Reprocessing CFHT Deep fields with the (LSST DM) stack

Ph.Gris
Laboratoire de Physique de Clermont
IN2P3/CNRS
(with help/ideas from P.Astier, D.Boutigny, N.Chotard, D. Fouchez, J.P. Reyes)

- Goals
- Attempts and achievements
- Next steps

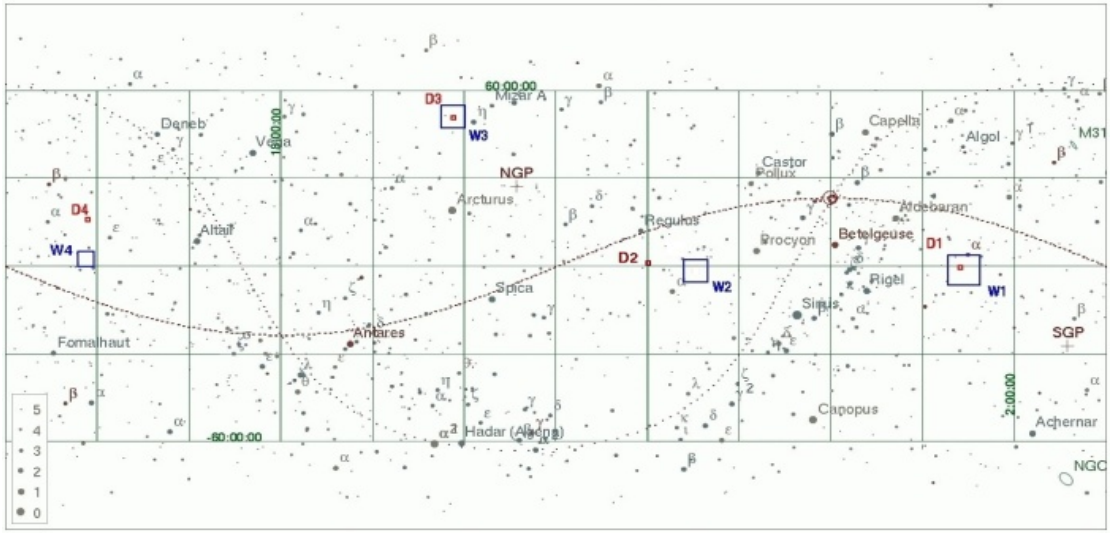
Reprocessing CFHT Deep fields with the stack

- **Goals:**
  - **technical:**
    - **practise/exercise the production chain (easiness, relialibility, reproducibility, …)**
    - **identify potential problems (if any/in case of bad luck)**
    - **effect of data size**
  - **science requirements:**
    - **quality of the production (performance in astrometry and photometry) for each step of the processing.**

- **Science case: extract the light curves of the Supernovae collected by CFHT**
  - **are we able to extract these light curves from the stack ?**
  - **Are they of sufficient quality to include the corresponding SN in the Hubble diagram ?**
  - **How does the quality compare with CFHT LCs ?**

CFHT D3 field

- One of the four deep fields of CFHT
- 1x1 deg2 – 36 ccds
- RA= 14:19:27 (214.8625 deg) DEC= +52:40:56 (52.6822 deg)
- 121 SN candidates detected



D3 reprocessing is the first step. Others will be processed too

Reprocessing: the steps

Level 1
Raw images
+corrections (bias, dark, flat)

Level 2

ProcessCcd

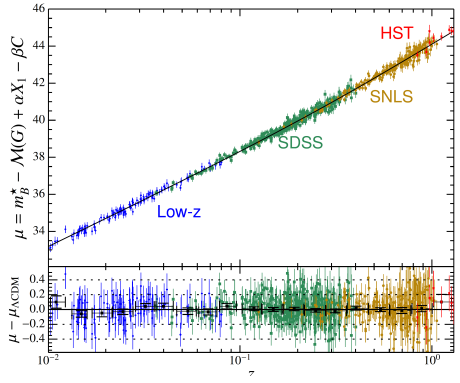
jointcal(simultaneous astrometry)

Coaddition

ForcedPhotometry → Image Subtraction

Level 3

SN light curves



Cosmology

Reprocessing: the steps

Level 1
Raw images
+corrections (bias, dark, flat)

Level 2

ProcessCcd

With an "old" release (~ july 2015)

jointcal(simultaneous astrometry)

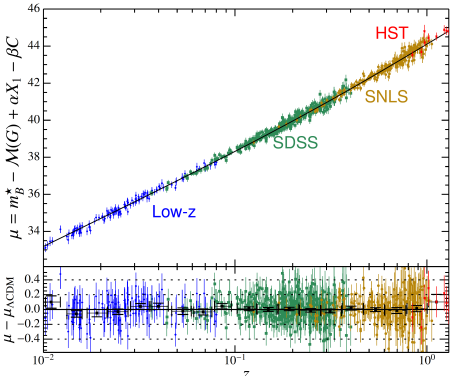
Coaddition

ForcedPhotometry

Image Subtraction

Level 3

SN light curves



Cosmology

Reprocessing: the steps

Level 1
Raw images
+corrections (bias, dark, flat)

Level 2

With weekly w20 (2017)

ProcessCcd

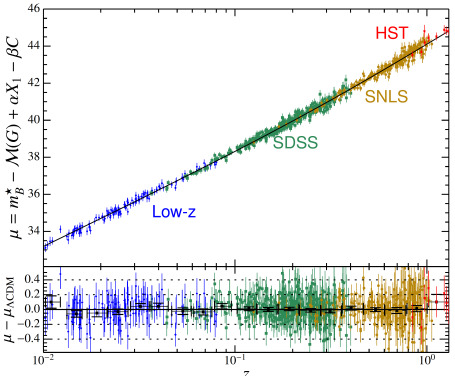
jointcal(simultaneous astrometry)

Coaddition

ForcedPhotometry → Image Subtraction

Level 3

SN light curves



Cosmology

D3 : calexps production status

- First step of L2: production of "Calibrated Exposure":
  - Calexp : A fully-qualified LSST image, which includes a science pixel array, and concomitant data including a quality mask and a variance array, in addition to a PSF characterization and metadata (including calibration metadata) about the image (https://confluence.lsstcorp.org/display/LSWUG/DMS+Glossary#DMSGlossary-C)
  - calibration
    - astrometry : reference catalog : SDSS ("old" release), PanSTARRS (w20)
    - photometry : reference catalog: SDSS
  - production of catalogs (sources) (stars and galaxies)

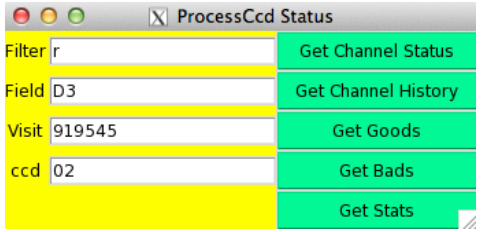
- Calexps to be produced
  - u : 153 x 36 = 5508
  - g : 542 x 36 = 19512
  - r : 645 x 36 = 23220
  - i : 713 x 36 = 25668
  - z : 682 x 36 = 25582

2736*36 = 98460 images

D3 : Calexps production status

"old" release

- filter u,g,r,i,z : > 99.6 % of the production completed
- We tried to get a clear view of the status of the processing:
  - python script accessible from ccage: python /sps/lsst/dev/lsstprod/CFHT/D3/utils/Production_Status/processCcd/processCcdgui.py &



Number of events processed: 5508 / 5508 Nimages= 153
Number of events ok : 5499 frac = 0.998
Number of events not ok : 9 frac = 0.002
Fatal errors : 9 Problem : 1 event(s)
No input matches : 2 event(s)
Unable to match sources : 4 event(s)
No sources remaining in match list after magnitude : 2 event(s)

Given the quality of input images, these numbers should, at some point, be equal to 100% (P.Astier)

  - from the wiki page of the DESC reprocessing TF: https://github.com/DarkEnergyScienceCollaboration/ReprocessingTaskForce/wiki/CFHT-D3-reprocessing-@ccin2p3

- Calexps: 8.5 Tb, catalogue: 761 Gb (fits)

D3 : Calexps production status

Weekly w20

- **Simplified log file analysis:**

```
object visit filter runId date ccd expTime isr Ndetected aperture_corr psf_sigma Ndeblended Nsources astrometry_scatter photometry_zp status
D3 788113 g 05AL01 2005-04-0 21 225.162 Yes 181 Yes 2.35 116 2144 "0.058 +- 0.029 arcse" 32.22821 Ok
D3 788113 g 05AL01 2005-04-0 22 225.162 Yes 215 Yes 2.31 143 1729 "0.066 +- 0.040 arcse" 32.253874 Ok
D3 788113 g 05AL01 2005-04-0 23 225.162 Yes 156 Yes 2.37 133 1833 "0.073 +- 0.045 arcse" 32.24846 Ok
D3 788113 g 05AL01 2005-04-0 24 225.162 Yes 153 Yes 2.38 114 1928 "0.056 +- 0.033 arcse" 32.235651 Ok
D3 788113 g 05AL01 2005-04-0 25 225.162 Yes 181 Yes 2.39 123 1628 "0.058 +- 0.025 arcse" 32.245081 Ok
D3 788113 g 05AL01 2005-04-0 26 225.162 Yes 169 Yes 2.44 129 1466 "0.077 +- 0.039 arcse" 32.240595 Ok
D3 788113 g 05AL01 2005-04-0 27 225.162 Yes 127 Yes 2.46 94 1977 "0.061 +- 0.035 arcse" 32.262429 Ok
D3 788113 g 05AL01 2005-04-0 28 225.162 Yes 171 Yes 2.4 136 1559 "0.058 +- 0.035 arcse" 32.261827 Ok
D3 788113 g 05AL01 2005-04-0 29 225.162 Yes 160 Yes 2.39 132 1952 "0.058 +- 0.031 arcse" 32.251591 Ok
D3 788113 g 05AL01 2005-04-0 30 225.162 Yes 152 Yes 2.39 130 1642 "0.061 +- 0.041 arcse" 32.243828 Ok
D3 788113 g 05AL01 2005-04-0 31 225.162 Yes 146 Yes 2.37 78 2010 "0.062 +- 0.030 arcse" 32.239847 Ok
D3 788113 g 05AL01 2005-04-0 32 225.162 Yes 173 Yes 2.34 127 1695 "0.050 +- 0.027 arcse" 32.232635 Ok
D3 788113 g 05AL01 2005-04-0 33 225.162 Yes 189 Yes 2.37 135 1801 "0.068 +- 0.038 arcse" 32.240328 Ok
D3 788113 g 05AL01 2005-04-0 34 225.162 Yes 174 Yes 2.45 119 1624 "0.063 +- 0.034 arcse" 32.23116 Ok
D3 788113 g 05AL01 2005-04-0 35 225.162 Yes 183 Yes 2.57 126 1634 "0.055 +- 0.035 arcse" 32.253916 Ok
```

- **Fraction of remaining files to be produced**

Band	Frac
g	0.15%
r	0.15%
i	0.13%
z	0.07%

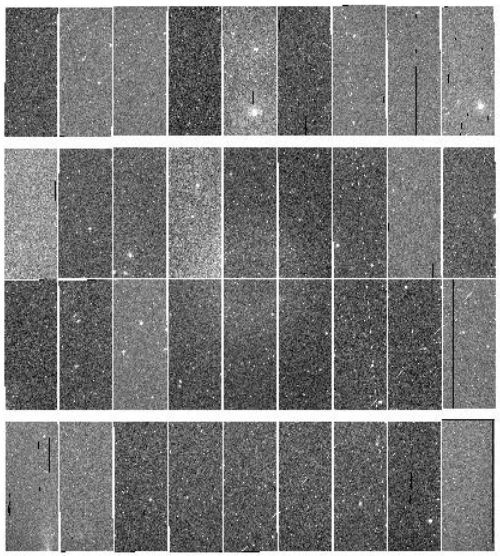
Need to be understood and solved (cf Pierre's comment) -> need to classify errors

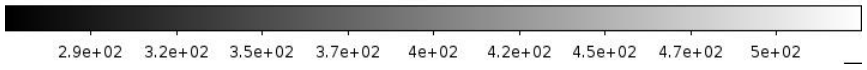
D3 : calexps production status

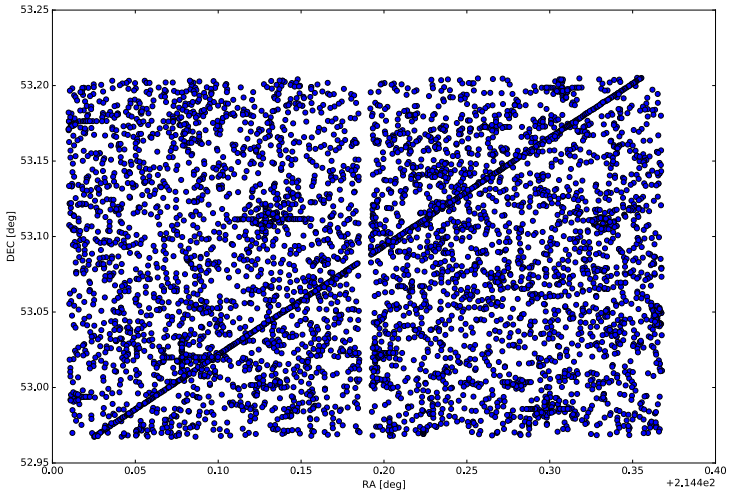
- Two major problems encountered:
  - attributed to afs problems:
    - on these workers: impossible to setup the release
    - systematic crash of the job submitted
    - this problem is transient, random and difficult to diagnose
  - Memory problems:
    - at the deblender level
    - identification: “bad::alloc” in error messages
    - appear in images crossed by KFO (P. Astier)

“old” release

May be partly solved by increasing job memory (up to 16 Gb !)







A FOI is being implemented in the stack

jointcal

- **simultaneous astrometry : fit astrometry positions (wcs, distorsions) using a large number of images -> see Pierre's presentation tomorrow**
- **primary wcs are given by the calexps**
- **from the calexps tagged as "good": apply jointcal.py (in the stack since w14 – 2017)**
- **bunches of 50 to 60 images to perform simultaneous astrometry**
- **Works fine, except for some cases where we get:**

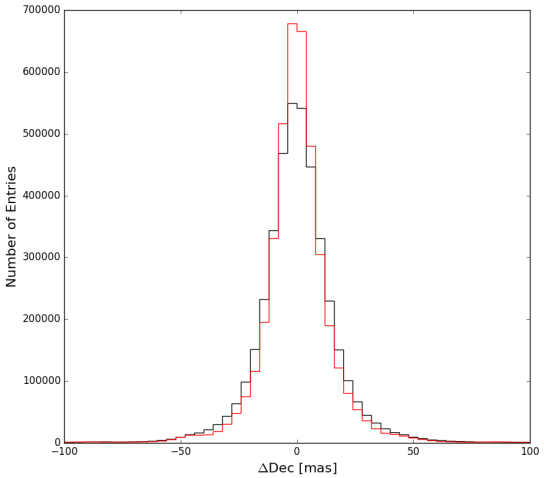
  jointcal INFO: Chi2/ndof : -nan/478730=-nan
  jointcal INFO: Chi2/ndof : nan/485930=nan
  jointcal INFO: Chi2/ndof : -nan/447770=-nan

  Weekly w20
- **This is not due to jointcal itself but to some sources that have bad (nan) values for centroids sigma and/or for shape_xx/yy**
- **Can be cured by modifying astrometrySourceSelector.py in meas_algorithms**
- **jointcal files (wcs) : 5.7 Gb**

Astrometry from processCcd+jointcal

- miltimatch (1") on src -> groups of sources
- selection:
  - median_mag < 21.5
  - at least 20 sources per group

Weekly w20



black : without jointcal
red : with jointcal

band	RMS(Ra*cos(Dec)) [mas]		RMS(Dec) [mas]		Nsources in [-50,50] mas
		jointcal		jointcal	
g	17.7	17.3	14.8	12.9	+53609 (0.2%)
r	14.9	14.2	14.0	12.4	+19686 (0.5%)
i	16.4	13.2	13.0	11.7	+15595 (0.5%)
z	15.6	14.1	15.0	13.3	+14770 (0.9%)

Science requirements

- **Science requirements necessary to achieving scientific measurements of specific accuracy have been defined in four main science areas of LSST:**
  - **Dark matter and dark energy constraints (WL, SN)**
  - **Inventory of the Solar system**
  - **transient optical sky**
  - **mapping the Milky Way**

- **For each requirement, three reference values:**
  - **minimum specification : minimun capability or accuracy required in order to achieve its scientific aims**
  - **design specification: system design point; basis for developping engineering tolerances**
  - **stretch specification: enhance scientific return if achieved; pursued if no increase in cost, schedule or risk**

- **https://docushare.lsstcorp.org/docushare/dsweb/Get/LPM-17**

Photometric quality (I)

- Repeatibility:
  - rms of the unresolved source magnitude distribution around the mean value
  - fraction of sources that deviate by more than x mmag from the mean

Quantity	design	minimum	stretch
PA1 [mmag]	5	8	3
PF1 (%)	10	20	15
PA2 [mmag]	15	15	10

- zeropoints:
  - rms of the internal photometric zero-point error
  - fraction of sources that deviate by more than x mmag from the mean

Quantity	design	minimum	stretch
PA3 [mmag]	10	15	5
PF2 (%)	10	20	5
PA4 [mmag]	15	20	15

Photometric quality (II)

- band-to-band(flux ratio) photometric calibration
  - accuracy of absolute band-to-band zero-point transformations (color zero-points) for main-sequence stars

Quantity	design	minimum	stretch
PA5 [mmag]	5	8	3
PA5 (with u)	10	15	5

- overall external absolute photometry
  - accuracy of photometric transformation to a physic scale

Quantity	design	minimum	stretch
PA6 [mmag]	10	20	5

Astrometric quality (I)

- relative astrometry:
  - rms of the astrometric distance distribution for stellar pairs with separation D arcmin will not exceed xx mas
  - fraction of the sample that deviate by more than xx mas from the median

	Quantity	design	minimum	stretch
D = 1 arcmin	AM1 [mas]	10	20	5
	AF1 (%)	10	20	5
	AD1 [mas]	20	40	10
D = 20 arcmin	AM2 [mas]	10	20	5
	AF2 (%)	10	20	5
	AD2[mas]	20	40	10
D = 200 arcmin	AM3 [mas]	15	30	10
	AF3(%)	10	20	5
	AD3 [mas]	30	50	20

Astrometric quality (II)

- relative astrometry:
  - match an image (other than r) to a r band image
  - rms of the distance between the positions on the two frames:
    - will not exceed x mas
    - fraction of measurements deviating by more than x mas from the mean

Quantity	design	minimum	stretch
AB1 [mas]	10	20	5
ABF1 (%)	10	20	5
AB2 [mas]	20	40	10

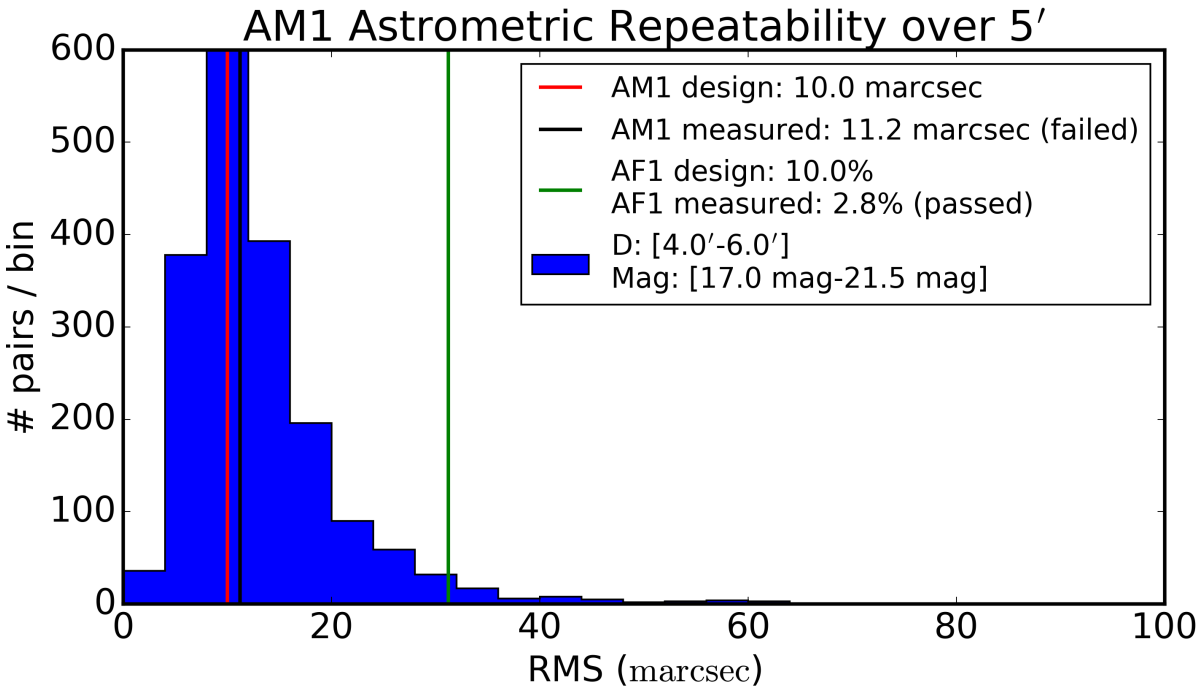
- Absolute astrometry:
  - transformation to an external system with a median accuracy

Quantity	design	minimum	stretch
AA1 [mas]	50	100	20

Science requirements

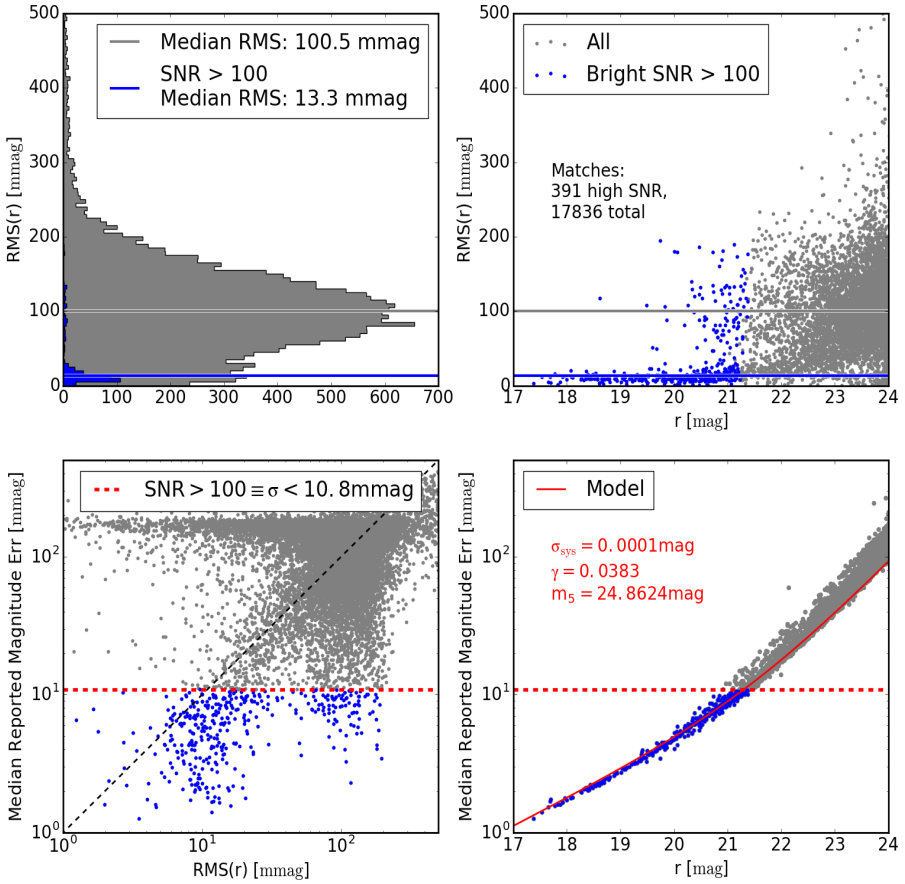
- **validate_drp : a python package that estimates some quantities related to science requirements (https://github.com/lsst/validate_drp.git):**
  - **photometry: PA1, PA2, PF1**
  - **astrometry: AM1->3, AF1->3, AD1->3**
- **On the calexps produced with w20:**

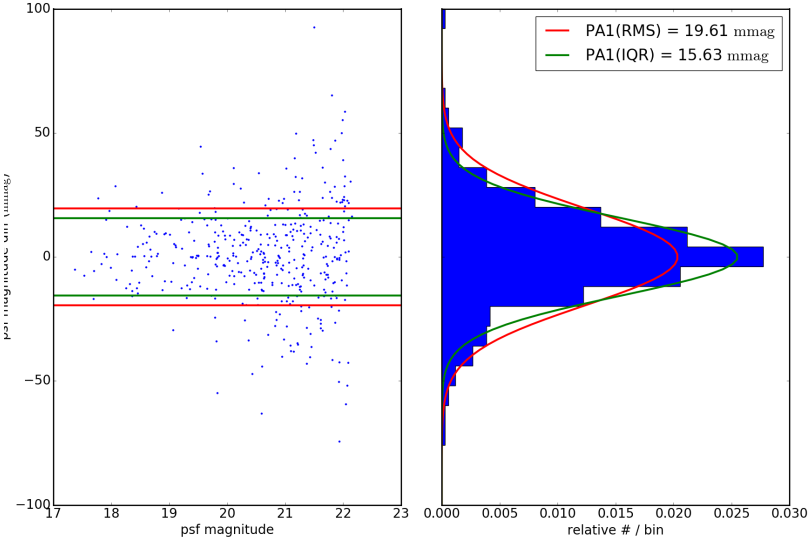
band r



validate_drp on calexps: photometry

Photometry Check : Output/r/0/sps_lsst_data_CFHT_D3_output_w_2017_20_r



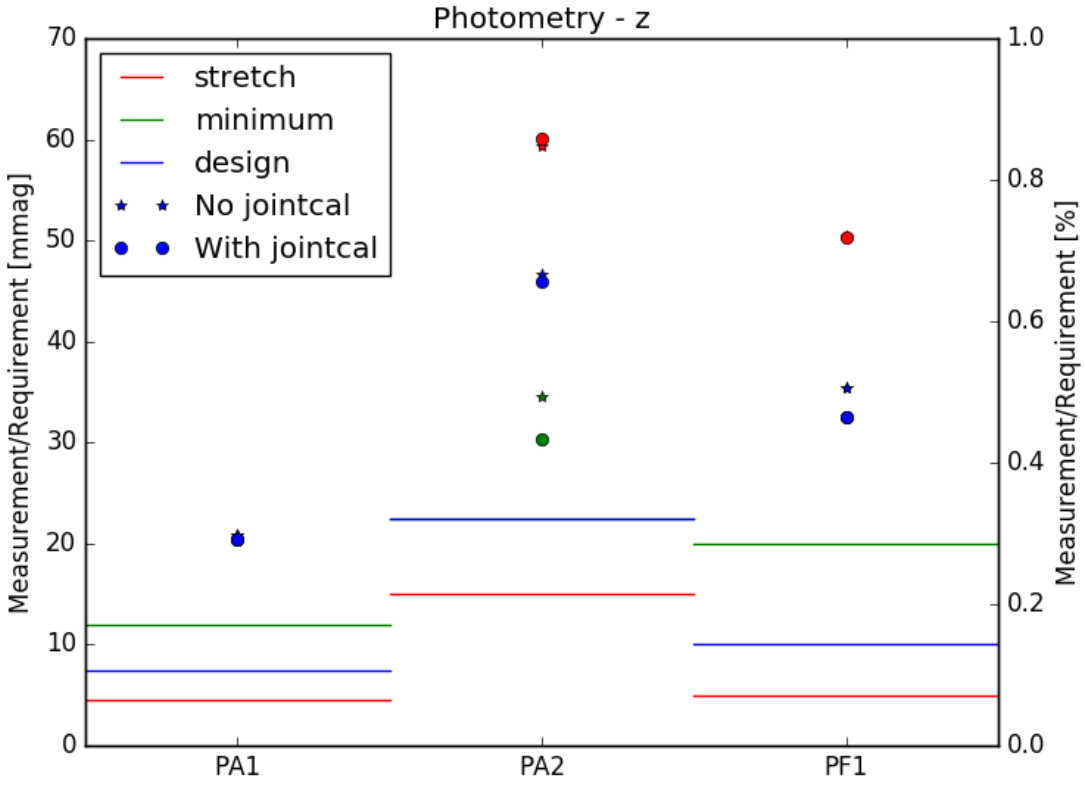


$$\sigma^2 = \sigma^2_{sys} + (0.04 - \gamma)x + \gamma x^2$$

$$x = 10^{0.4(m-m5)}$$

LSST overview paper
http://arxiv.org/abs/0805.2366v4

Science requirements: photometry



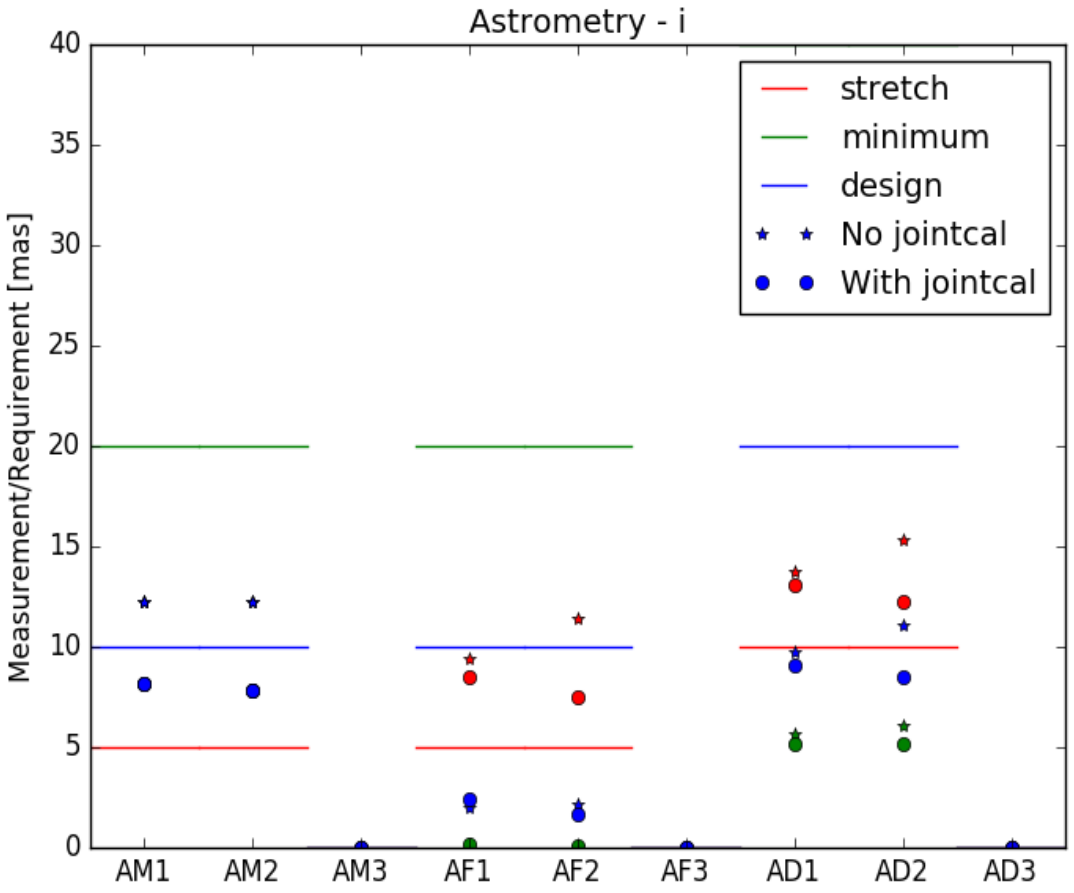
Same kind of results for r and i bands : minimum requirements not reached

Science requirements: photometry

		r		i		z	
		meas	ref	meas	ref	meas	ref
	stretch		3		3		4.5
PA1	minimum	15.3	8	14.6	8	20.8	12
[mmag]	design		5		5		7.5
	stretch	52.9	5	47.3	5	50.5	5
PF1	minimum	34.4	20	33.6	20	35.4	20
[%]	design	34.4	10	33.6	10	35.4	10
	stretch	42.4	10	41.1	10	59.3	15
PA2	minimum	22.7	15	20.9	15	34.5	22.5
[mmag]	design	32.6	15	32.2	15	46.7	22.5

Science requirements: astrometry

systematically better with jointcal



Science requirements: astrometry

black : without jointcal
red : with jointcal

Quantity	r			i			z		
	design	minimum	stretch	design	minimum	stretch	design	minimum	stretch
AM1	N/Y	Y/Y	N/N	N/Y	Y/Y	N/N	N/N	N/Y	N/N
AF1	Y/Y	Y/Y	N/Y	Y/Y	Y/Y	N/N	N/N	Y/Y	N/N
AD1	Y/Y	Y/Y	N/Y	Y/Y	Y/Y	N/N	N/N	Y/Y	N/N
AM2	N/Y	Y/Y	N/N	N/Y	Y/Y	N/N	N/N	Y/Y	N/N
AF2	Y/Y	Y/Y	N/Y	Y/Y	Y/Y	N/N	Y/Y	Y/Y	N/N
AD2	Y/Y	Y/Y	N/Y	Y/Y	Y/Y	N/N	Y/Y	Y/Y	N/N

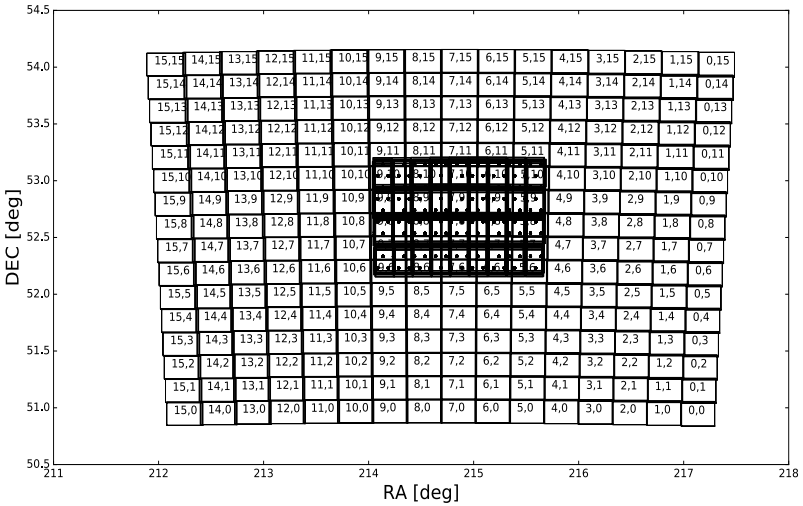
minimum requirement : ok (jointcal or not)
design requirement : ok r and y with jointcal, partly for z
stretch requirement : partly for r and i, with jointcal

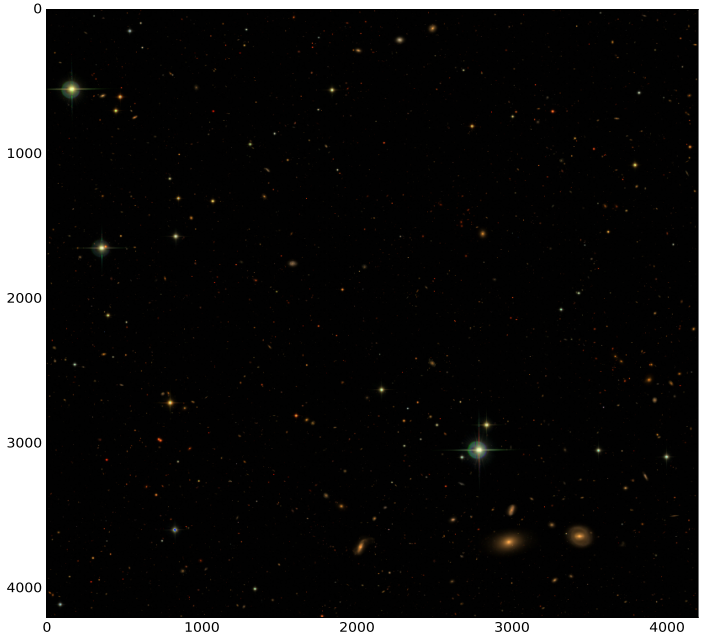
"old" release

Coaddition

- Coaddition steps:
  - CoaddTempExp -> ok
  - AssembleCoadd -> ok
  - MultiBandProcessing:
    - detectCoaddSources -> ok
    - mergeCoaddDetections -> ok
    - measureCoaddSources -> ok
    - mergeCoaddMeasurements -> ok

"ok" here just mean "went up to the end with no crash". It is not a "quality" statement.
At some point, this "quality" will have to be estimated (validation)



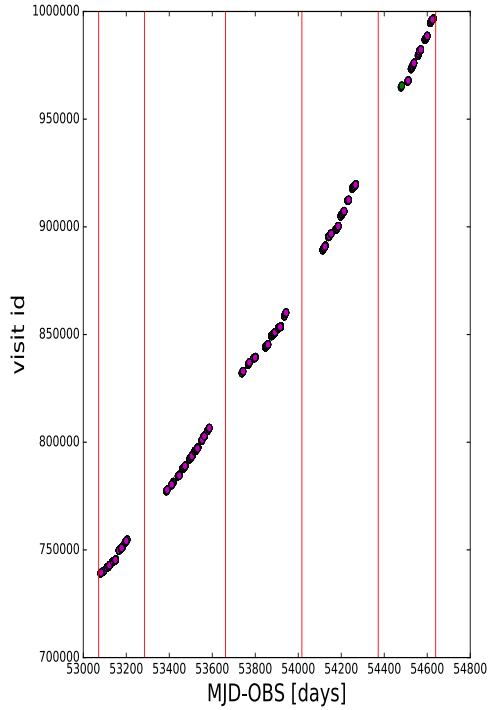


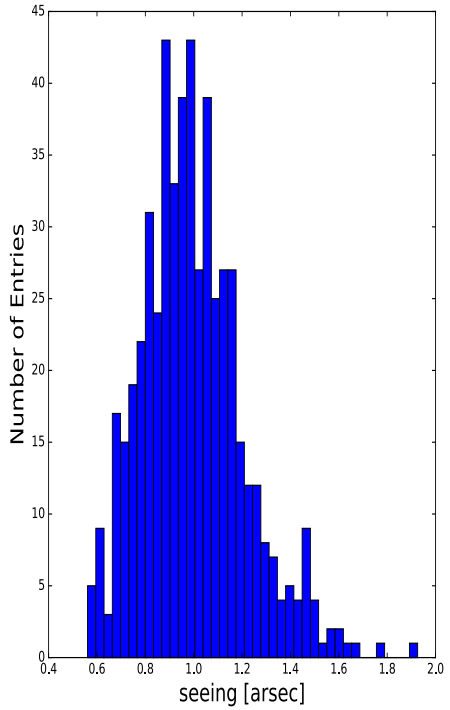
coadd g+r+i

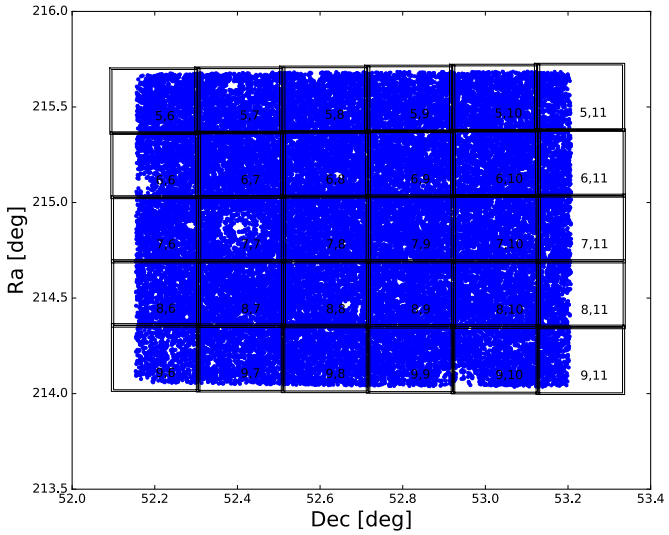
"old" release

Coaddition

- Coaddition step performed for one season:
  - season 2
  - calexps selected : lowest psf (30%)



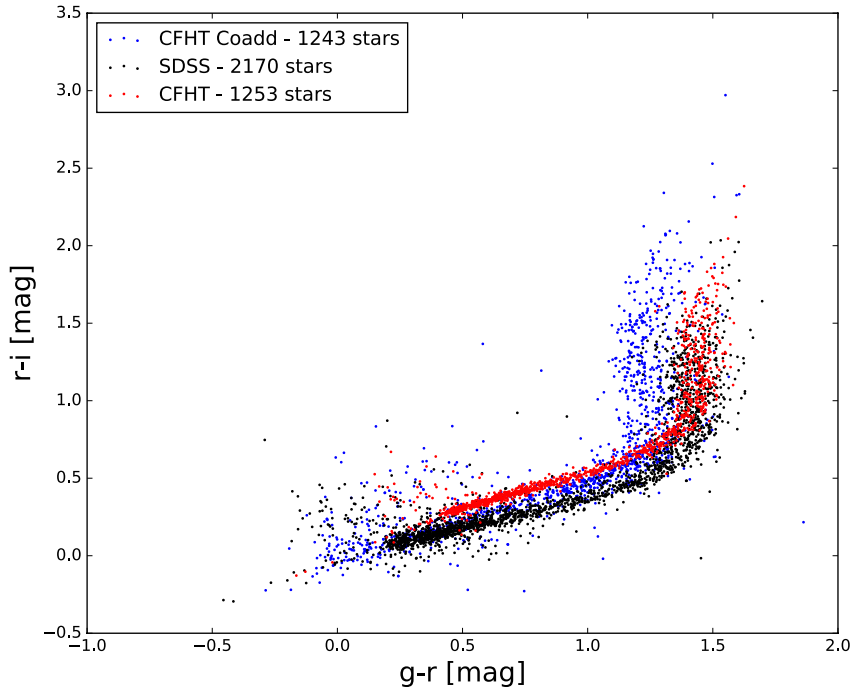


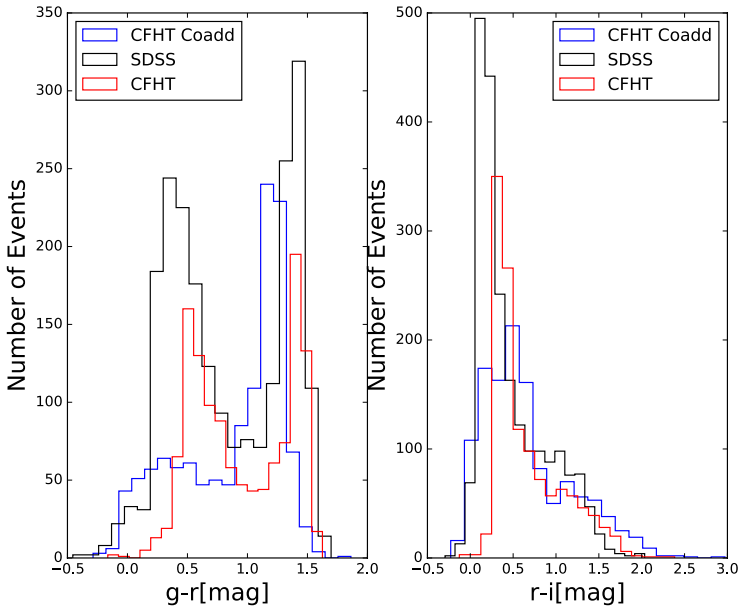


“old” release

Coaddition

- **Coaddition step validation : stellar locus plot**

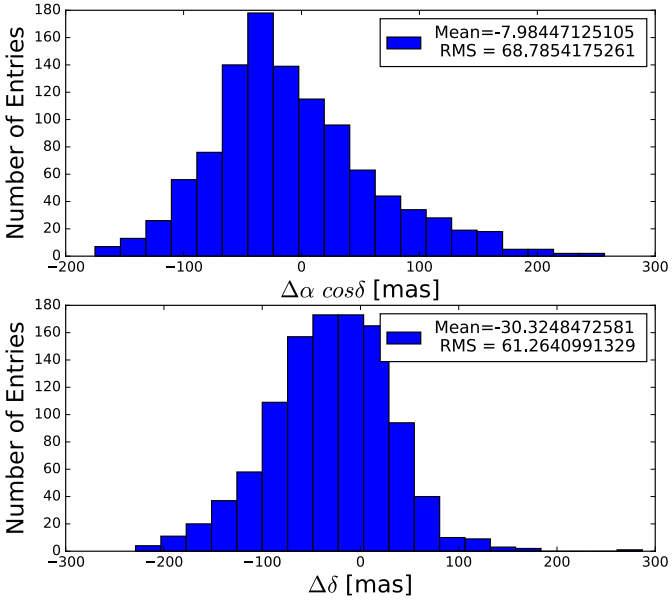


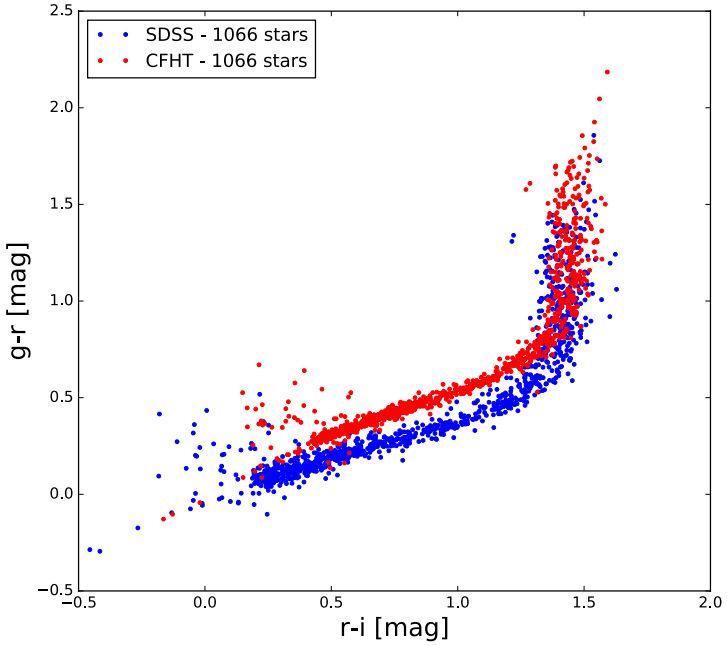


"old" release

Coaddition

- **Coaddition step validation : multimatch (afw,1") between catalogs:**
  - **SDSS and CFHT**

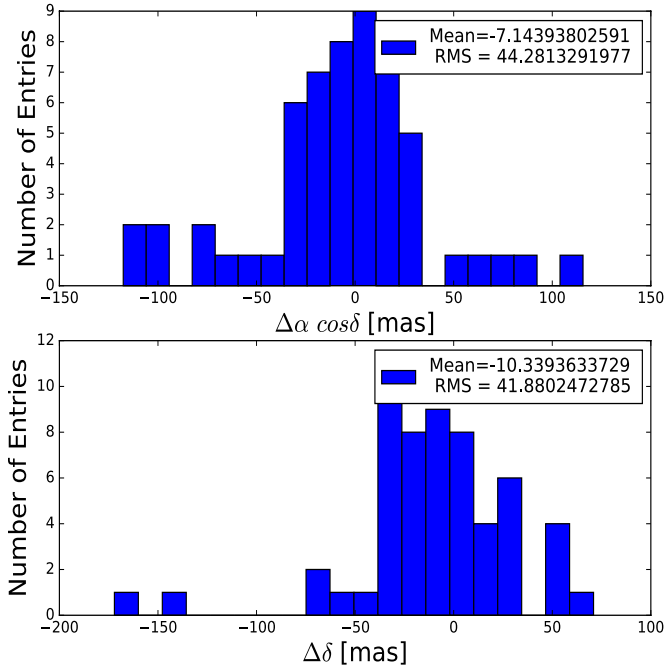


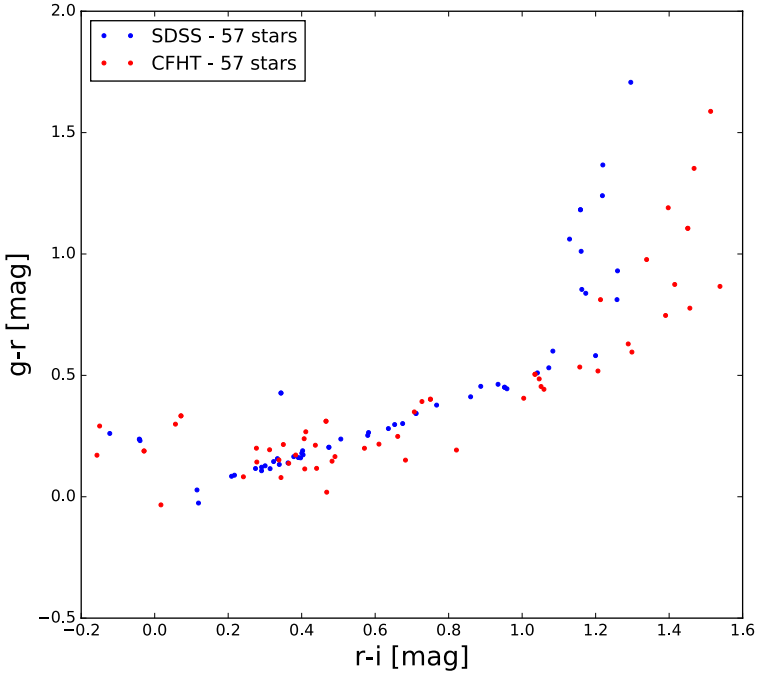


"old" release

Coaddition

- **Coaddition step validation : multimatch (afw, 1") between catalogs:**
  - **SDSS and Stack**





Only 57 sources matched !?!

"old" release

Forced photometry

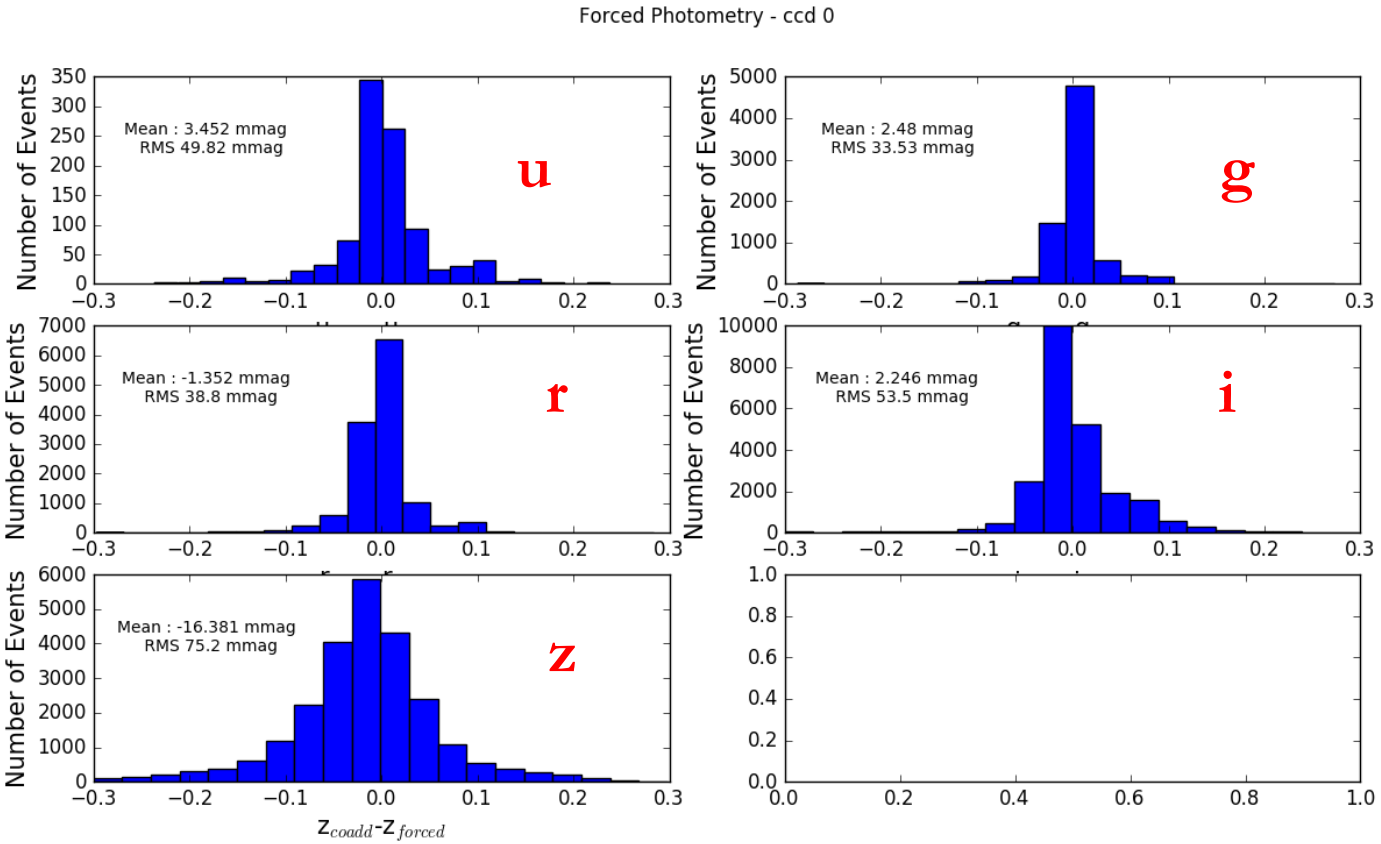
- Used the coadded (multiband) sources to perform a forced photometry
- Quite a lot of warnings/errors during processing:

```
forcedPhotCcd.measurement WARNING: Error in base_CircularApertureFlux.measure on record 5498900345556:
  File "src/image/Image.cc", line 92, in static lsst::afw::image::ImageBase<PixelT>::_view_t
lsst::afw::image::ImageBase<PixelT>
::_makeSubView(const Extent2I&, const Extent2I&, const _view_t&) [with PixelT = float;
lsst::afw::image::ImageBase<PixelT>::_vie
w_t =
boost::gil::image_view<boost::gil::memory_based_2d_locator<boost::gil::memory_based_step_iterator<boost::gil::pixel<flo
at,
 boost::gil::layout<boost::mpl::vector1<boost::gil::gray_color_t> > >*> > >; lsst::afw::geom::Extent2I =
lsst::afw::geom::Extent
<int, 2>]
   Box2I(Point2I(-1114,-4627),Extent2I(0,0)) doesn't fit in image 15x15 {0}
lsst::pex::exceptions::LengthError: 'Box2I(Point2I(-1114,-4627),Extent2I(0,0)) doesn't fit in image 15x15'
```

"old" release

Forced photometry

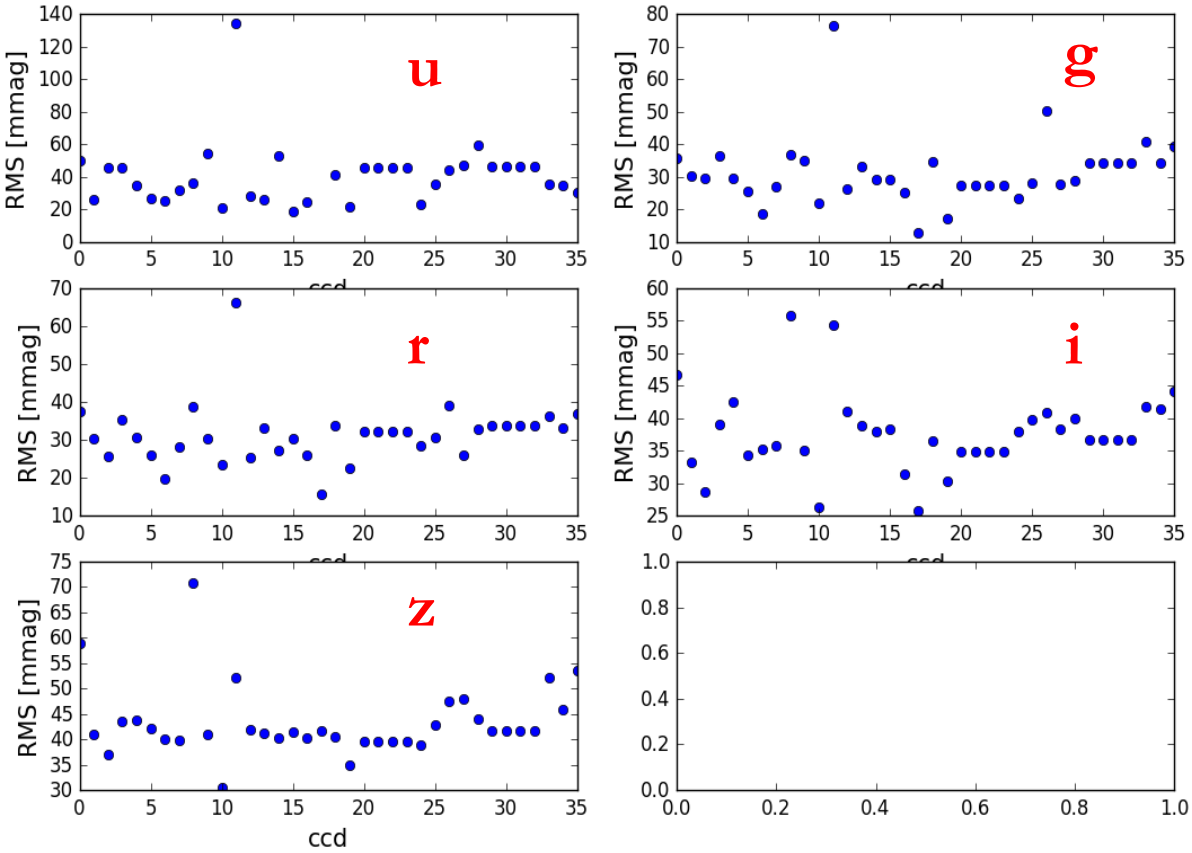
- **Photometry checks : coadded sources <-> forced sources**



"old" release

Forced photometry

Photometry checks : coadded sources <-> forced sources



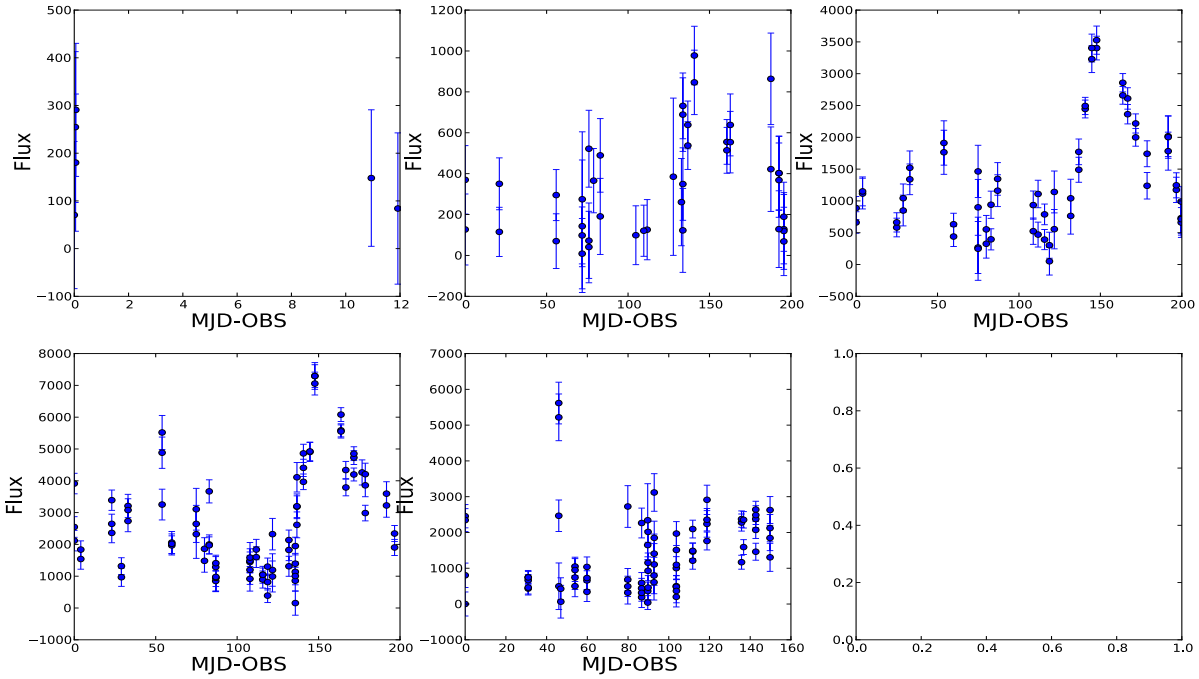
To be done vs mag

"old" release

Forced photometry

- Let us try to grab a type Ia supernova
- From the list of JLA SN, did a (ra,dec) match to identify Sne in our D3 (season 2) sample
- Get the LC from forced photometry sources

ra_ref=215.550590
dec_ref=53.184275
T0 = 53541.012444

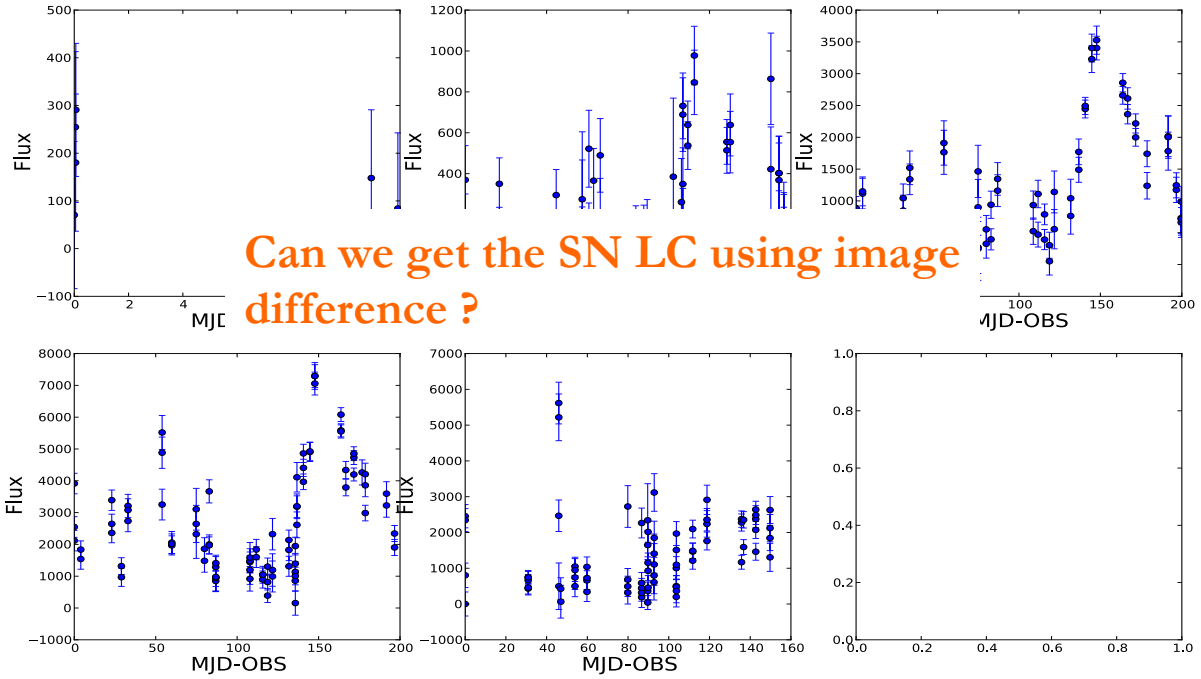


"old" release

Forced photometry

- Let us try to grab a type Ia supernova
- From the list of JLA SN, did a (ra,dec) match to identify Sne in our D3 (season 2) sample
- Get the LC from forced photometry sources

ra_ref=215.550590
dec_ref=53.184275
T0 = 53541.012444



Can we get the SN LC using image difference ?

"old" release

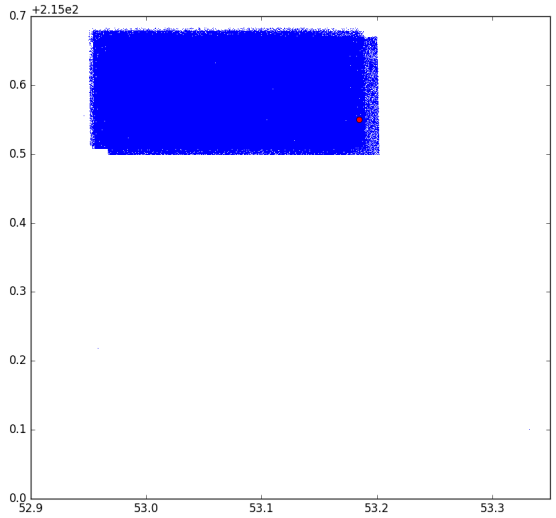
Image differencing

- For each of the five bands:
  - subtract the calexps from one reference image of the period (ImageDifference)
  - lots of errors during the processing:

imageDifference.measurement WARNING: Error in base_PsfFlux.measure on record 123127297808007983:
File "src/PsfFlux.cc", line 132, in virtual void
lsst::meas::base::PsfFluxAlgorithm::measure(lsst::afw::table::SourceRecord&,
const lsst::afw::image::Exposure<float>&) const
Invalid pixel value detected in image. {0}
lsst::meas::base::PixelValueError: 'Invalid pixel value detected in image.'

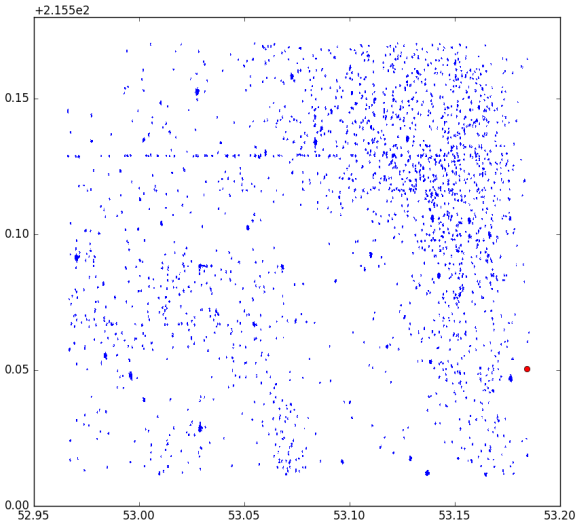
  - multimatch (afw, 1") of the diaSources (per ccd)

At least three measurements in g,r,i



Ra

Dec



Ra

Dec

"old" release

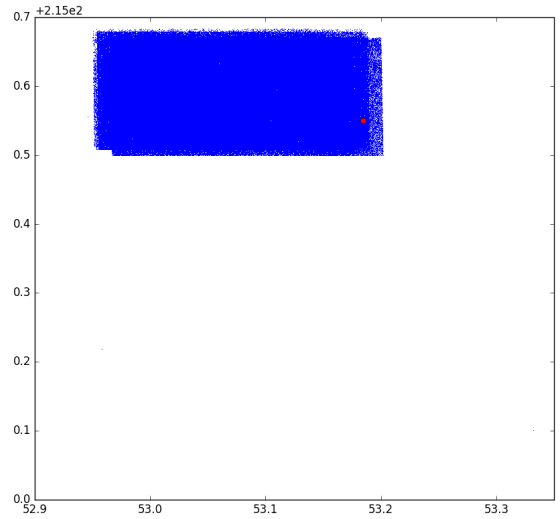
Image differencing

- For each of the five bands:
  - subtract the calexps from one reference image of the period (ImageDifference)
  - lots of errors during the processing:

```
imageDifference.measurement WARNING: Error in base_PsfFlux.measure on record 123127297808007983:
  File "src/PsfFlux.cc", line 132, in virtual void
lsst::meas::base::PsfFluxAlgorithm::measure(lsst::afw::table::SourceRecord&,
const lsst::afw::image::Exposure<float>&) const
    Invalid pixel value detected in image. {0}
lsst::meas::base::PixelValueError: 'Invalid pixel value detected in image.'
```

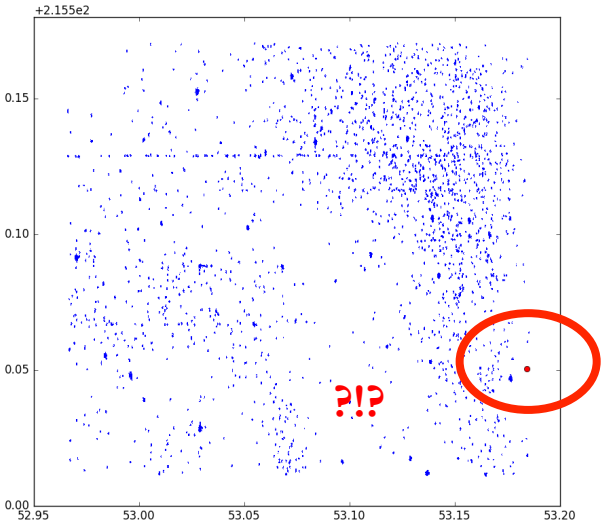
  - multimatch (afw, 1") of the diaSources (per ccd)

At least three measurements in g,r,i



Ra

Dec



Ra

Dec

Automated Data Reprocessing

- For the results presented in this talk : processing steps were performed using a bunch of python (and shell) scripts
- A DESC work-flow engine has been developped @SLAC and is currently used for Twinkles
- SRS pipeline : "Software mechanism for organizing and executing massively parallel computing projects"
- Pipeline main characteristics
  - Automated submission and monitoring of batch jobs
  - Web interface for monitoring and submiting tasks
  - Processing steps : XML files
  - Can run jobs at SLAC and CC-IN2P3
- We have started testing it @CC
- Not that easy – Some tuning probably necessary…but could be an efficient tool to process data (full chain) with the LSST stack

Summary

- **The LSST DM processing chain has been exercised up to the image differencing step.**
- **Exercise being repeated with recent releases**
- **Some tools (validate_drp) are available to estimate the quality of the data produced.**
- **Target : processing the D3 field (up to forced photometry) in ~ one week and a half (now calexps in two-three days) -> automated data processing ?**
- **What may be difficult : to grab a release for which all the processing steps are functional (or to have few releases - compatible - where some of the steps are working).**