th of July 2012

Specialities of high pressure data

- Limited resolution in sinθ/λ
- Limited access in reciprocal space in certain directions
- intensities are affected by errors
  - diamond dips
  - falsified due to components of the DAC
  - overlap with diamond reflections
  - shadowed reflections

→ Outliers may be present!

→ High pressure data are of poor quality

Measure for the quality of a dataset: the internal R-value

$$R_{int} = \sum_i \sum_j \frac{(I_j - \bar{I}_i)}{\bar{I}_i}$$

where i runs over all independent reflections and j over all symmetry equivalent reflections corresponding to the i-th independent reflection.

Rule of thumb: the final agreement factor for the refinement should be below the internal R-value

Redundancy

$$\text{Redundancy} = \frac{\text{No. measured reflections}}{\text{No. crystallographically independent reflection}}$$

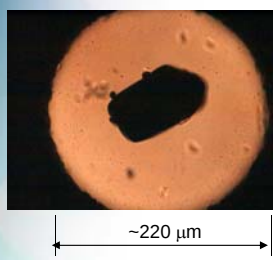
The more reflections are merged

- the smaller the importance of outliers
- the easier the identification of outliers

→ measure as many reflections as possible!

Example Dataset

(Laboratory source; Mo Kα, 5.82 GPa, 2 runs, Almax-type DAC, 1:1 pentane to isopentane)



BaV_6O_{11} (Z=2)

Lattice parameter: a ≈ 5.8 Å, c ≈ 13.2 Å

Space group $P6_3/mmc$ at ambient conditions:
1 Barium ($56e^-$), 3Vanadium ($23e^-$), 3 Oxygen ($8e^-$)

Space group $P6_3mc$ at high pressures (> 3 GPa)

1 Barium, 4 Vanadium, 5 Oxygen

Scaling of datasets

Run 1	Run 2	Run 3	Run 4
Scale factor 1	Scale factor 2	Scale factor 3	Scale factor 4

Intensities of run 2, 3, and 4 have to be multipled by scale factors to fit the intensities of run 1

Advantage: increase of redundancy

All runs
Scale factor 1

scaling is usually done via common reflections in the partial datasets

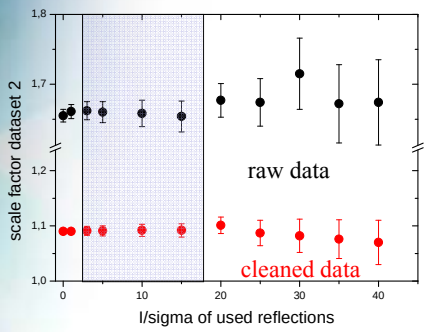
only reflections over a certain I/σ ratio limit are chosen

Falsified reflection intensities introduce errors in the scale factors

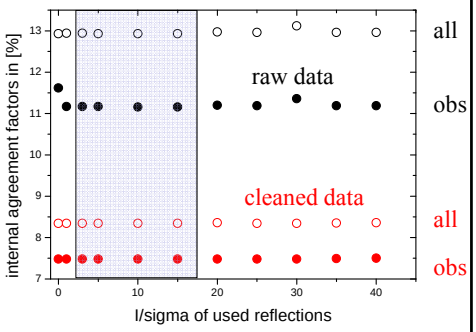
Any error in the scale factor influences all reflections in the run !!!

Which I/σ limit should be used for choosing the reflections for scaling?

Scale factor in dependence of I/σ of chosen reflections



Internal agreement factors in dependence of I/σ of chosen reflections



According to our experience I/σ should be between 3 to 15

Improving the Dataset

obs/all

11.16/12.93

- Internal R-value after integration and scaling

Improving the Dataset

obs/all

11.16/12.93

↓

10.12/11.34

- Internal R-value after integration and scaling
- after correction for diamond anvils (no shadowing by gasket)

Improving the dataset

obs/all

11.16/12.93

↓

10.12/11.34

↓

10.04/11.25

- Internal R-value after integration and scaling
- after correction for diamond anvils (no shadowing by gasket)
- after correction for absorption of crystal

Identification of Outliers

- On the basis of symmetry equivalent reflections

  the more reflections are averaged, the easier to find the outliers

  → the higher the symmetry and redundancy, the better

  in the initial stages one can use "approximate" symmetries to make identification of outliers easier (e.g. Laue symmetry)

- On the basis of the refinement

  F(obs)/F(calc) diagrams

  F(obs)/ w| F(obs)-F(calc) | diagrams

Identification of Outliers: Symmetry Equivalent Reflections

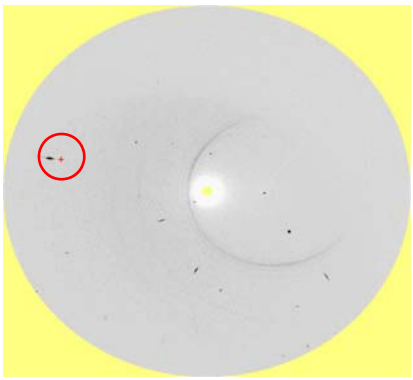
First stage: reflections with I-I(average)>25σ(I(ave))

			I	σ(I)
?	0	5 -8	473549.1	9500.2
	-5	5 -8	926865.1	16891.2
	5	-5 -8	20233.0	8700.6
	-1	2 -2	235417.7	4534.3
	-2	1 -2	307915.0	9683.5
	-2	1 -2	290463.4	8889.1
	-1	2 -2	1271.9	7205.8
	-2	1 -2	342020.5	10210.5
	0	0 8	395706.3	3593.2
	0	0 8	393350.8	14390.3
	0	0 8	415274.4	9004.3
	0	0 8	369744.4	10256.1
	0	0 8	385701.7	12387.9
	0	0 8	438368.4	14201.6
	0	0 8	391667.8	12969.5
	0	0 8	450817.9	10454.7
	0	0 8	411253.3	8985.0
	0	0 8	301623.6	8961.2
	0	0 8	399260.8	10202.7

strongest reflection in the dataset

shadowed reflections

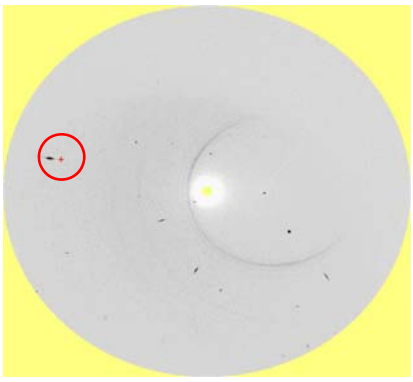
Checking reflections on the original frames

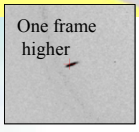


$\bar{5}5\bar{8}$ I=926865.1 σ(I)=16891.2

$5\bar{5}\bar{8}$ I=20233.0 σ(I)= 8700.6

Checking reflections on the original frames





$\bar{5}5\bar{8}$ I=926865.1 σ(I)=16891.2

$5\bar{5}\bar{8}$ I=20233.0 σ(I)= 8700.6

Increasing redundancy: adding a center of symmetry

$P6_3mc$				
0	5	-8	473549.1	9500.2
-5	5	-8	926865.1	16891.2
5	-5	-8	20233.0	8700.6

$P6_3/mmc$				
0	5	8	237917.1	5993.2
5	-5	8	1046.5	11860.2
5	-5	-8	20290.2	8725.2
-5	5	-8	929489.1	16939.0
-5	5	8	842.5	8424.3

Improving the dataset

obs/all

11.16/12.93

↓

10.12/11.34

↓

10.04/11.25

↓

8.28/9.57

- Internal R-value after integration and scaling
- after correction for diamond anvils (no shadowing by gasket)
- after correction for absorption of crystal
- initial exclusion of falsified reflections (3 shadowed + 1 diamond)

Identification of Outliers: Symmetry Equivalent Reflections

Next stages: reflections with I-I(average)>xxσ(I(ave))

I σ(I)

h k l ……. ……
…… ……. ……
…….. ……. ……

h k l ……. ……
…… ……. ……
…… ……. ……
…… ……. ……

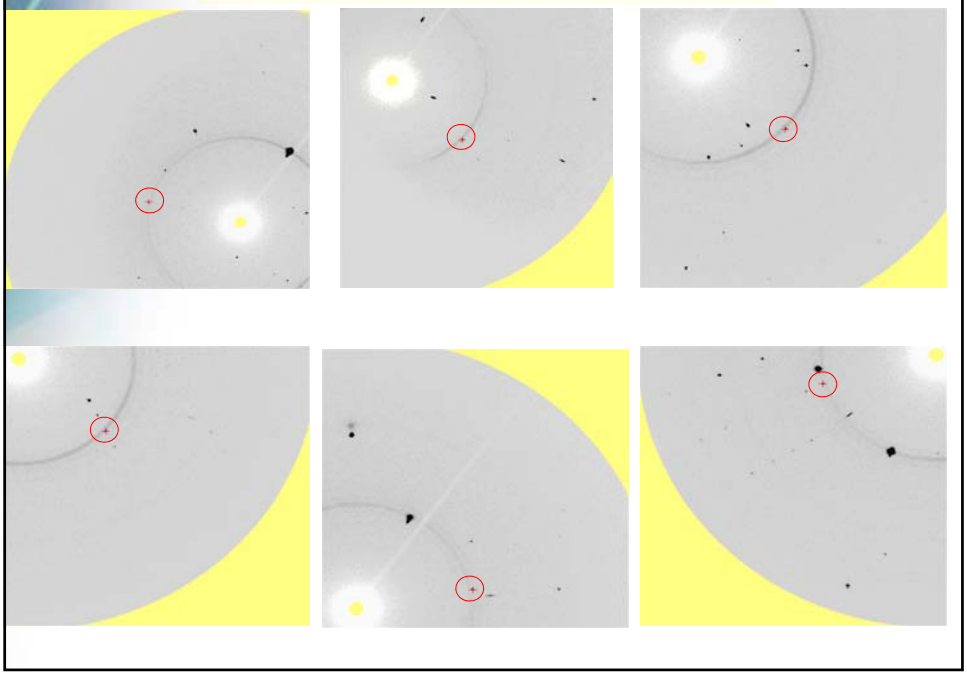
- **Change the criteria and repeat**

Reflections falling on a gasket ring

First gasket ring sinθ/λ=0.245 Å^{-1}

sinθ/λ 0.249Å^{-1}									
	0	1	-6	37999.4	2508.3				
	1	0	-6	43012.5	10080.6	0	-1	-6	33644.9	8301.3
	1	0	-6	42179.6	8953.9	1	0	-6	36127.9	9828.2
	1	0	-6	736.9	7368.7	0	-1	-6	39565.2	9850.8
	-1	1	-6	60834.1	10070.0	0	-1	-6	40667.8	9114.2
	0	-1	-6	41443.9	9130.0	0	1	-6	40718.1	9734.4
	-1	0	-6	48138.3	8794.1	0	-1	-6	41116.5	11286.6
	1	0	-6	35725.2	8622.5	0	1	-6	28081.3	9637.7

Checking reflections on the original frames

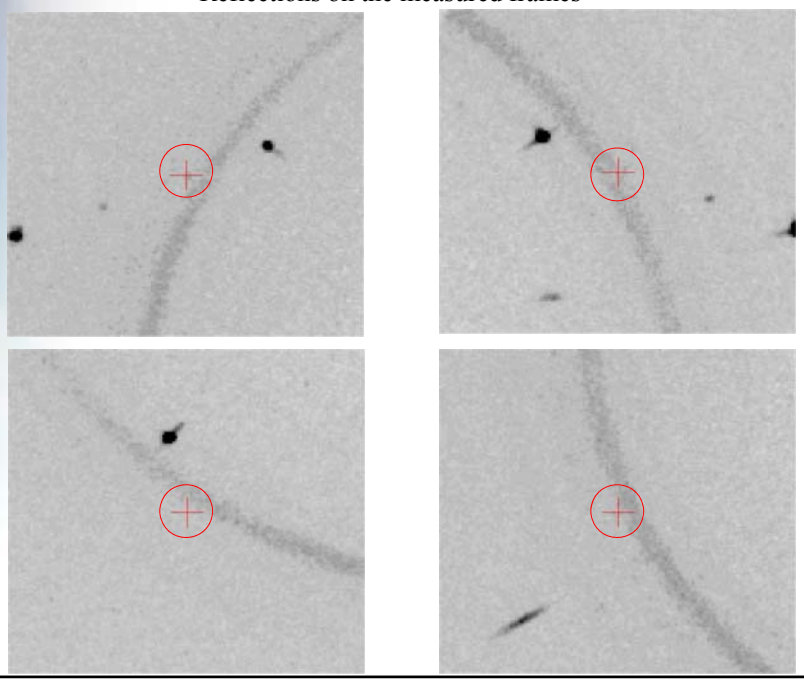


Weak or unobserved reflections on gasket

0	2	4	6221.3	2515.8	2.5σ				
2	-2	4	742.4	7424.6	-2	2	4	23129.5	8921.0
-2	2	4	745.6	7456.2	2	-2	4	1066.8	10667.7
-2	2	4	10668.6	15560.3	2	-2	4	3319.6	7047.4
2	-2	4	703.0	7029.9	2	-2	4	11725.8	9544.0
2	-2	4	932.5	9325.4	-2	2	4	16548.5	7099.7
2	-2	4	944.1	9440.0	-2	2	4	6599.6	6936.5
-2	2	4	3751.0	7659.5					

>1σ

Reflections on the measured frames



Identification of Outliers

- On the basis of symmetry equivalent reflections

  the more reflections are averaged, the easier to find the outliers

  → the higher the symmetry and redundancy, the better

  in the initial stages one can use “approximate” symmetries to make identification of outliers easier (e.g. Laue symmetry)

- On the basis of the refinement

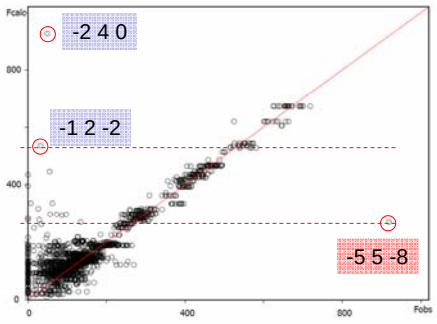
  F(obs)/F(calc) diagrams

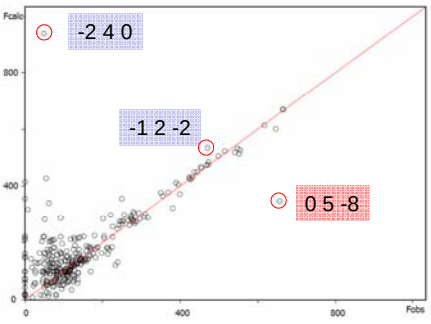
  F(obs)/ w| F(obs)-F(calc) | diagrams

Outliers from refinement: F(obs)/F(calc) plots

Shadowed reflections:
F(obs)<<F(calc)

Diamond reflections:
F(obs)>>F(calc)





raw data

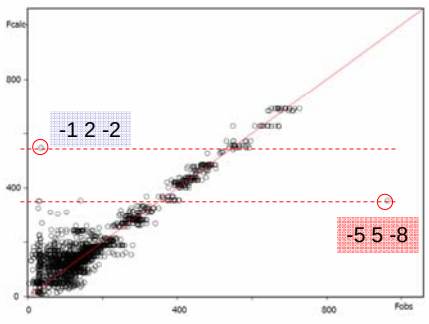
non-averaged
6.61/21.62

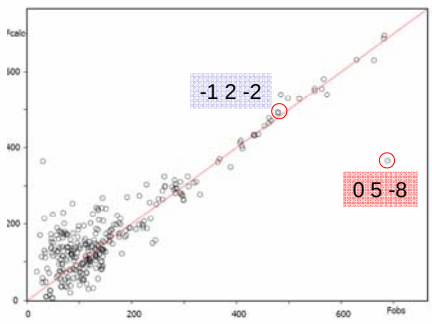
averaged
10.56/33.26

Outliers from refinement: F(obs)/F(calc) plots

Shadowed reflections:
F(obs)<<F(calc)

Diamond reflections:
F(obs)>>F(calc)



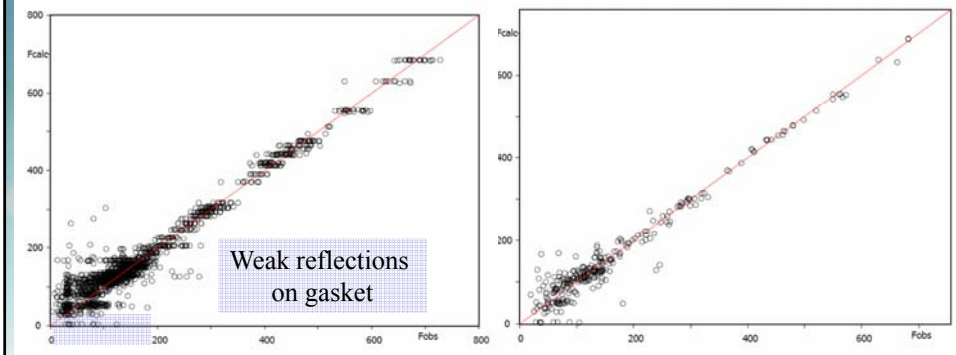


Absorption corrected data

non-averaged
6.25/19.23

averaged
10.73/25.10

Outliers from refinement: F(obs)/F(calc) plots

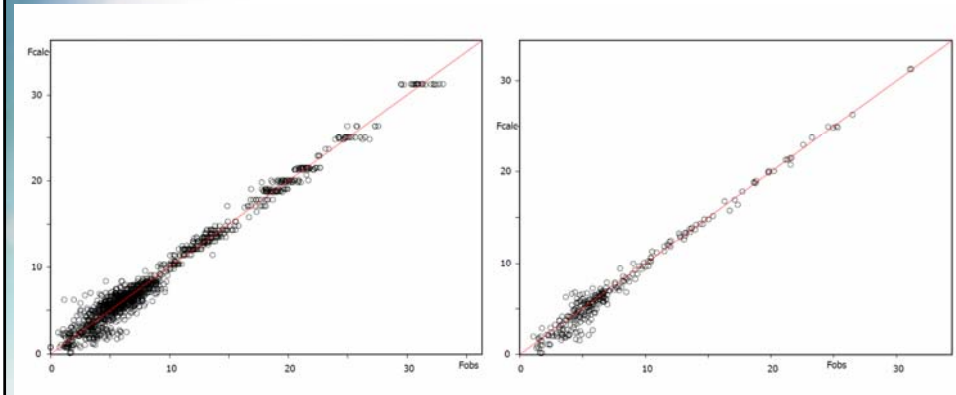


Absorption corrected data without biggest outliers

non-averaged
4.48/11.65

averaged
4.66/11.22

Outliers from refinement: F(obs)/F(calc) plots



Final dataset without outliers

non-averaged
3.63/7.93

averaged
2.51/8.23

Improving the dataset

obs/all

11.16/12.93

↓

10.12/11.34

↓

10.04/11.25

↓

8.28/9.57

↓

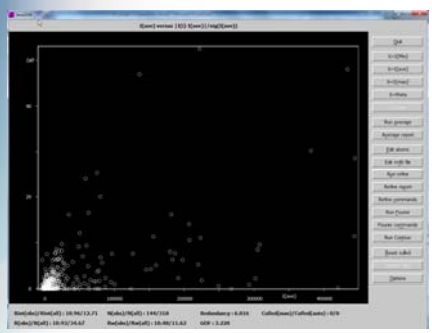
7.48/8.34

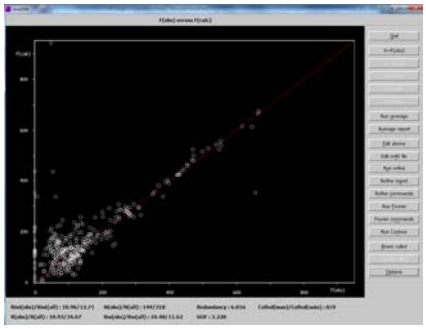
- Internal R-value after integration and scaling
- after correction for diamond anvils (no shadowing by gasket)
- after correction for absorption of crystal
- initial exclusion of falsified reflections (3 shadowed + 1 diamond)
- Further rejection of outliers

The whole procedure

Integrated data		Absorption correction (crystal), Shadowed reflections
decay correction		Only if synchrotron
Absorption corrections shadowed reflections		Absorb-program
Scaling + averaging	repeat	use higher (pseudo)symmetry if necessary and possible
elimination of outliers		
Scaling + averaging		use real symmetry
Structure solution		if there is no structural model
refinement		limit the number of parameters
elimination of outliers	repeat	F(obs)/F(calc) plots
Scaling + averaging		
final reflection data		
final refinement		

Identification of Outliers: Helpful options in Jana2006





→ Refinement session

Comparison of the results with good and bad datasets

	wR(all) [%] uncleaned	wR(all) [%] cleaned
• Solution (Sir97)	17.16 2 oxygen missing	8.34 1 oxygen missing
• Scale factor	18.27	7.05
• Coordinates	18.18	4.05
• Inclusion of missing atom(s)	13.74	2.84
• Coordinates of missing atom	14.45 refinement unstable	2.64
• U_{iso} Ba, V	13.26 refinement unstable	2.52
• U_{iso} O	10.57 1V and 3 O negative	2.45 1V and 1 O negative
• Ba aniso (1V negative)	10.57	2.38
• Trial: V aniso (2 V negative)	10.23	2.29
• U_{iso} of part of V/O set equal	11.17 4 O negative	2.50
Number of symmetrically independent reflections:	318	292

Ideal parameter to data ratio

Rule of thumb: 10 data points for 1 parameter

May be difficult to reach in high pressure experiments

Two solutions:

- increase the number of data points

  Shift to shorter wavelengths: synchrotron
  Choose a cell with a maximum opening angle
  Reduce the number of bad reflections

- limit the number of parameters

Ideal parameter to data ratio

Rule of thumb: 10 data points for 1 parameter

May be difficult to reach in high pressure experiments

Two solutions:

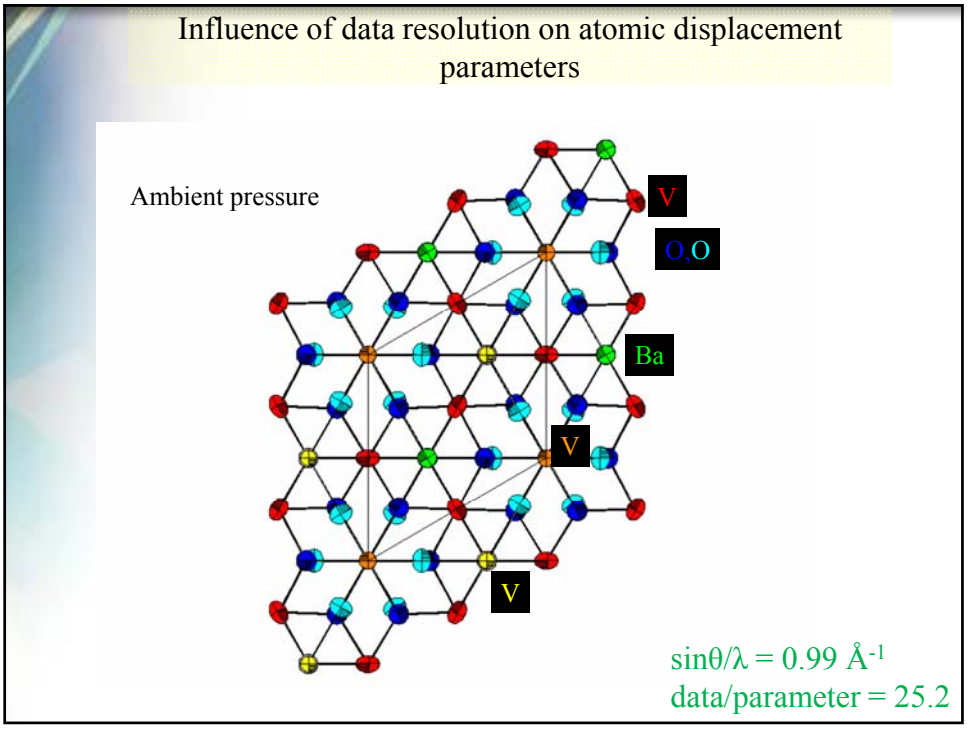
- increase the number of data points

  Shift to shorter wavelengths: synchrotron
  Choose a cell with a maximum opening angle
  Reduce the number of bad reflections

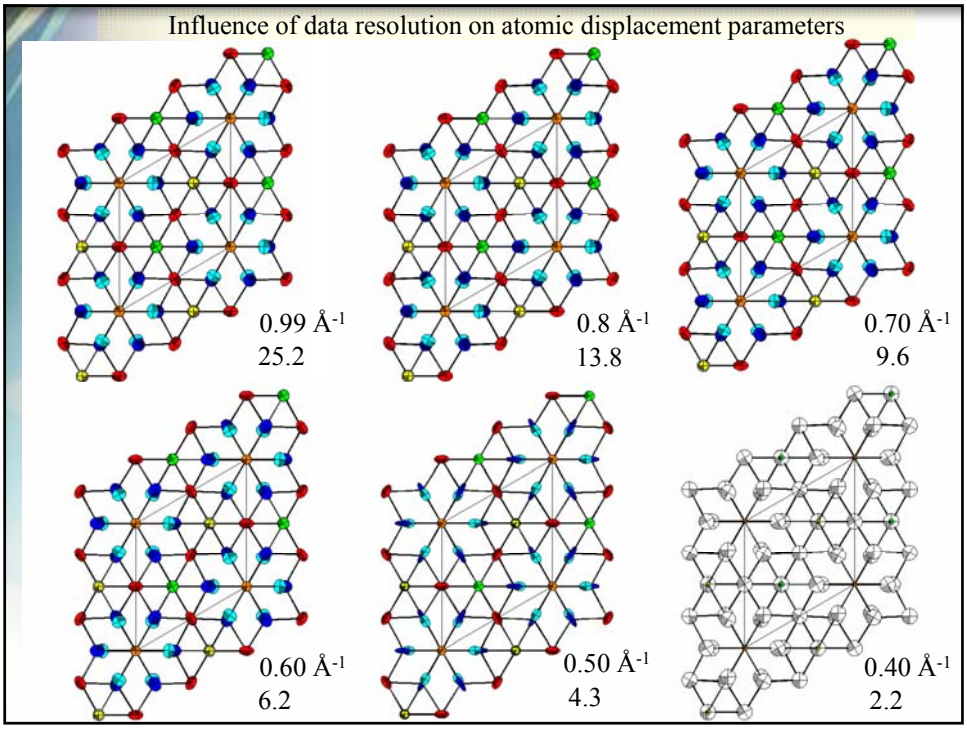
- limit the number of parameters

Which Ones?

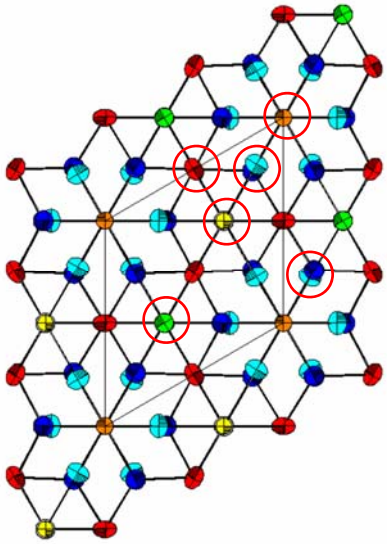
Influence of data resolution on atomic displacement parameters



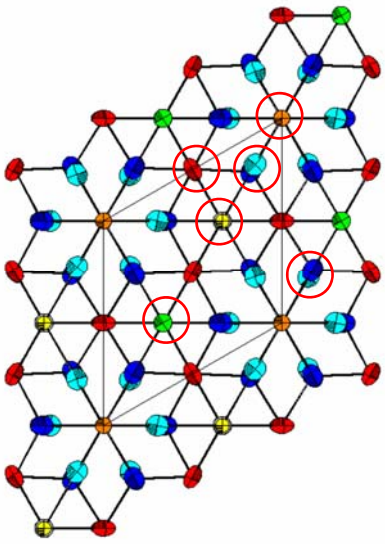
Influence of data resolution on atomic displacement parameters



Influence of data resolution on atomic displacement parameters



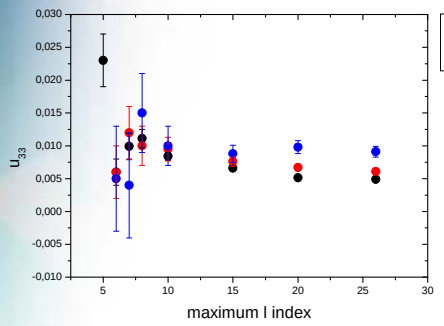
Ambient pressure: Resolution 0.99
Data/Parameter 25.2

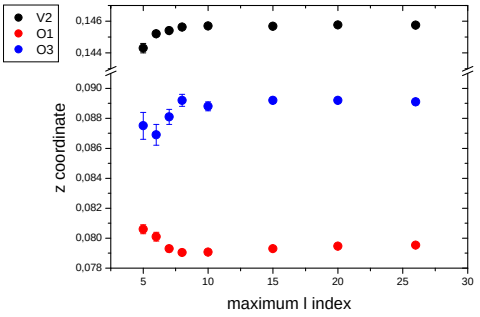


1.18 GPa: Resolution 0.67
Data/Parameter 12.6

Influence of limited resolution in certain directions

Example: limited resolution in c*

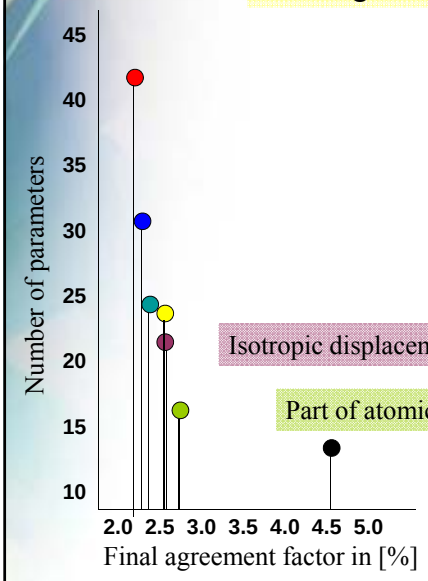




anisotropic displacement parameters

atomic coordinates

Limiting the number of parameters



all atoms anisotropic

O atoms isotropic

V atoms isotropic

Ba atom isotropic

Isotropic displacement parameters restricted via pseudosymmetry

Part of atomic coordinates fixed to higher (pseudo)symmetry

Refined with higher (pseudo)symmetry

292 symmetrically independent reflections

Limiting the number parameters in the refinement

Displacement Parameters: 6 per atom on a general position

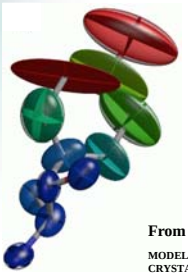
- fix the displacement parameters to reasonable values (e.g ambient pressure)
- use isotropic displacement parameters
- use higher pseudosymmetry (if present) to restrict the number of parameters
- TLS refinement (Translation-Libration-Screw-Motion)
  (Schomaker & Trueblood, Acta Cryst. B24, 63,1968)

Limiting the number parameters in the refinement

Displacement Parameters:

- fix the displacement parameters to reasonable values (e.g ambient pressure)
- use isotropic displacement parameters
- use higher pseudosymmetry (if present) to restrict the number of parameters
- TLS refinement: Translation- Libration-Screw-Motion
  (Schomaker & Trueblood, Acta Cryst. B24, 63,1968)

Unrestrained: 12 atoms~72 Parameters



TLS: 20 Parameters

From IUCr99 COMPUTING SCHOOL

MODELLING RIGID-BODY THERMAL MOTION IN MACROMOLECULAR CRYSTAL STRUCTURE REFINEMENT. Ian J. Tickle and David S. Moss, Department of Crystallography, Birkbeck College, University of London

Limiting the number parameters in the refinement

Geometrical constraints:

- restrict bond lengths
- restrict molecular/polyhedral geometry

Limiting the number parameters in the refinement

Approximate the structure

- use rigid units
- refine an average structure with higher symmetry (if present)
- fix part of the atomic positions
- in the case of a phase transition: mode refinement

Limiting the number parameters in the refinement

Approximate the structure (serious cases)

- use rigid units
- refine an average structure with higher symmetry (if present)
- fix part of the atomic positions
- in the case of a phase transition: mode refinement

Secondary Modes

P6_3/mmc[0,0,0]GM1+ (a) 194 P6_3/mmc

0.000(11)	[V2:e]A1(a)
0.03(10)	[O1:k]A'_1(a)
-0.01(5)	[O1:k]A'_2(a)
-0.10(4)	[O2:h]A1(a)
0.02(4)	[O3:f]A1(a)

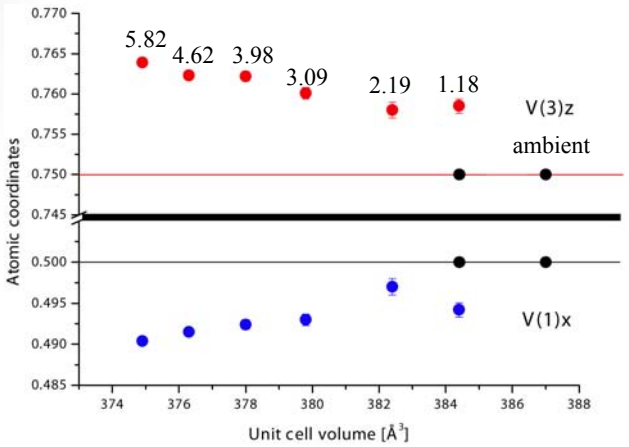
Primary Modes

P6_3/mmc[0,0,0]GM2- (a) 186 P6_3mc

0.00000	[Ba:c]A2"(a)
0.23(2)	[V1:g]Bu_1(a)
0.232(19)	[V1:g]Bu_2(a)
0.297(11)	[V2:e]A1(a)
0.260(9)	[V3:d]A2"(a)
0.04(10)	[O1:k]A'_1(a)
0.13(5)	[O1:k]A'_2(a)
0.15(6)	[O2:h]B1(a)
0.17(4)	[O3:f]A1(a)

Mode amplitudes for the phase transition $P6_3/mmc \rightarrow P6_3mc$ in BaV_6O_{11} calculated with ISODISTORT (Stokes, Campbell & Hatch)

Looking at data trends



$P6_3mc$

$P6_3/mmc$

$P6_3/mmc$

$P6_3mc$

Conclusions

- Invest time and effort in the experiment
- Collect data at different pressure points
- Make reconstructions of reciprocal space
- Check carefully for outliers
- Refine carefully and stepwise: make sure adding parameters improves the model (Hamilton Test)
- Limit the number of parameters
- Be critical about the data

The End

Comparing different structural models : Hamilton's test

W.C. Hamilton, Acta Crystallogr. 18, 502-510 (1965)
Significance tests on the Crystallographic R-Factor

Does the increase of parameters to a model lead to a significant improvement of the model?

Comparison of an R-factor ratio to tabulated values

R-factor ratio: wR(model B)/wR(model A)
Model B: model with restriction
Model A: model without restrictions

If the R-factor ratio is larger than the tabulated value
→ the hypothesis can be rejected

Examples for the use of the Hamilton test

- independent structure refinements
- different structural models
  e.g. anisotropic/isotropic/partially anisotropic
- structural models with refined and fixed (=estimated) coordinates
- comparison of two absolute configurations
- two refinements: one with fixed molecular geometry, the other with free geometry
- refinements with different space group symmetries

Some points which have to be observed

- Test is based on wR(F)
- the number of reflections in the two models has to be equal
- If you use geometrical constraints, think carefully about the number of parameters
- the tabulated values correspond to a certain probability level

e.g. $R_{b,n-m,0.50}$ indicates that the hypothesis cannot be rejected /can be rejected at the 50% level i.e. we are wrong half the time if we reject (or accept) a hypothesis at this level.

Example: What is the correct space group at a pressure of 5.82 GPa?

From the refinment:

Number of reflections n=292

Model A ($P6_3/mmc$): 14 Parameters = m_A R_A = wR(all) = 0,0311

Model B ($P6_3mc$): 22 Parameters = m_B R_B = wR(all) = 0,0255

- Hypothesis: Model A is better than model B

  Dimension of the hypothesis $m_B - m_A = 8$

  Number of degrees of freedom $n - m_B = 292 - 22 = 270$

Interpolated value at a 0.005 significance level:

$$R_{8,270,0.005} \cong 1 + 120/270(R_{8,120,0.005} - 1) = 1 + 120/270(1.093 - 1) = 1.0413$$

$$R = R_A/R_B = 1.219 > 1.0413$$

- Hypothesis can be rejected at a 0.005 probability level
  → model B is better → the structure is acentric