## 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

→ ひ ⋒ ① www.crystal.chem.uu.nl/distr/eval/

#### 

#### **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 <u>J. Appl. Cryst. 43</u>, (2010) 70-82

A reprint (Copyright © International Union of Crystallography J. Appl. Cryst. 43, (2010) 70-82) can be found here (PDF file, 1.1 Mb).

The eval14 method is described in:

A.J.M. Duisenberg, L.M.J. Kroon-Batenburg and A.M.M. Schreurs An intensity evaluation method: EVAL-14 *J. Appl. Cryst.* **36**, (2003) 220-229.

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

runs on Linux

## **EVAL flow diagram**





\*often  $O^T$  is used as orthogonalization matrix

Data evaluation, integration and analysis I

Х

## **Crystal and reciprocal lattices**



Data evaluation, integration and analysis I

## Thaumatin



ADSC detector Diamond Light Source Scan around horizontal spindle axis 0.5° increments





9

detector

## Ewald sphere



#### **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.

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

d\*

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

## Peakref: Refining peak positions



### Metadata: image header

ł 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;

ASCII image header of ADSC detector at Diamond



Center of incident beam in mm on detector

px= -(157.9/0.1026 -3072)=1533 py= -(157.48/0.1026-3072)=1537

## **VIEW:** datcol





Frame 200

## **VIEW:** datcol

Generation of shoeboxes for integration with EVAL15



## VIEW generated shoeboxes



### **EVAL** integration by SVD



18



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:

$$|F_{hkl}|^2 = I/Lp$$





#### Detector panel edge



overflow

## Background



Data evaluation, integration and analysis I

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

#### Dark current has been subtracted

	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	
294	6	8	9	5	3	6	8	4	4	2	5	
293	7	7	6	4	5	0	4	0	2	6	6	
292	5	1	0	1	3	1	3	2	6	7	5	
291	7	4	5	-2	4	-1	5	7	10	6	3	
290	5	4	4	1	11	6	10	8	1	2	2	
289	1	4	6	4	9	0	3	4	8	6	4	
288	4	5	5	14	7	3	6	3	6	3	2	
287	8	4	8	9	9	6	9	2	3	5	4	
286	4	6	6	7	3	3	5	2	3	7	6	
285	4	3	7	5	3	2	4	3	5	7	5	
284	5	6	7	5	-1	6	5	2	8	7	6	
ange	e from	n 1 to	0 14 1	Most :	freque	ent:	value	=6 tin	mes=1	9		
ixel	ls: 1	14 Sur	m: 57	7 Mean	n: 5.0	06 Sig	gma:	2.38	(Sigma	a^2)/I	Mean:	1.12

pixels: 114 aBG -0.07 bBG -0.02 cBG 5.56 sigBG 2.39

## Background and standard deviations of reflections

$$I_{net} = I_{bruto} - BG$$



$$I_{net} = J \sum_{i} P_i = \sum_{i} \rho_i - \sum_{i} (ax_i + by_i + c)$$

$$\sigma_I^2 = \sigma_{peak}^2 + \sigma_{BG}^2$$

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 ٠ Multiware array detectors ٠ phosphor amplifier ADC CCD CCD detectors ٠ photons xph<sup>-</sup> light<sup>-</sup> image e  $\sigma(I)$  $\sigma(I)$  $\sigma(I)$ ADU's Hybrid pixel detectors ٠ electrons  $\sigma_{dark-current}(t)$  $\sigma_{\text{read-out}}(n)$ Peltier Phosphor Liquid cooler X-ray Beryllium window CCD n++ n+ Taper **Readout ASIC** Vacuum Silicon sensor pixel Almost no point spread : . No electronic noise and dark current No read-out noise PC Power unit Controller

Data evaluation, integration and analysis I

## Unwarping images

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

860

1080

#### **ANY:** analysis

pg 4/mmm pointgroup set to 4/mmm nPg=16 (h,k,l) (k,-h,l) (-h,-k,l) (-k,h,l) (h,-k,-l) (-k,-h,-l) (-h,k,-l) (k,h,-l) (-h, -k, -1) (-k, h, -1) (h, k, -1) (k, -h, -1) (-h, k, 1) (k, h, 1) (h, -k, 1) (-k, -h, 1)RebuildDB ok centric=on Completeness and Rmerge for Shells Forbid: EDGEVER EDGEHOR EDGEROT BADUNIF OVERFLOW MAXSHIFT Allow: GOOD WEAK NEGATIVE Require: NONE theta from 0.0 to 27.709 Sh Theta Reso Meas Equi Obs Mis Lost Total Perc Cum Unil Uni2+ Nrsym Redun Rsym Rmeas Rpim Chil 1 12.46 3.447 43845 6609 50454 18 522 50994 98.9 98.9 25 3666 47609 12.99 0.066 0.069 0.019 637.25 2 15.78 2.736 41851 8787 50638 0 50638 100.0 99.5 2 3487 47107 13.51 0.087 0.090 0.024 313.74 0 3 18.14 2.390 2 50832 100.0 99.6 0 3448 47370 13.74 0.113 0.117 0.031 189.00 41134 9696 50830 0 4 20.04 2.172 40382 10650 51032 0 51032 100.0 99.7 0 3430 47237 13.77 0.106 0.110 0.029 188.12 0 5 21.66 2.016 39182 11526 50708 0 0 50708 100.0 99.8 0 3394 46466 13.69 0.109 0.114 0.031 72.70 0 3379 46058 13.63 0.121 0.126 0.034 6 23.09 1.897 38388 12382 50770 0 0 50770 100.0 99.8 50.32 13007 0 3360 45594 13.57 0.136 0.141 0.038 30.55 7 24.38 1.802 37693 50700 0 0 50700 100.0 99.8 8 25.57 1.724 37410 13852 51262 0 0 51262 100.0 99.9 0 3394 45715 13.47 0.154 0.160 0.044 25.30 9 26.68 1.657 44613 13.37 0.174 0.181 0.049 36203 14329 50532 0 0 50532 100.0 99.9 0 3336 22.72 10 27.71 1.600 26893 23925 50818 0 0 50818 100.0 99.9 0 3349 33563 10.02 0.200 0.211 0.066 17.08 \_\_\_\_\_ \_\_\_\_\_ \_\_\_\_ 27.71 1.600 382981 124763 507744 18 524 508286 99.9 99.9 27 34243 451332 13.18 0.094 0.098 0.027 161.28 Resolution 45.663-1.6 (1.657-1.6) Rsym 0.094 (0.2) Rmeas 0.098 (0.211) Rpim 0.027 (0.066) Reflections 451359 (33563) Unique 34270 (3349) Completeness 99.893 (100.0) Redundancy 13.18 (10.022)  $\chi^2 = \frac{\sum (I - \langle I \rangle)^2}{\sum \sigma^2} \approx 1$ 



22

## **ANY:** analysis

<u> </u>					0.5.5.5										
Int	Intensity distribution for Shells, unmerged and merged														
For	Forbid: EDGEVER EDGEHOR EDGEROT BADUNIF OVERFLOW MAXSHIFT														
All	Low: Go	DOD WEA	AK NEGA	TIVE						$\frown$	$\frown$				
Red	Require: NONE														
Sh	Theta	Reso	N	<1>	<s></s>	<i s=""></i>	Nmerge	<1>	<s></s>	KI/s>	¢c1/2	cc*	npair		
1	12.46	3.447	47634	601.48	2.28	206.72	3691	641.34	12.65	55.49	0.998	0.999	3666		
2	15.78	2.736	47109	390.19	2.38	127.40	3489	389.65	9.90	38.08	0.997	0.999	3487		
3	18.14	2.390	47370	197.47	2.04	76.98	3448	198.03	6.54	29.17	0.994	0.999	3448		
4	20.04	2.172	47243	165.13	2.10	63.20	3430	165.93	5.04	31.67	0.995	0.999	3430		
5	21.66	2.016	46460	129.79	2.09	49.81	3394	129.67	4.11	30.06	0.995	0.999	3394		
6	23.09	1.897	46061	94.66	2.04	37.99	3379	95.05	3.34	27.06	0.994	0.998	3379		
7	24.38	1.802	45591	61.65	1.96	26.23	3360	62.53	2.46	23.60	0.993	0.998	3360		
8	25.57	1.724	45719	45.77	2.00	19.62	3394	46.59	2.08	20.62	0.990	0.998	3394		
9	26.68	1.657	44609	34.83	2.09	14.52	3336	35.53	1.78	17.82	0.990	0.997	3336		
10	27.71	1.600	33563	27.44	2.24	10.91	3349	27.97	2.09	12.82	0.975	0.994	3349		
	27.71	1.600	451359	181.55	2.12	65.69	34270	184.43	5.10	28.95	0.998	0.999	34243		

Next: scaling and error model for esd

Merged I/ $\sigma$ 

## SADABS: scaling



#### **Detector systematic errors**



## Error model

True standard deviations can be described by second order polynomial in I \*



\*Popov & Bourenkov Acta Cryst D59 (2003) 1145

## Statistics after scaling

Completeness and Rmerge for Shells																				
Forbid: EDGEVER EDGEHOR EDGEROT BADUNIF OVERFLOW MAXSHIFT SADABS																				
Allow: GOOD WEAK NEGATIVE																				
Require: NONE																				
theta	fro	m 0.0	to 27.7	09												$\langle \rangle$			$\langle \rangle$	
Sh The	ta	Reso	Meas	Equi	Obs	Mis 1	Lost T	otal	Per	c Cu	ım Uni	1 Uni	2+ Nrs	ym Re	dun	Rsyn	Rmeas	Rpim	/Chi2	
1 12.	46 3	3.447	42278	8176	50454	18	522 5	0994	98.	9 98.	.9 2	5 36	66 455	35 12	.42	0.044	0.046	0.013	0.88	
2 15.	78 2	2.736	39330	11308	50638	0	05	0638	100.	0 99.	. 5	2 348	37 436	08 12	.51	0.051	0.053	0.014	1.10	
3 18.	14 2	2.390	37346	13484	50830	0	25	0832	100.	0 99.	6	1 344	47 422	51 12	.26	0.059	0.061	0.017	1.23	
4 20.	04	2.172	39188	11828	51016	0	16 5	1032	100.	0 99.	.7	1 342	28 453	23 13	.22	0.052	0.054	0.015	0.98	
5 21.	66 2	2.016	38348	12360	50708	0	05	0708	100.	0 99.	. 8	0 339	94 450	50 13	.27	0.051	0.053	0.014	0.89	
6 23.	09 :	1.897	37636	13134	50770	0	05	0770	100.	0 99.	. 8	0 33	79 447	44 13	.24	0.056	0.058	0.016	0.88	
7 24.	38 :	1.802	37061	13639	50700	0	05	0700	100.	0 99.	. 8	0 33	60 445	31 13	.25	0.065	0.067	0.019	0.86	
8 25.	57 :	1.724	36689	14573	51262	0	05	1262	100.	0 99.	. 9	0 339	94 445	25 13	.12	0.077	0.080	0.022	0.84	
9 26.	68 :	1.657	35308	15224	50532	0	05	0532	100.	0 99.	. 9	0 333	36 432	04 12	.95	0.093	0.097	0.02	0.83	
10 27.	71 :	1.600	25987	24815	50802	0	16 5	0818	100.	0 99.	. 9	5 334	43 320	99 9	.60	0.115	0.121	0.039	0.84	
																+/				
27.	71 :	1.600	369171	138541	507712	18	556 50	8286	99.	9 99.	.9 3	4 3423	34 4308	70 12	.59	Q.052	0.055	0.015	0.98	
		_		_					_		$\langle \rangle$					$\bigcirc$			$\bigcirc$	
Sh The	eta	Reso	N	<i></i>	<s></s>	<i s=""></i>	Nmerge	3	<i></i>	<s></s>	KI/s	cc1/2	2 CC*	npair	r	C		2		27
1 12.	.46	3.447	45560	184.07	12.56	14.16	369:	1 191	.78 4	4.31	47.76	0.999	1.000	366	6	$\sigma^2 =$	$= K   \sigma$	f + (	g(I)	2
2 15.	.78	2.736	43610	123.88	8.54	13.60	3489	9 119	.40 2	2.41	45.99	0.999	9 1.000	348	/				0 ( //	-
3 18.	.14	2.390	42252	63.44	4.49	12.65	3448	3 62	.20 1	L.34	42.10	0.998	3 1.000	344	/					
4 20.	.04	2.172	45330	52.73	3.81	12.23	3429	9 52	.24 1	L.07	43.36	0.999	9 1.000	342	3	Fitted	in SAD	ABS t	o mini	mize $\gamma^2$
5 21.	.66	2.016	45044	41.57	3.11	11.37	3394	4 41	.19 (	0.87	40.60	0.999	9 1.000	3394	4				-	
6 23.	.09	1.897	44747	30.25	2.40	10.44	3379	9 30	.25 (	0.68	37.36	0.999	9 1.000	337	9					
7 24.	.38	1.802	44528	19.67	1.78	8.91	. 3360	) 19	.94 (	0.51	32.16	0.999	1.000	336	)					
8 25.	.57	1.724	44529	14.57	1.53	7.68	3394	4 14	.85 (	0.44	27.66	0.998	0.999	3394	4					
9 26.	. 68	1.657	43200	11.07	1.42	6.35	333	6 11	.32 (	0.41	22.78	0.997	0.999	333	6					
10 27.	.71	1.600	32104	8.83	1.41	5.19	334	3 9	0.02 (	5.51	15.55	0.994	0.998	3343	3					
		1 600	420004		4 00	10 41	2406			=====			1 000	2402	-					
27.	. / ⊥	T.000	430904	50.50	4.20	10.41	. 34260	5 56		1.29	41.69	0.999	, T.000	3423	+					
											$\bigcirc$									