Recent Developments in xia2

Strubi early 2012 update

Graeme Winter

Diamond Light Source

1 March 2012

Overview

- CCP4 2012 presentation
- Removal of frame count limitations
- Current state of multi-crystal analysis
- Other developments
- Conclusions & plans

Removal of frame count limitations

- Use Aimless in place of Scala
- Use Pointless in place of Reindex[1]
- Run -3daii or -3da
- No more batch limitations

---

[1]This is still not as efficient as it could be - but it works

Example: 20 sweeps from insulin

- Insulin test sample from I04-1
- 20 sweeps: 0.2 degree / frame x 900: 18000 images
- Radiation damage very likely
- Substantial variation between data sets
- Processing rather time consuming, use a big workstation

```
High resolution limit           1.66    7.43    1.66
Low resolution limit           55.18   55.18    1.70
Completeness                   100.0   99.9  100.0
Multiplicity                   314.3  346.7  262.8
I/sigma                         81.7  228.5    7.6
Rmerge                         0.106  0.071  5.227
Rmeas(I)                       0.106  0.071  5.240
Rmeas(I+/-)                    0.106  0.071  5.247
Rpim(I)                        0.006  0.004  0.322
Rpim(I+/-)                     0.008  0.005  0.450
Wilson B factor                31.457
Total observations             2982419 44723 180512
Total unique                   9488 129 687
```

Effects of Running Aimless

- Output effectively the same as Scala
- SD correction slightly more effective
- C++ not Fortran - no need to recompile for lots of batches
- Parallel (openMP) version in development

diamond

Effects of Running Aimless in xia2

- Since scaling performed by XSCALE very little difference in results

Current state of multi-crystal analysis

- Developed over ~1 year or so
- Hiatus for most of that time
- Still very rough around the edges
- Requires use of XDS processing
- Requires R (at the moment) to generate plot

diamond

Basic principles

- Assume in processing that everything comes from the same crystal
- Try to assemble a data set from everything
- Worry about isomorphism in scaling rather than from unit cell etc.
- Look at CC's between sweeps as measure of isomorphism
- Scale CC's to ersatz distance as $d = \frac{1}{CC} - 1^2$

[2]Which means you can *interpret* the distances.

xia2 -3daii /here/are/my/data

# Time passes... let's talk about options

Useful options

- -microcrystal - switch to microcrystal mode
- -failover - ignore failure of processing individual sweeps
- -spacegroup - set the spacegroup as e.g. P23
- -cell $a, b, c, \alpha, \beta, \gamma$ - needs -spacegroup
- -min_images 3 - minimum images to count as a sweep
- -xparm GXPARM.XDS - refined model of experimental geometry

diamond

After processing

```
> cd DEFAULT/scale
> x1335 XSCALE.LP
> R --no-save < x1335.R
```

Results

Cluster Dendrogram

d
hclust (*, "ward")
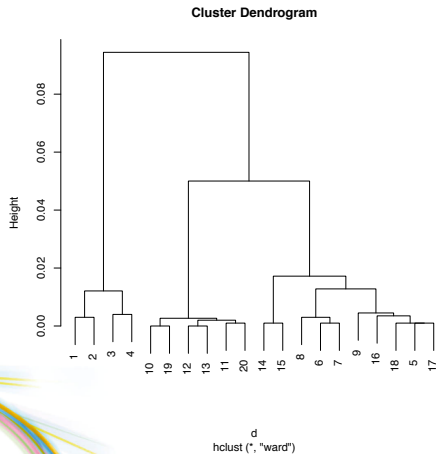
```
No.  Av. CC  I/sig  File Name
  1  0.9796   2.25  NATIVE_SWEEP11.HKL
  2  0.9822   3.61  NATIVE_SWEEP22.HKL
  3  0.9807   7.49  NATIVE_SWEEP33.HKL
<snip>
 18  0.9925   2.64  NATIVE_SWEEP9.HKL
 19  0.9930   1.69  NATIVE_SWEEP10.HKL
 20  0.9826   3.39  NATIVE_SWEEP12.HKL
```
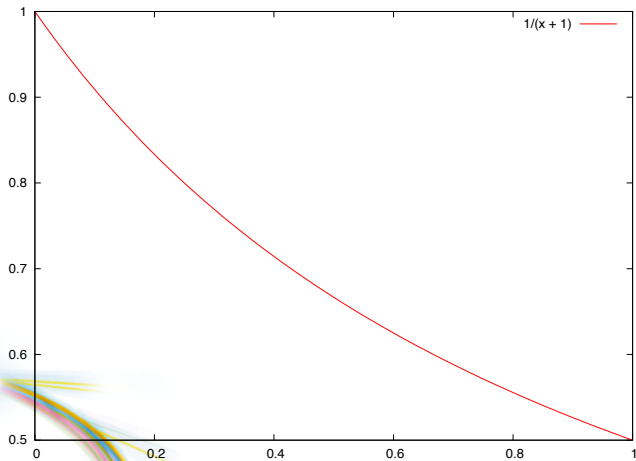
Interpretation

- $d = \frac{1}{CC} - 1$
- $CC = \frac{1}{d+1}$
- $CC = 0.95 \rightarrow d < 0.05$
- Up to you to decide on your limits... at least you know what they mean

Interpretation

Next steps

- Book keeping - take the sweeps you wish to merge from dendogram
- Edit xinfo file
- Run again
- This needs automating, and also calculations performing to tell you how complete results will be

Trivial example - Thaumatin

- Split 450 image Thaumatin data set into 5 x 90 image chunks
- Process as usual, very good results and grouping comes in data collection order
- Distances are fantastically small indicating very high CC's

Results

Cluster Dendrogram

d
hclust (*, "ward")
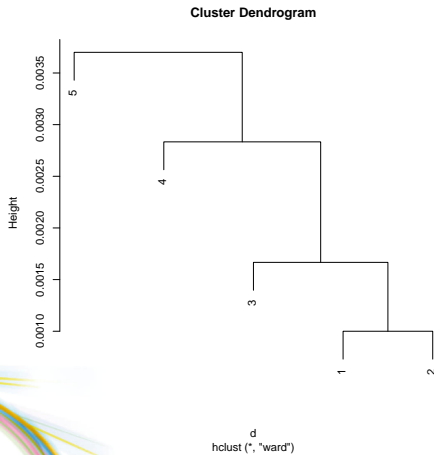
Developments planned from mid April

- Apply to more cases
- Work on clustering
- Include comparison of R_{merge} as a function of batch number
- Work on incorporating radiation damage: split sweeps into blocks
- Automate analysis, also allowing for completeness

Other developments

- Development of xia2 for small molecule crystallography
- Support for complex experimental geometry from imgCIF data
- Full scaling model search with scaling with Aimless and Scala