Data evaluation, integration and analysis I

L.M.J. Kroon-Batenburg
Utrecht University
The Netherlands
Content

Data processing with EVAL:
• Crystal orientation and goniometer
• Indexing
• Cell matrix refinement and metadata
• Reflection profiles and integration
• Intensities, corrections, scaling, merging
• Error estimates and contributions to esd
• Special cases:
  • Multiscan data
  • Twin lattices
  • incommensurate
  • Diffuse scattering
Diffractometer

© Garland Science 2010
Eval Program Suite

Eval14 and eval15 are integration methods for single crystal X-ray diffraction on area detectors. Both methods use knowledge about the exact experimental setup.

- Eval14 applies boxsummation (BPB) with predicted reflection boundaries.
- Eval15 applies profile fitting with predicted reflection profiles

The eval15 method is described in:

A.M.M. Schreurs, X. Xian and L.M.J. Kroon-Batenburg
EVAL15: a diffraction data integration method based on ab initio predicted profiles


The eval14 method is described in:

An intensity evaluation method: EVAL-14


A reprint (Copyright © International Union of Crystallography J. Appl. Cryst. 36, 220-229) can be found here (PDF file, 820 kb).

runs on Linux
EVAL flow diagram
Lattice matrices and orientations

\[ O = \begin{pmatrix} a & 0 & 0 \\ b \cos \gamma & b \sin \gamma & 0 \\ c \cos \beta & \frac{c(\cos \alpha - \cos \beta \cos \gamma)}{\sin \gamma} & \frac{v}{ab \sin \gamma} \end{pmatrix} \]

\[ D = R_0 \cdot O \]

\[ R_\omega R_\chi R_\phi R_0 \cdot O \]

*often \( O^T \) is used as orthogonalization matrix

Data evaluation, integration and analysis I
Crystal and reciprocal lattices

\[ D = d_{hk} b_{hk} c_{hk} \]

\[ R = D^{-1} \]

\[ d_{*hk} = R \begin{pmatrix} h \\ k \\ l \end{pmatrix} = h a^* + k b^* + l c^* = \frac{1}{d_{hk}} \]
Thaumatin

ADSC detector Diamond Light Source
Scan around horizontal spindle axis 0.5° increments
\[ \eta = k_1 - k_0 \]

Bragg reflection if \( d^* \) co-aligns with \( \eta \)

\[ |k_0| = 1/\lambda \]

Data evaluation, integration and analysis I
Indexing: find vectors $d^*$ and index these in $(a^*, b^*, c^*)$

Data evaluation, integration and analysis I
Indexing with Dirax

Peaks in 2D on the detector

200 c-vectors from file pk2.drx.

<table>
<thead>
<tr>
<th>Acl</th>
<th>nH</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>alpha</th>
<th>beta</th>
<th>gamma</th>
<th>Volume</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
<td>19</td>
<td>66.401</td>
<td>49.954</td>
<td>242.935</td>
<td>90.73</td>
<td>96.13</td>
<td>92.31</td>
<td>800438</td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>17</td>
<td>45.308</td>
<td>57.681</td>
<td>238.449</td>
<td>96.49</td>
<td>87.23</td>
<td>103.79</td>
<td>601201</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>17</td>
<td>36.578</td>
<td>58.401</td>
<td>225.349</td>
<td>97.39</td>
<td>89.40</td>
<td>104.43</td>
<td>462223</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>111</td>
<td>81.498</td>
<td>172.526</td>
<td>204.769</td>
<td>98.93</td>
<td>95.47</td>
<td>89.79</td>
<td>2831121</td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>118</td>
<td>56.728</td>
<td>57.489</td>
<td>150.502</td>
<td>90.88</td>
<td>91.66</td>
<td>90.99</td>
<td>490485</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>119</td>
<td>56.778</td>
<td>57.490</td>
<td>150.510</td>
<td>90.88</td>
<td>91.60</td>
<td>90.98</td>
<td>490958</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>107</td>
<td>56.772</td>
<td>57.550</td>
<td>150.147</td>
<td>89.25</td>
<td>91.89</td>
<td>89.01</td>
<td>490174</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>97</td>
<td>56.820</td>
<td>57.474</td>
<td>112.751</td>
<td>96.29</td>
<td>95.28</td>
<td>90.96</td>
<td>364300</td>
<td></td>
</tr>
</tbody>
</table>

Solutions limited by voltest

108 119 56.778 57.490 150.510 90.88 91.60 90.98 490958

selected ACL 108
Indexing with Dirax

DIRAX: 57.373 57.565 149.279 90.41 89.44 90.13 492982

3D peak search over a range of 100°
### Peakref: Refining peak positions

<table>
<thead>
<tr>
<th>ref</th>
<th>current</th>
<th>initial</th>
<th>change</th>
<th>accumulate</th>
<th>shift</th>
<th>refl</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Yes</td>
<td>57.64815</td>
<td>57.37284</td>
<td>0.27530</td>
<td>57.64815</td>
<td>0.11475</td>
</tr>
<tr>
<td>b</td>
<td>Con</td>
<td>57.64815</td>
<td>57.37284</td>
<td>0.27530</td>
<td>57.64815</td>
<td>0.11475</td>
</tr>
<tr>
<td>c</td>
<td>Yes</td>
<td>149.61322</td>
<td>149.27934</td>
<td>0.33388</td>
<td>149.61322</td>
<td>0.29856</td>
</tr>
<tr>
<td>alpha</td>
<td>Fix</td>
<td>90.00000</td>
<td>90.41067</td>
<td>0.00000</td>
<td>90.00000</td>
<td>0.20000</td>
</tr>
<tr>
<td>beta</td>
<td>Fix</td>
<td>90.00000</td>
<td>90.44347</td>
<td>0.00000</td>
<td>90.00000</td>
<td>0.20000</td>
</tr>
<tr>
<td>gamma</td>
<td>Fix</td>
<td>90.00000</td>
<td>90.13022</td>
<td>0.00000</td>
<td>90.00000</td>
<td>0.20000</td>
</tr>
<tr>
<td>orx</td>
<td>Yes</td>
<td>0.79530</td>
<td>0.78922</td>
<td>0.00608</td>
<td>0.79530</td>
<td>0.00022</td>
</tr>
<tr>
<td>oty</td>
<td>Yes</td>
<td>0.59701</td>
<td>0.60462</td>
<td>-0.00762</td>
<td>0.59701</td>
<td>0.00028</td>
</tr>
<tr>
<td>ora</td>
<td>Yes</td>
<td>-51.16824</td>
<td>-51.19841</td>
<td>-0.03017</td>
<td>-51.16824</td>
<td>0.20000</td>
</tr>
<tr>
<td>xtalx</td>
<td>No</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.10000</td>
</tr>
<tr>
<td>xtalz</td>
<td>No</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.10000</td>
</tr>
<tr>
<td>zover</td>
<td>Yes</td>
<td>-0.16770</td>
<td>0.00000</td>
<td>-0.16770</td>
<td>0.16770</td>
<td>0.10000</td>
</tr>
<tr>
<td>zover</td>
<td>Yes</td>
<td>0.42149</td>
<td>0.00000</td>
<td>0.42149</td>
<td>0.01610</td>
<td>0.10000</td>
</tr>
<tr>
<td>detrot</td>
<td>Yes</td>
<td>-0.57659</td>
<td>0.00000</td>
<td>-0.57659</td>
<td>-0.46698</td>
<td>0.10000</td>
</tr>
<tr>
<td>detrot</td>
<td>Yes</td>
<td>0.14871</td>
<td>0.00000</td>
<td>0.14871</td>
<td>0.14871</td>
<td>0.20000</td>
</tr>
<tr>
<td>detrot</td>
<td>Yes</td>
<td>0.27821</td>
<td>0.00000</td>
<td>0.27821</td>
<td>0.27821</td>
<td>0.20000</td>
</tr>
<tr>
<td>detrot</td>
<td>Yes</td>
<td>0.36010</td>
<td>0.00000</td>
<td>0.36010</td>
<td>0.36010</td>
<td>0.20000</td>
</tr>
</tbody>
</table>

| Vol  | 497210.97 | 491337.31 | 5873.66 | 497210.97 | 20152 |

| mm   | 0.03903 | 0.70960 | -0.67057 | 0.03903 | 20152 |
| mmAng| + 0.01495 | 0.28118 | -0.26623 | 0.01495 | 20152 |
| rotoutside| 0.03599 | 0.25945 | -0.22346 | 0.03599 | 742 |
| rotinside | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 19410 |
| rotall | + 0.00133 | 0.09680 | -0.09548 | 0.00133 | 20152 |
| res   | 0.01628 | 0.37799 | -0.36171 | 0.01628 | 13 |

- **Constrain Bravais lattice**
- **Refine cell parameters**
- **Refine instrumental parameters**
- **Pixel size 0.102 mm**
- **Scan width 0.5°**
- **Peaks from 100 images**
Metadata: image header

```
! HEADER_BYTES=512;
DIM=2;
BYTE_ORDER=little_endian;
TYPE=unsigned_short;
PIXEL_SIZE=0.1026;
BIN=2x2;
ADC=fast;
DETECTOR_SN=920;
DATE=Fri Jul 16 16:27:51 2010;
TIME=0.50;
DISTANCE=116.000;
OSC_RANGE=0.500;
PHI=0.000;
OSC_START=0.000;
TWOTHETA=0.000;
AXIS=phi;
WAVELENGTH=1.4880;
BEAM_CENTER_X=157.999;
BEAM_CENTER_Y=157.484;
CREV=1;
CCD=TH7899;
BIN_TYPE=HW;
ACC_TIME=1750;
UNIF_PED=1500;
IMAGE_PEDESTAL=40;
SIZE1=3072;
SIZE2=3072;
CCD_IMAGE_SATURATION=65535;
```
VIEW: datcol

Frame 200
Generation of shoeboxes for integration with EVAL15
VIEW generated shoeboxes

hkl = (12, 11, 38)

hkl = (17, -7, -16)

hkl = (14, -2, -27)

hkl = (21, 15, -17)
EVAL integration by SVD

\[ \chi^2 = \sum_{i=1}^{N} w_i \left[ \rho_i - JP_i - \sum_{m} M J_m P_{im} - ax_i - by_i - c \right]^2 \]

\[ I = J \sum_{i=1}^{N} P_i \]

\[ \sigma_i = \sqrt{\rho_i} \]

\[ w_i = \frac{1}{\sigma_i^2} \]

\( \rho_i \) observed pixel intensities

\( P_i \) model pixel intensities

background neighbours

J\( P_i \) + \( ax_i \) + \( by_i \) + \( c \) + \( J_{nb1} P_i_{nb1} \) + …

Poisson statistics
Corrected for Lorentz and polarization

χ
Data reduction

\[ I \propto L(\eta)p(\eta)|F_{hkl}|^2 \]

- Lorentz factor accounts for the relative time that a reciprocal lattice point spends on the Ewald sphere during the data collection
- Polarization factors accounts for incident beam polarization and polarization caused by reflection

After correction we get the squared structure factors:

\[ |F_{hkl}|^2 = I/Lp \]
Data evaluation, integration and analysis

- mosaic -> 0.5
- $\sigma(\lambda)$ -> 0.002
- divergence -> 1.3 mrad
- Xtal -> *2
Data evaluation, integration and analysis I

Detector panel edge
Background

- crystal droplet (solvent)
- fibre (nylon) loop
- air scattering

From the image header:
- Image pedestal=40 (base-line offset)
- avoids having to store negative numbers in 16 bits

Dark current has been subtracted

Data evaluation, integration and analysis
Background and standard deviations of reflections

\[ I_{\text{net}} = I_{\text{bruto}} - BG \]

\[ I_{\text{net}} = J \sum_i P_i = \sum_i \rho_i - \sum_i (ax_i + by_i + c) \]

\[ \sigma_i^2 = \sigma_{\text{peak}}^2 + \sigma_{BG}^2 \]

According to Poisson statistics, the higher the pixel intensities the higher the background noise.

Try to avoid background as much as you can.
X-ray detectors

- Film and image plates
- Multiware array detectors
- CCD detectors
- Hybrid pixel detectors

Almost no point spread
No electronic noise and dark current
No read-out noise

Data evaluation, integration and analysis
Pixels on the CCD detectors are distorted, largely due to the fibre optic taper. The equipment's software will have a distortion table that maps every pixel in the image file to the true position on the detector. If you want to take the images away from the equipment, you’d better unwarp the images.
Flood field

Non-uniformity correction image of CCD detector

In the centre larger phosphor thickness
Calibrated with homogeneous scatterer

Platinum 135 CCD
\[ \chi^2 = \frac{\sum (I - \langle I \rangle)^2}{\sum \sigma^2} \approx 1 \]  

\[ \sigma_I^2 \] from Poisson statistics
ANY: analysis

<table>
<thead>
<tr>
<th>Sh</th>
<th>Theta</th>
<th>Reso</th>
<th>N</th>
<th>&lt;I&gt;</th>
<th>&lt;s&gt;</th>
<th>&lt;I&gt;/s</th>
<th>Nmerge</th>
<th>&lt;I&gt;</th>
<th>&lt;s&gt;</th>
<th>&lt;I&gt;/s</th>
<th>cc1/2</th>
<th>cc*</th>
<th>npair</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.46</td>
<td>3.447</td>
<td>47634</td>
<td>601.48</td>
<td>2.28</td>
<td>206.72</td>
<td>3691</td>
<td>641.34</td>
<td>12.65</td>
<td>55.49</td>
<td>0.998</td>
<td>0.999</td>
<td>3666</td>
</tr>
<tr>
<td>2</td>
<td>15.78</td>
<td>2.736</td>
<td>47109</td>
<td>390.19</td>
<td>2.38</td>
<td>127.40</td>
<td>3489</td>
<td>389.65</td>
<td>9.90</td>
<td>38.08</td>
<td>0.997</td>
<td>0.999</td>
<td>3487</td>
</tr>
<tr>
<td>3</td>
<td>18.14</td>
<td>2.390</td>
<td>47370</td>
<td>197.47</td>
<td>2.04</td>
<td>76.98</td>
<td>3440</td>
<td>198.03</td>
<td>6.54</td>
<td>29.17</td>
<td>0.994</td>
<td>0.999</td>
<td>3448</td>
</tr>
<tr>
<td>4</td>
<td>20.04</td>
<td>2.172</td>
<td>47243</td>
<td>165.13</td>
<td>2.10</td>
<td>63.20</td>
<td>3430</td>
<td>165.93</td>
<td>5.04</td>
<td>31.67</td>
<td>0.995</td>
<td>0.999</td>
<td>3430</td>
</tr>
<tr>
<td>5</td>
<td>21.66</td>
<td>2.016</td>
<td>46460</td>
<td>129.79</td>
<td>2.09</td>
<td>49.81</td>
<td>3394</td>
<td>129.67</td>
<td>4.11</td>
<td>30.06</td>
<td>0.995</td>
<td>0.999</td>
<td>3394</td>
</tr>
<tr>
<td>6</td>
<td>23.09</td>
<td>1.897</td>
<td>46061</td>
<td>94.66</td>
<td>2.04</td>
<td>37.99</td>
<td>3379</td>
<td>95.05</td>
<td>3.34</td>
<td>27.06</td>
<td>0.994</td>
<td>0.998</td>
<td>3379</td>
</tr>
<tr>
<td>7</td>
<td>24.38</td>
<td>1.802</td>
<td>45591</td>
<td>61.65</td>
<td>1.96</td>
<td>26.23</td>
<td>3360</td>
<td>62.53</td>
<td>2.46</td>
<td>23.60</td>
<td>0.993</td>
<td>0.998</td>
<td>3360</td>
</tr>
<tr>
<td>8</td>
<td>25.57</td>
<td>1.724</td>
<td>45719</td>
<td>45.77</td>
<td>2.00</td>
<td>19.62</td>
<td>3394</td>
<td>46.59</td>
<td>2.08</td>
<td>20.62</td>
<td>0.990</td>
<td>0.998</td>
<td>3394</td>
</tr>
<tr>
<td>9</td>
<td>26.68</td>
<td>1.657</td>
<td>44609</td>
<td>34.83</td>
<td>2.09</td>
<td>14.52</td>
<td>3336</td>
<td>35.53</td>
<td>1.78</td>
<td>17.82</td>
<td>0.990</td>
<td>0.997</td>
<td>3336</td>
</tr>
<tr>
<td>10</td>
<td>27.71</td>
<td>1.600</td>
<td>33563</td>
<td>27.44</td>
<td>2.24</td>
<td>10.91</td>
<td>3349</td>
<td>27.97</td>
<td>2.09</td>
<td>12.82</td>
<td>0.975</td>
<td>0.994</td>
<td>3349</td>
</tr>
</tbody>
</table>

Next: scaling and error model for esd

Merged I/σ
SADABS: scaling

Bruker propriety software
Detector systematic errors

Data evaluation, integration and analysis I
True standard deviations can be described by second order polynomial in $I$ *

$$\sigma^2 = bgnoise^2 + \sigma_p^2 + (gI)^2$$

**Random noise**
- Read-out noise
- Dark current
- Incoherent scattering

**Systematic**

From Poisson statistics

$$= I$$

**Instrument errors:**
- Error in pixel size
- Profile error
- Counting error
- Inhomogeneity phosphor
- Radiation damage

**SADABS:**

$$\sigma^2 = K[\sigma_I^2 + (g\langle I\rangle)^2]$$

*Popov & Bourenkov  Acta Cryst D59 (2003) 1145*
Statistics after scaling

\[
\sigma^2 = K[\sigma_i^2 + (g\langle I \rangle)^2]
\]

Fitted in SADABS to minimize \(\chi^2\)