Difference Image Analysis in LSST stack

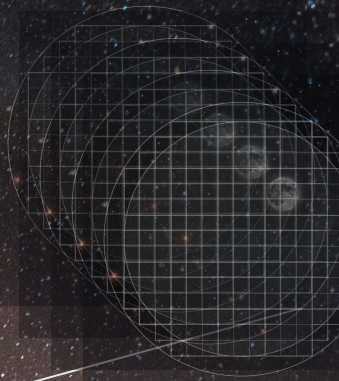
Synthetic source injection for LSST commissioning

Bruno S., Ian S. & AP Team

Time domain Cosmology: from pixel to Constraints

Observation Scheduler / Target follow-up

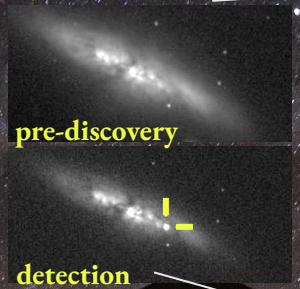
Time domain Cosmology: from pixel to Constraints

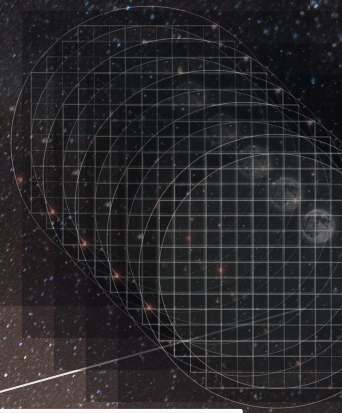


Observation Scheduler / Target follow-up

Time domain Cosmology: from pixel to Constraints

Difference Image Analysis (DIA)

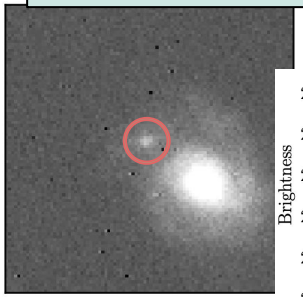




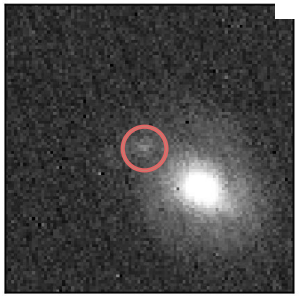
Observation Scheduler / Target follow-up

Time domain Cosmology: from pixel to Constraints

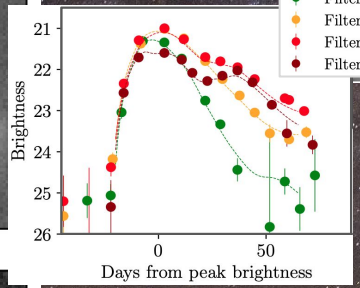
Follow-up observations
Accurate photometry & Calibration



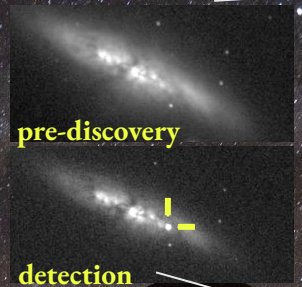
MJD:60144.39

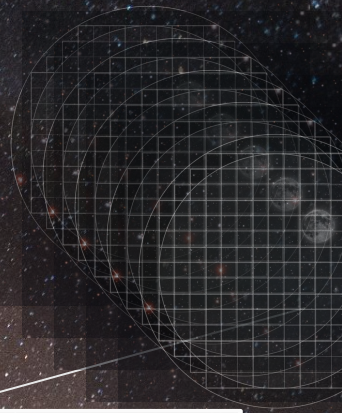


MJD:60150.39



Difference Image Analysis (DIA)



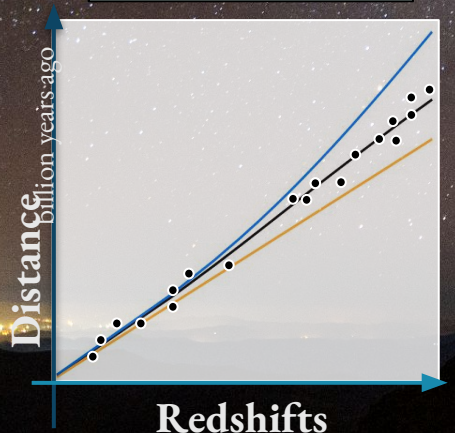


Observation Scheduler / Target follow-up

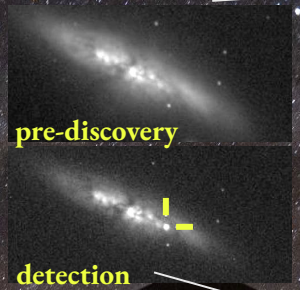
Time domain Cosmology: from pixel to Constraints

Follow-up observations
Accurate photometry &
Calibration

Light-curve
standardization
Cosmology data vector



Difference Image Analysis (DIA)

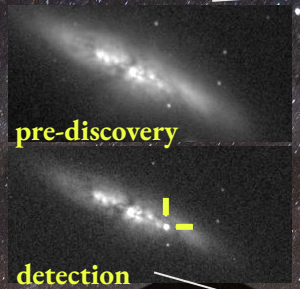


Observation
Scheduler /
Target follow-up

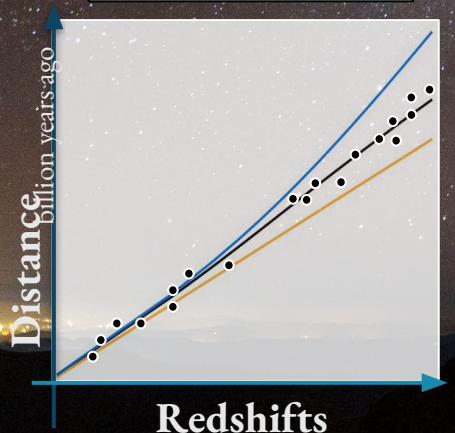
Time domain Cosmology: from pixel to Constraints

Follow-up observations
Accurate photometry &
Calibration

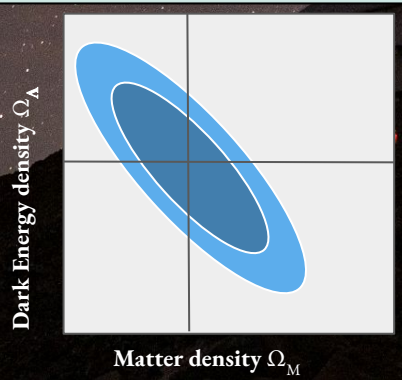
Difference Image Analysis (DIA)



Light-curve
standardization
Cosmology data vector



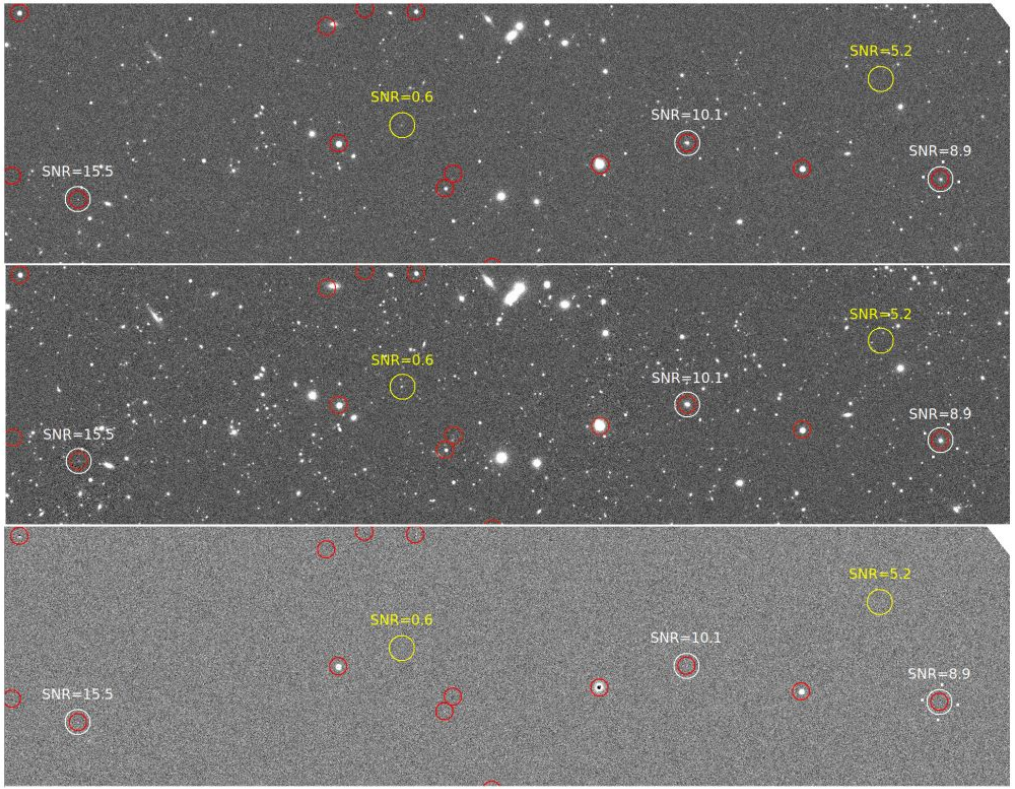
Cosmological Parameter
inference



Observation
Scheduler /
Target follow-up

Difference Image Analysis

- **DIA is the workhorse behind** a large portion of Vera Rubin LSST **Time Domain science**
- LSST Data volume *is already making* tiny or rare effects (corner cases) a massive daily problem to solve
- There is new and exciting science where the limit cases take place
  - Very low SNR transients
  - Very noisy environments
  - …

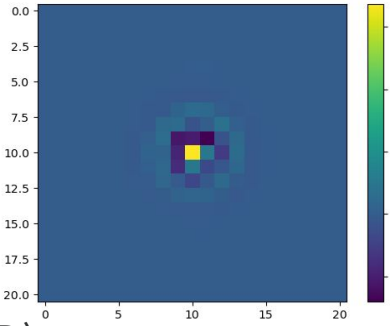


lp_diffim A&L Subtraction tasks

Difference Image Analysis consists on pixel-to-pixel comparisons.

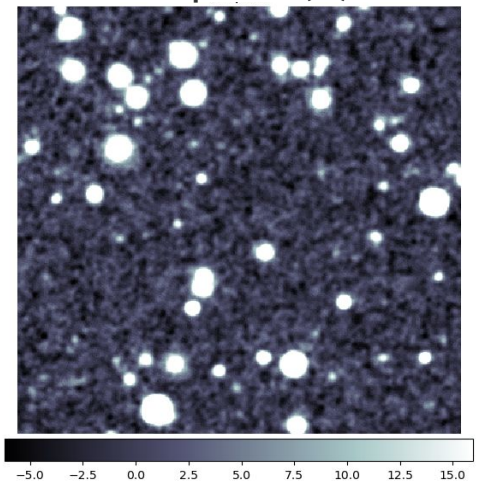
- Nominal Alard & Lupton 98/2000 DIA algorithm

$$D = N - k * R$$

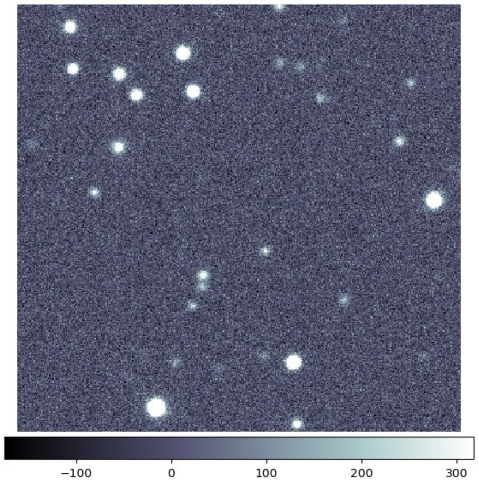


Kernel

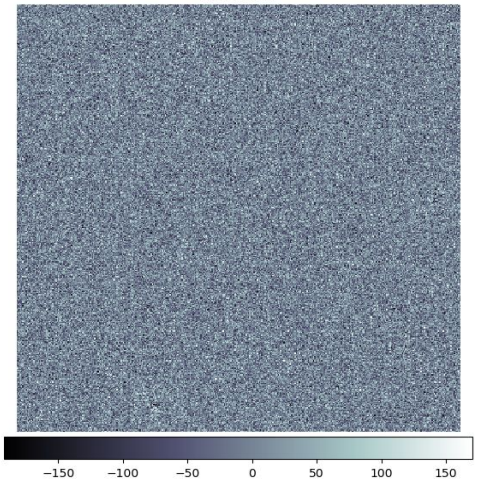
Template (R)



Science (N)



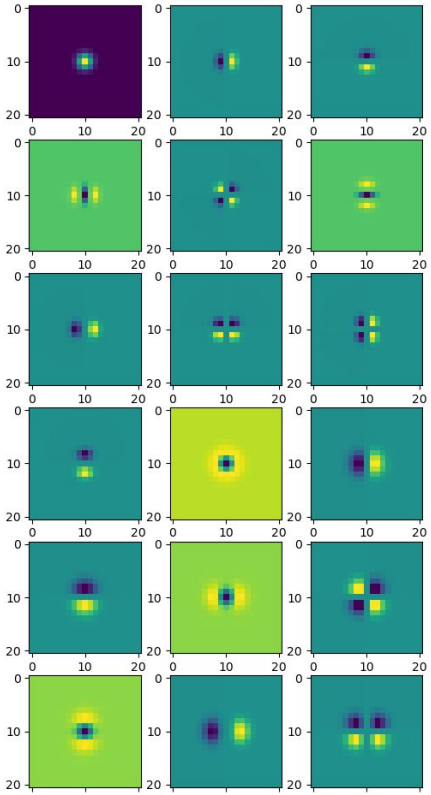
Difference (D)



Inspection of the `psfMatchKernel`

The kernel is constructed using Gaussian basis functions, modulated by low-order polynomials.
Gauss-Hermite basis functions.

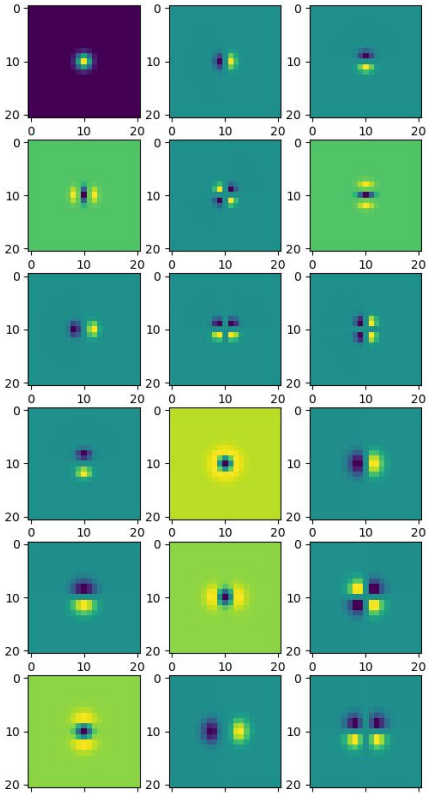
Some of the fitted components look like this:

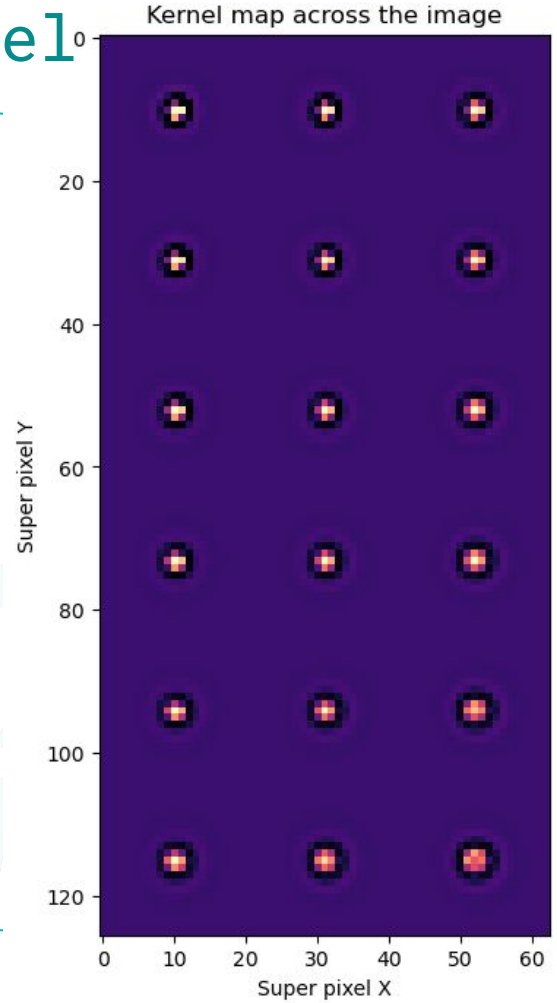


Inspection of the psfMatchKernel

The kernel is constructed using Gaussian basis functions, modulated by low-order polynomials.
Gauss-Hermite basis functions.

Some of the fitted components look like this:

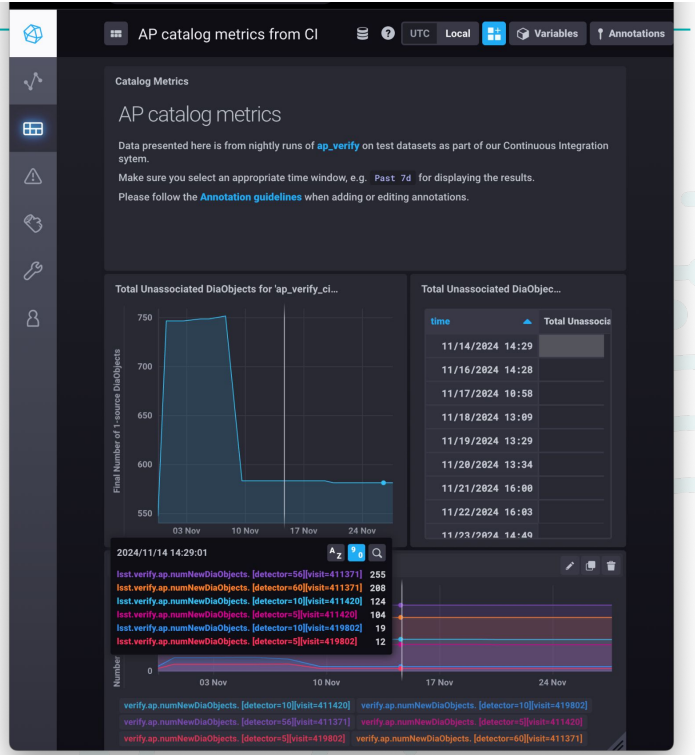




Testing DIA in the stack

For testing DIA we use several levels and scales:

- Single **unit tests** that verify code integrity
- **Continuous Integration** tests (CI) that verify pipeline interlocking blocks
- CI with datasets (we call them “verify”) that report **metrics** (Chronograph)
  - HSC images
  - DECam images
  - DC2 simulation images
  - LSSTComCam (soon a new CI set)
- There are also, development testing that we run.
  - Production results on LSSTCam



LSST Commissioning Camera

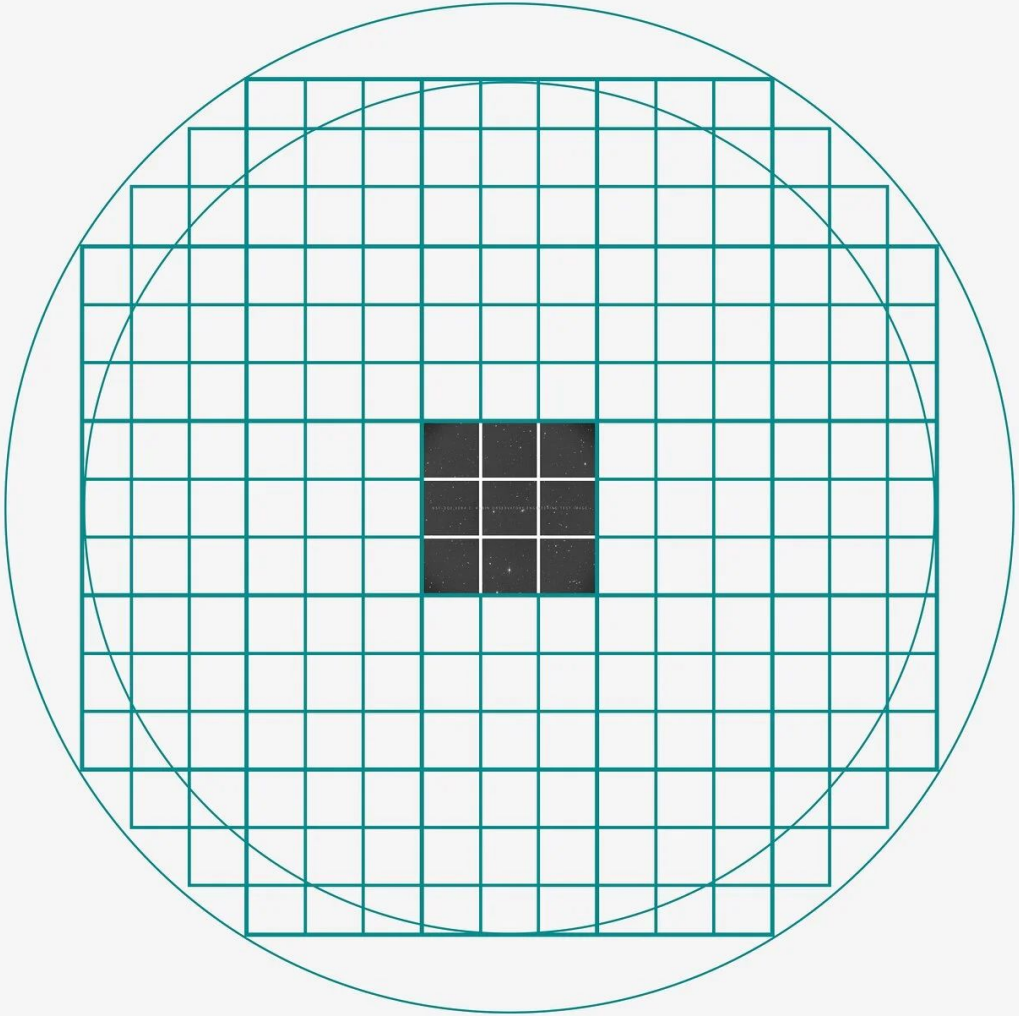
LSSTComCam is 9 CCDs (1 raft)

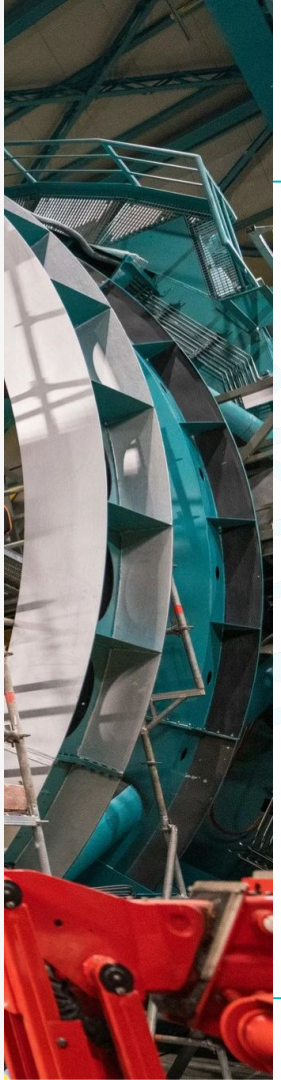
LSSTCam is 189 CCDs (x20)

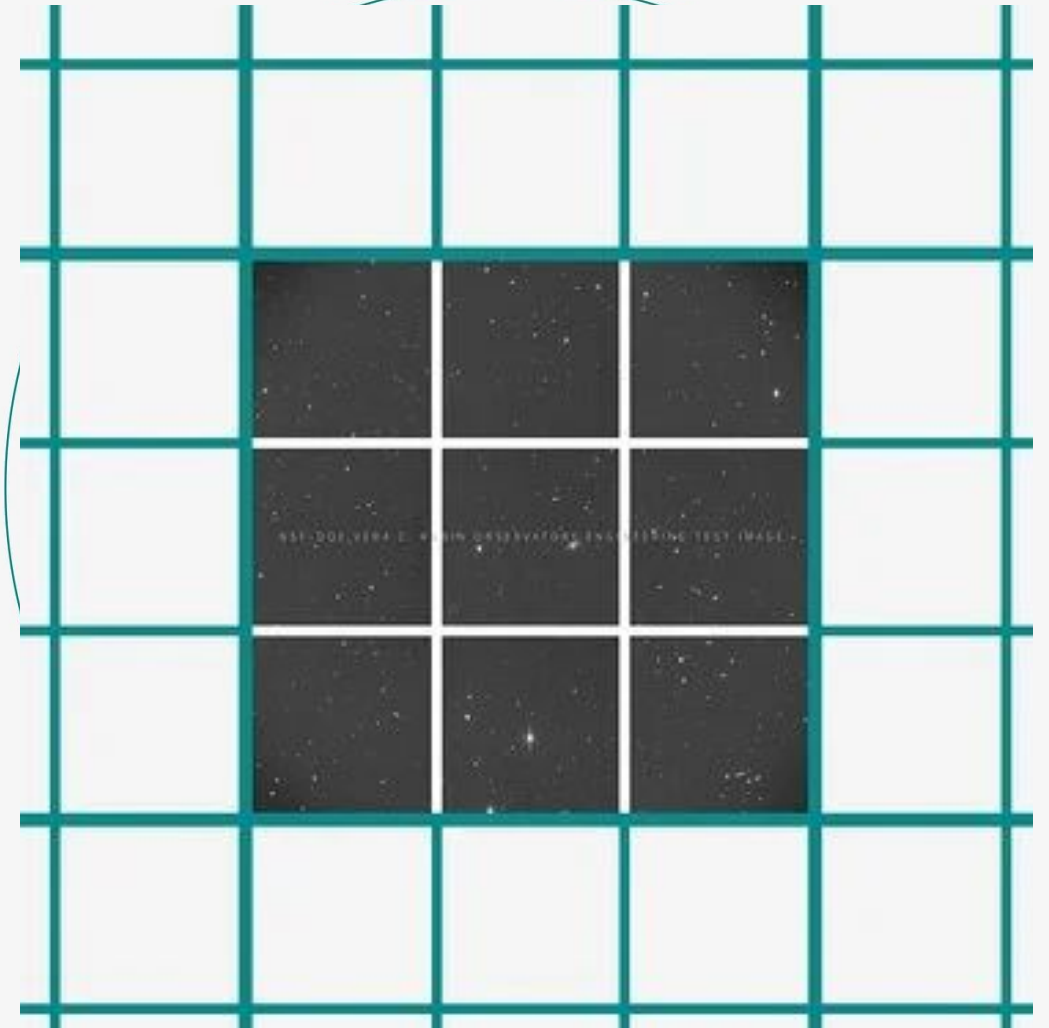
Also: CCDs are replacement parts, and not as free of artifacts as LSSTCam ones

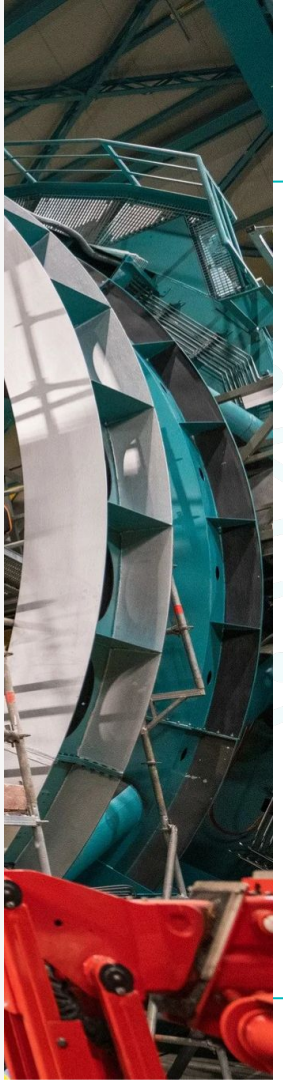

Layher
Layher

VERA C
OBSERV





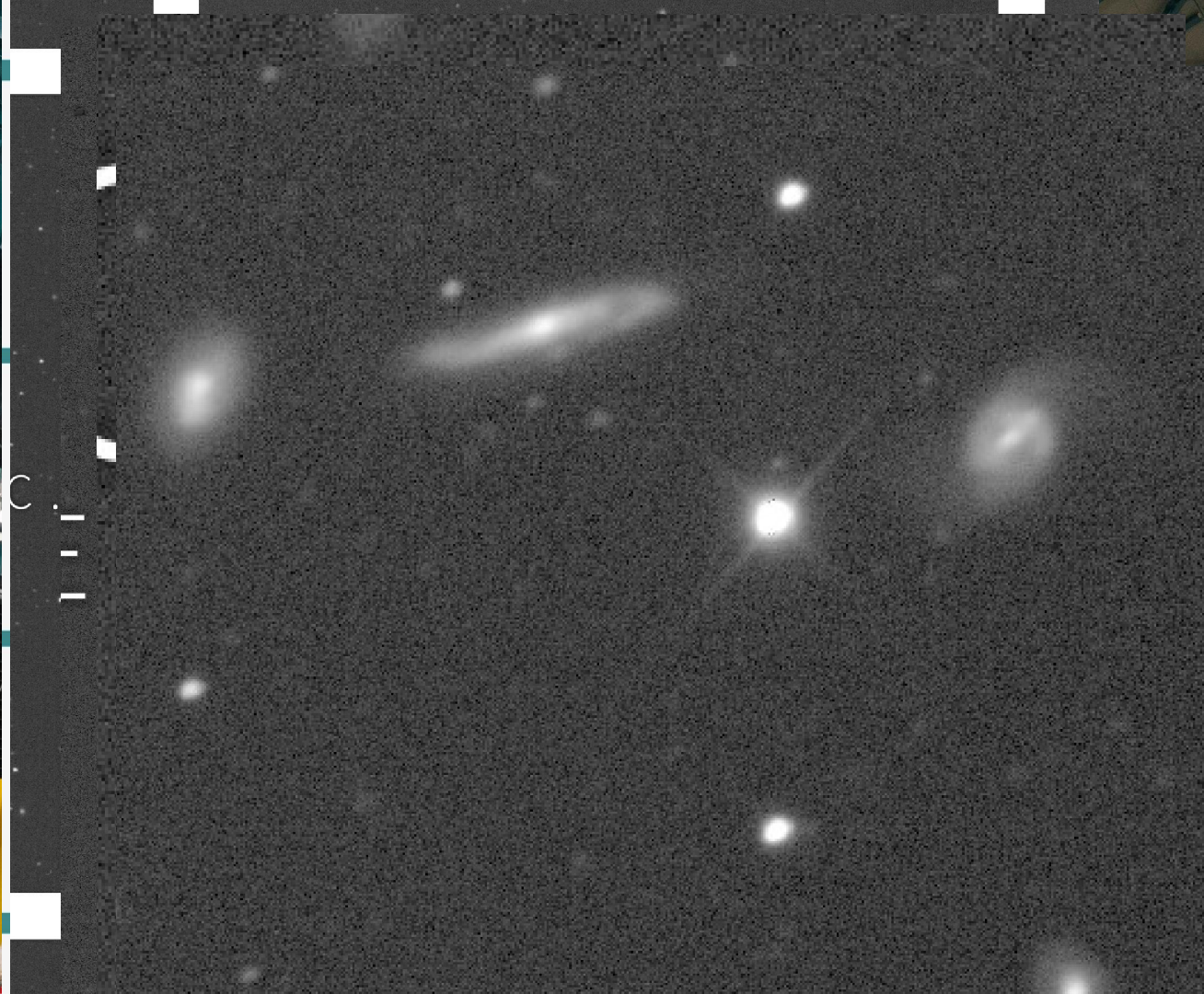





NSF-DOE VERA C. RUBIN OBSERVATORY ENGINEERING TEST IMAGE

C. RUBIN OBSERVATORY ENGINE


D R Y



Time domain observations LSSTComCam

Several fields with LSSTComCam observations at different epochs.

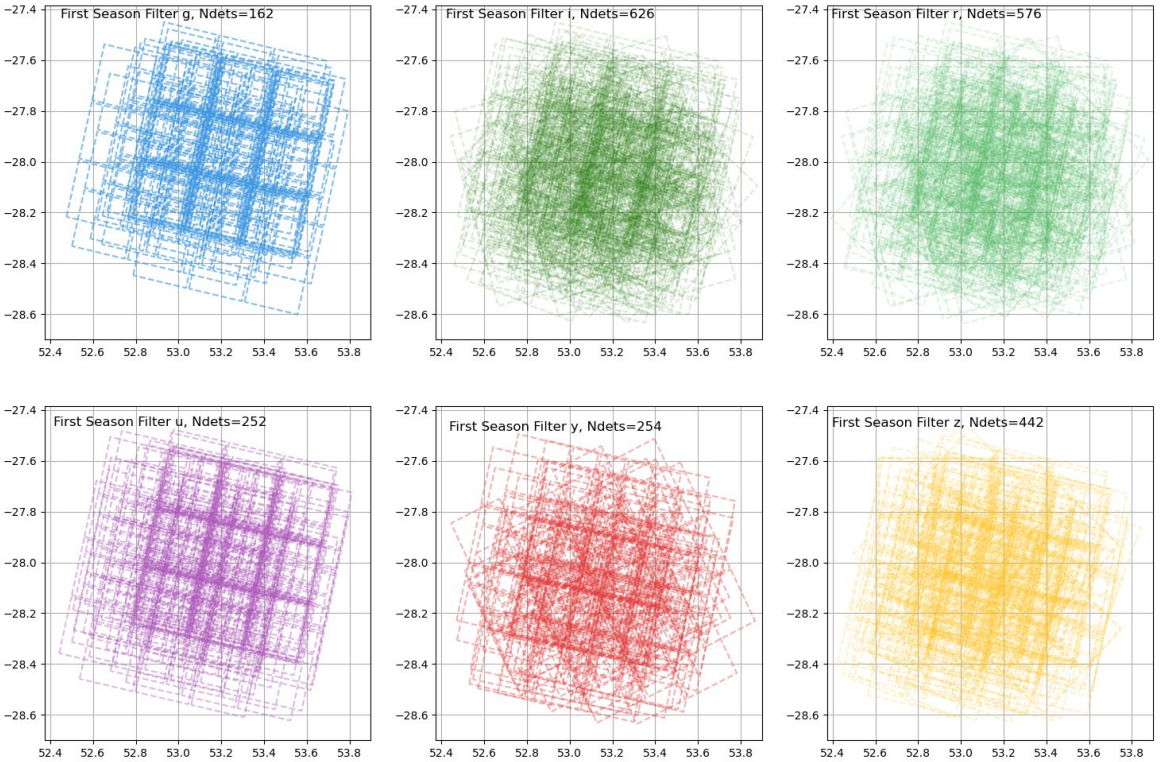
- Validation runs for Difference Image Analysis
- Solar System Science pipeline testing
- Validation with fake source injection

Target	RA	Dec
47 Tuc Globular Cluster (47 Tuc)	6,02	-72,08
Low Ecliptic Latitude Field (Rubin SV 38 7)	37,86	6,98
Fornax Dwarf Spheroidal Galaxy (Fornax dSph)	40,00	-34,45
Extended Chandra Deep Field South (ECDFS)	53,13	-28,10
Euclid Deep Field South (EDFS)	59,10	-48,73
Low Galactic Latitude Field (Rubin SV 95 -25)	95,00	-25,00
Seagull Nebula (Seagull)	106,23	-10,51

Time domain observations LSSTComCam

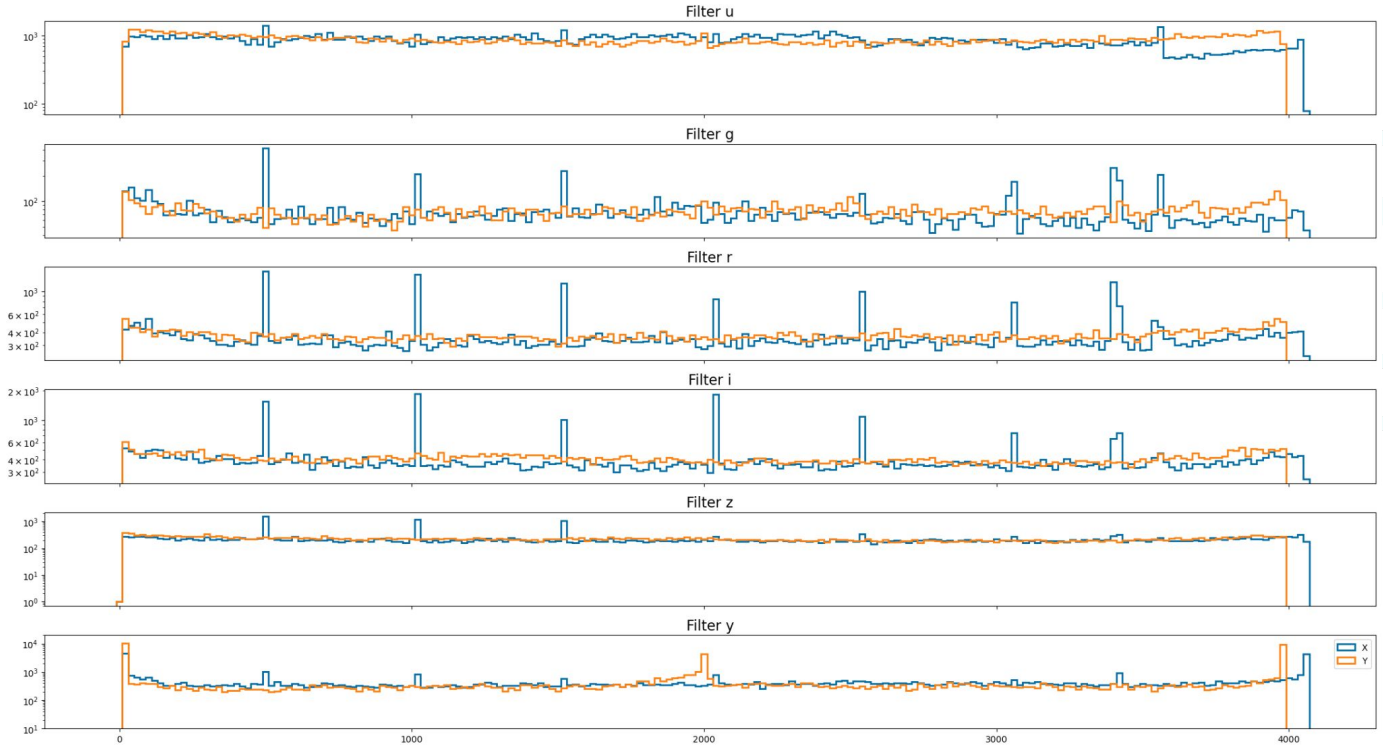
Several fields with LSSTComCam observations at different epochs.

- Validation runs for Difference Image Analysis
- Solar System Science pipeline testing
- Validation with fake source injection



Time domain observations LSSTComCam

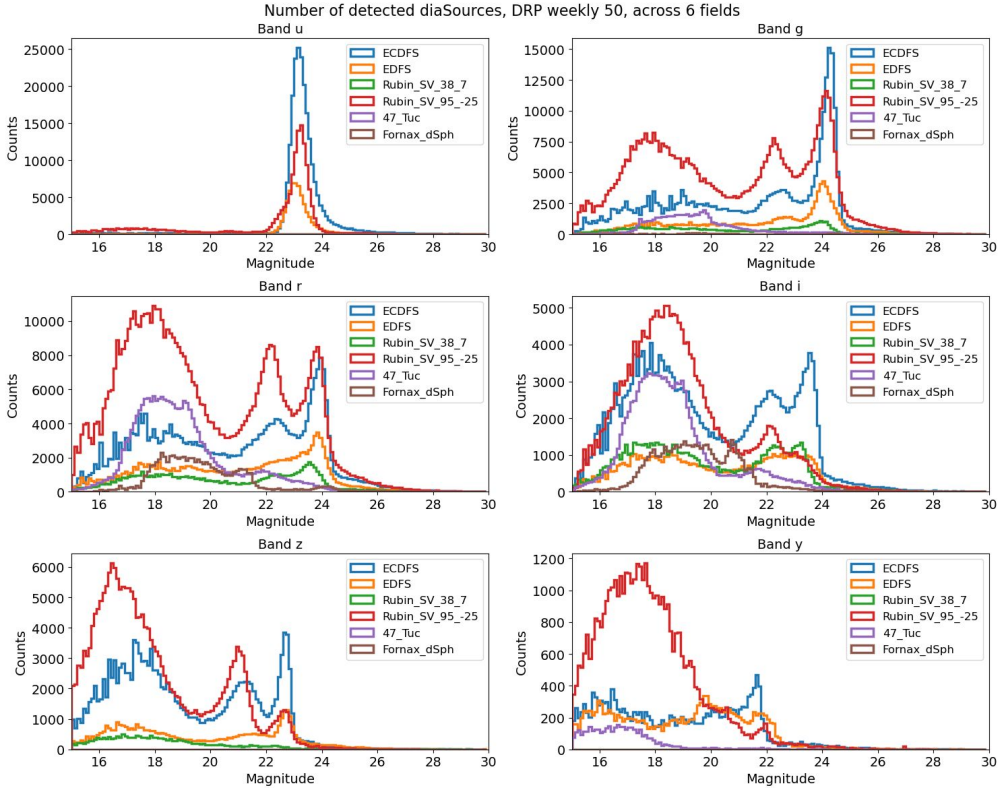
Diagnostics on LSSTComCam were released in the SITCOMTN-149 technical report were we gave a set of initial test results to the community



Synthetic sources: HSC example

Observations spanned several fields and we obtained extremely valuable information

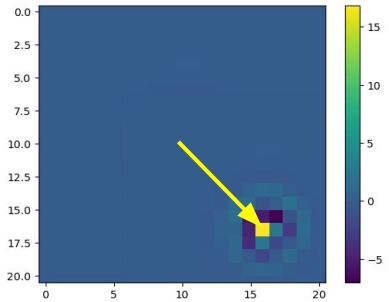
Individual detections for each field, in each band.
Also we gathered light-curves as well as all metadata information related to the data processing

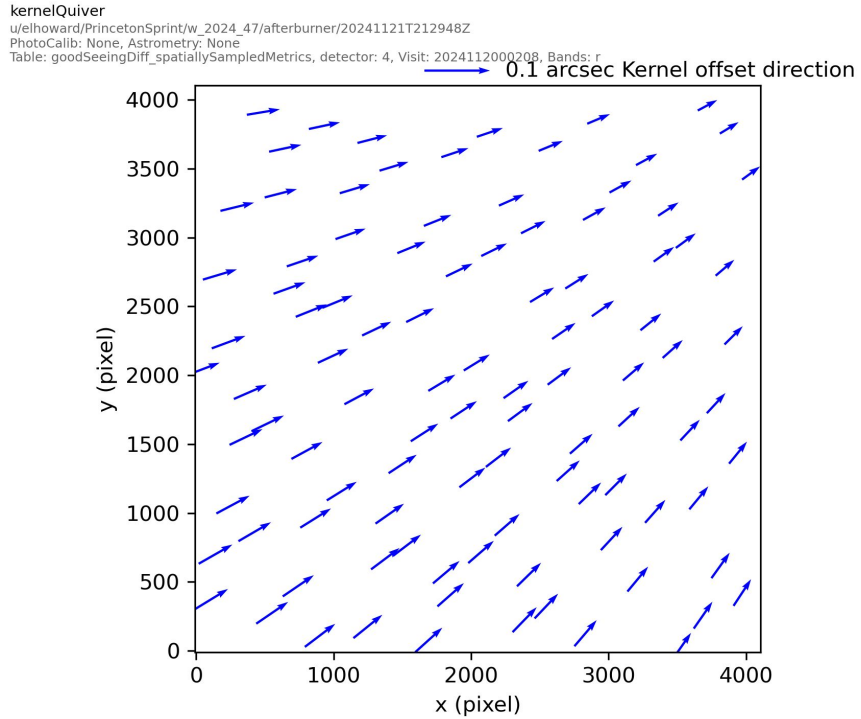


Inspection of the psfMatchKernel

The kernel contains information about the transformation needed to match both images.

For instance looking at its centroid we can generate maps of the smallest astrometric shift



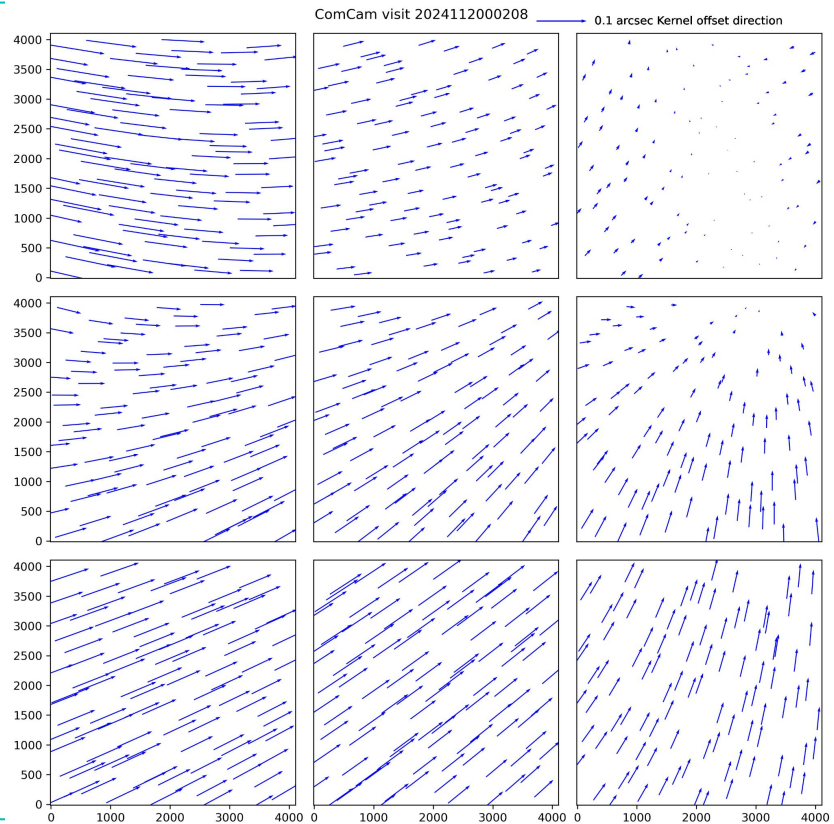


Inspection of the psfMatchKernel

The kernel contains information about the transformation needed to match both images.

For instance looking at its centroid we can generate maps of the smallest astrometric shift

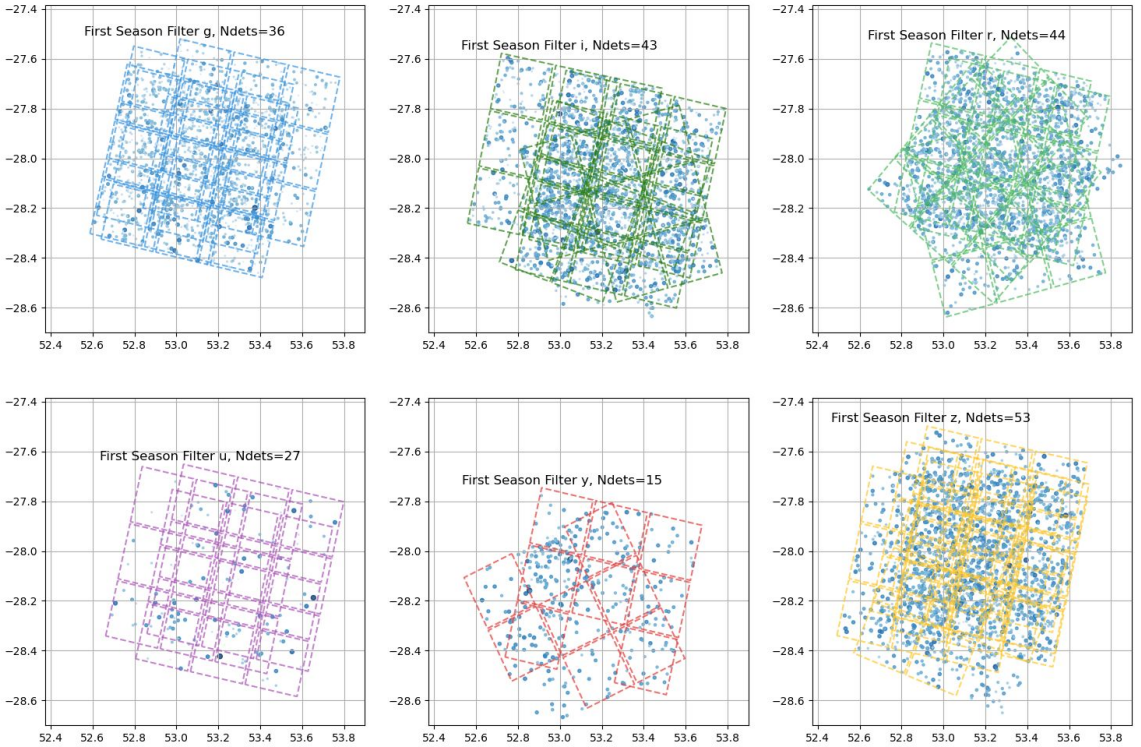
This is for LSSTComCam 9 CCDs



Synthetic sources in LSSTComCam images

Validation with fake source injection. A very large campaign of ~1.5 x 10^6 sources

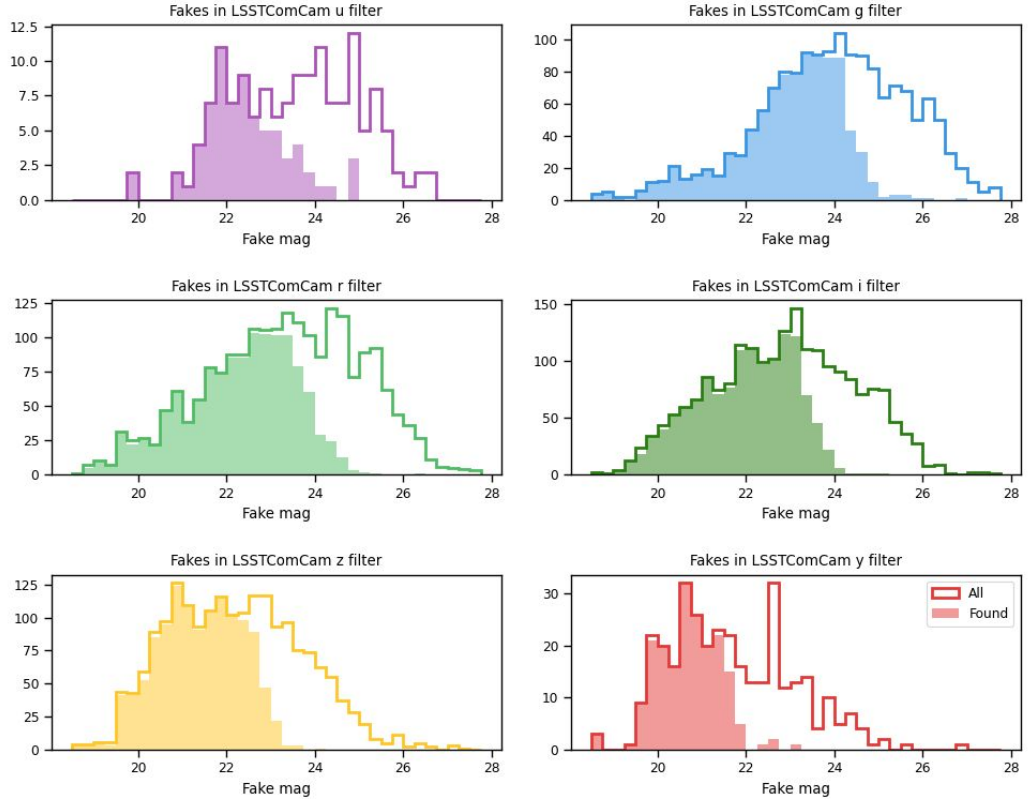
- Injection of point sources
- Reprocessing of all images through DRP/AP pipelines
- Validation of transient finding pipeline
- Training set building for Real/Bogus
- Preliminary DP1 preview report



Synthetic sources in LSSTComCam images

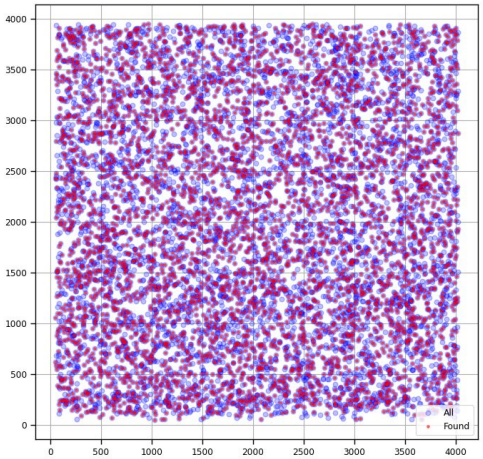
The first metric to check is the fraction of detections.

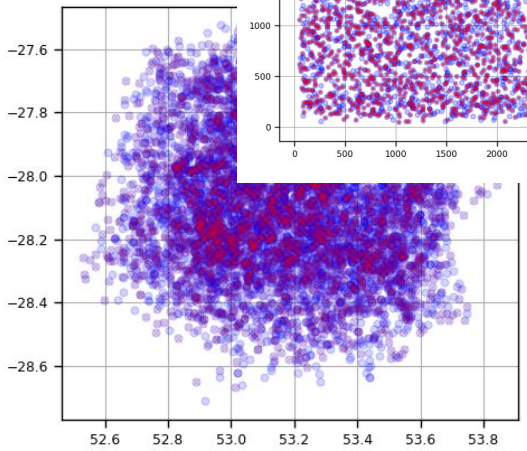
In this case we check as function of magnitudes



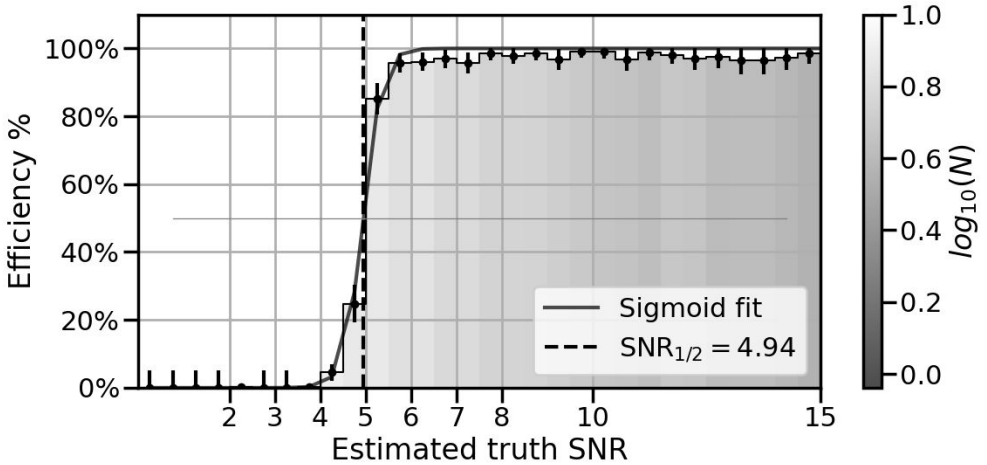
Synthetic sources in LSSTComCam images

CCD plane





Sky plane

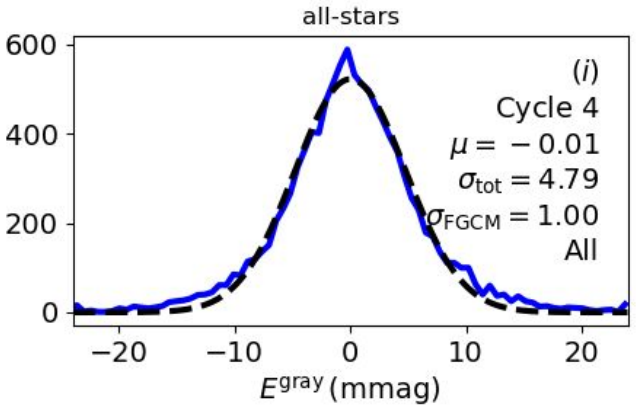


We define the value of
$S/N_{1/2}$:= SNR at which
Recall = 50%.
(Lower $SNR_{/2}$ is more sensitive)

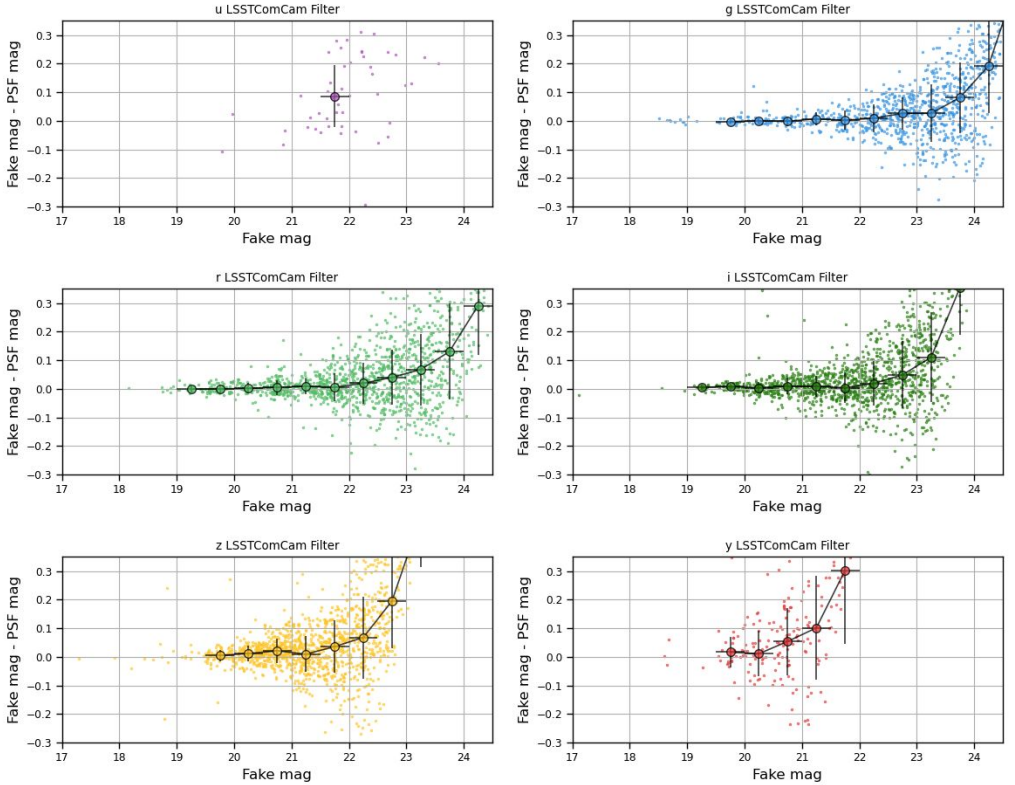
Synthetic sources in LSSTComCam images

Photometric measurements:

- PSF photometry vs fake true magnitudes



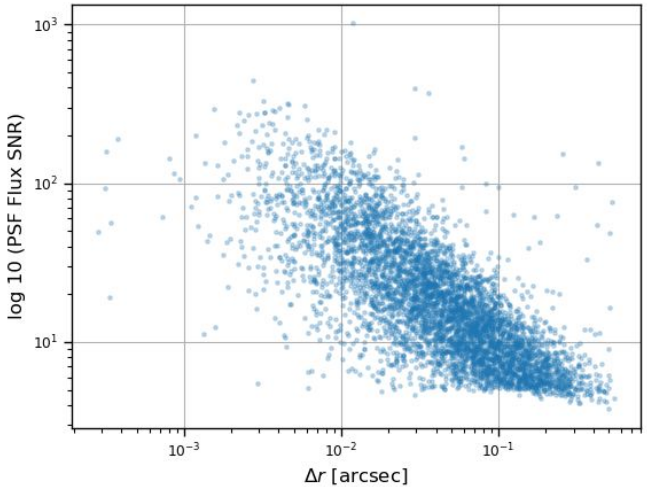
Calibration star repeatability

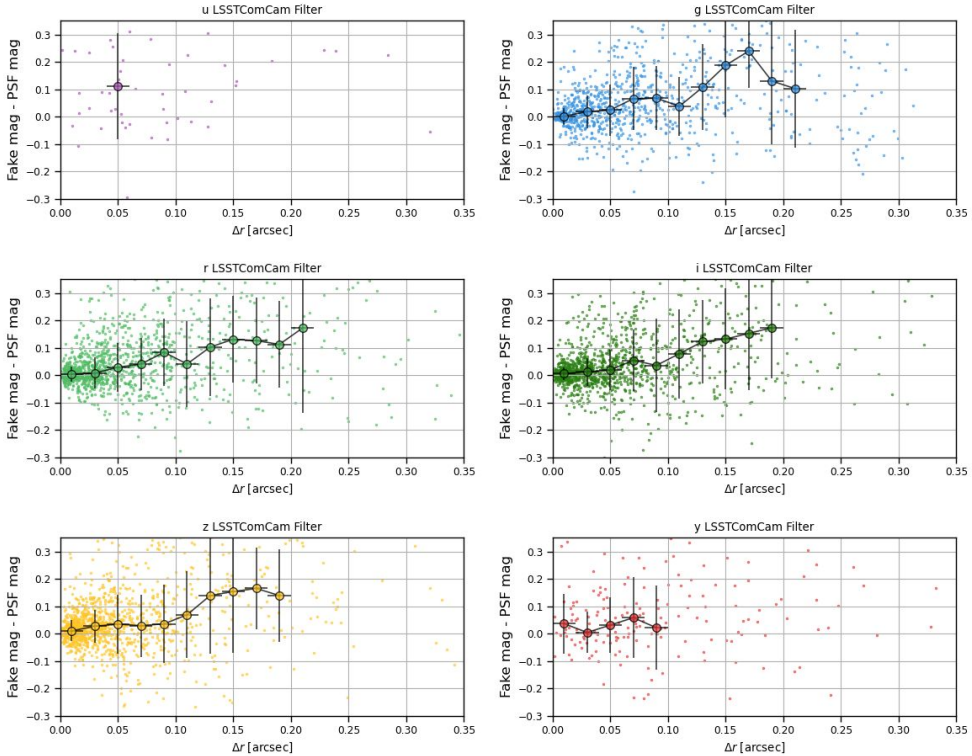


Synthetic sources in LSSTComCam images

Photometric measurements:

- PSF photometry vs. Astrometric centroid distance to true centroid.

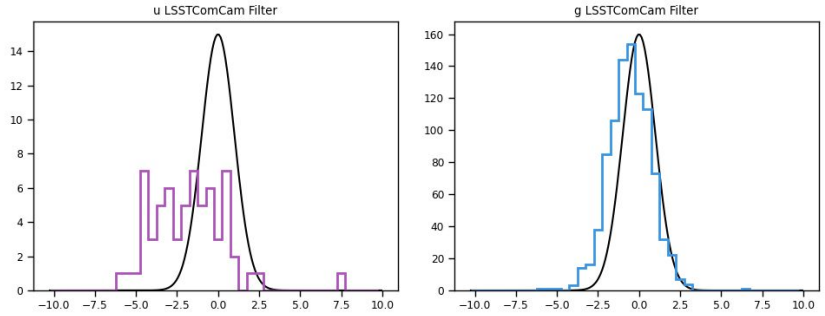


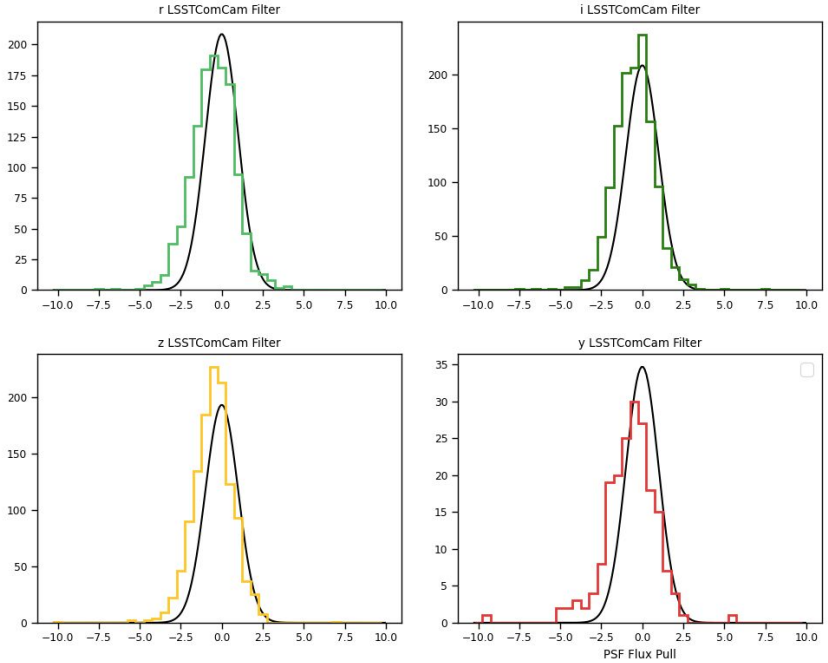


Synthetic sources in LSSTComCam images

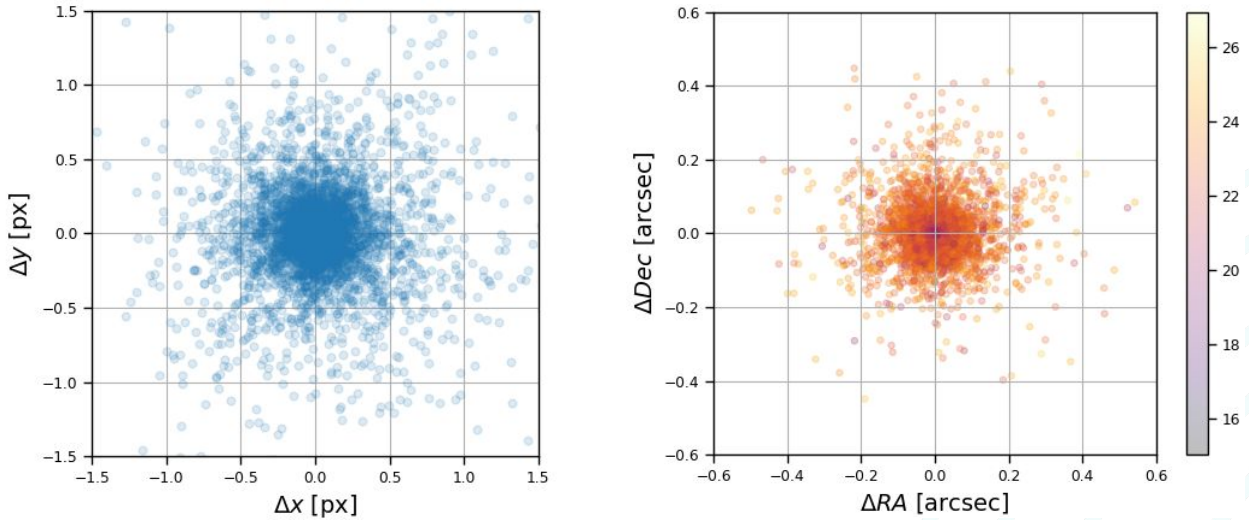
Photometric measurements:

- Flux pulls for PSF photometry. We find variable results and wavelength dependency.
- LSSTCam doesn't look like this





Synthetic sources in LSSTComCam images



Astrometric measurement of our Fakes position. In pixels, sky coordinates, and in combination with the brightness of the source

Synthetic sources in LSSTComCam images

Metrics not able to be shown at this point (check DP1 paper and data release!)

- The fraction of found fakes as function of multiple parameters
  - Hosted-hostless fakes
  - Distance to host center
  - Host brightness
  - Position in the focal plane
  - Observing conditions
  - Etc…

- Measurements on the synthetic source
  - Positions
  - Fluxes
  - Uncertainties (variance plane validation)
- Levels of contamination
- Background and foreground scenes

2025: LSSTCam is installed and running

- We are replicating the LSSTComCam injection campaign with LSSTCam data
- It takes more work, and lot more of processing which is at premium.
- Will roll-out very soon fakes in templates as well (negative transient detections)
- Currently designing and testing a day-time processing continuous injection campaigns for QA during operations

Summary so far and what's next

- We are conducting **regular testing and development of DIA algorithms** in the LSST Stack, we use diverse testing methodologies, and diverse datasets

- Valuable information both from visual inspection, low-level and high-level metrics
- The **synthetic source injection verification is a fundamental tool** to uncover problems and have a controlled scenario to assess improvements

- *We are currently applying all of these into LSSTCam data and preliminary results are already internally in discussion.*
- We are implementing day-time processing with fake injections


Thanks for listening!

Back-Up Slides

lp_diffim A&L Subtraction tasks

Difference Image Analysis consists on pixel-to-pixel comparisons.

- Nominal Alard & Lupton DIA algorithm
  $D = N - k * R$

- Auto-convolution mode: perform a re-assignment of the image to transform
  $D = k * N - R$ `if` $(\mathrm{PSF}_N < \mathrm{PSF}_R)$ `else` $D = N - k * R$

lp_diffim A&L Subtraction tasks

Difference Image Analysis consists on pixel-to-pixel comparisons.

- Nominal Alard & Lupton DIA algorithm
  $D = N - k * R$

- Auto-convolution mode: perform a re-assignment of the image to transform
  $D = k * N - R$ `if` ($PSF_N < PSF_R$) `else` $D = N - k * R$

- Pre-Convolution mode: convolution of N with a given known kernel v:
  $PSF_{N'} > PSF_R$; as result k is no longer a de-convolution kernel

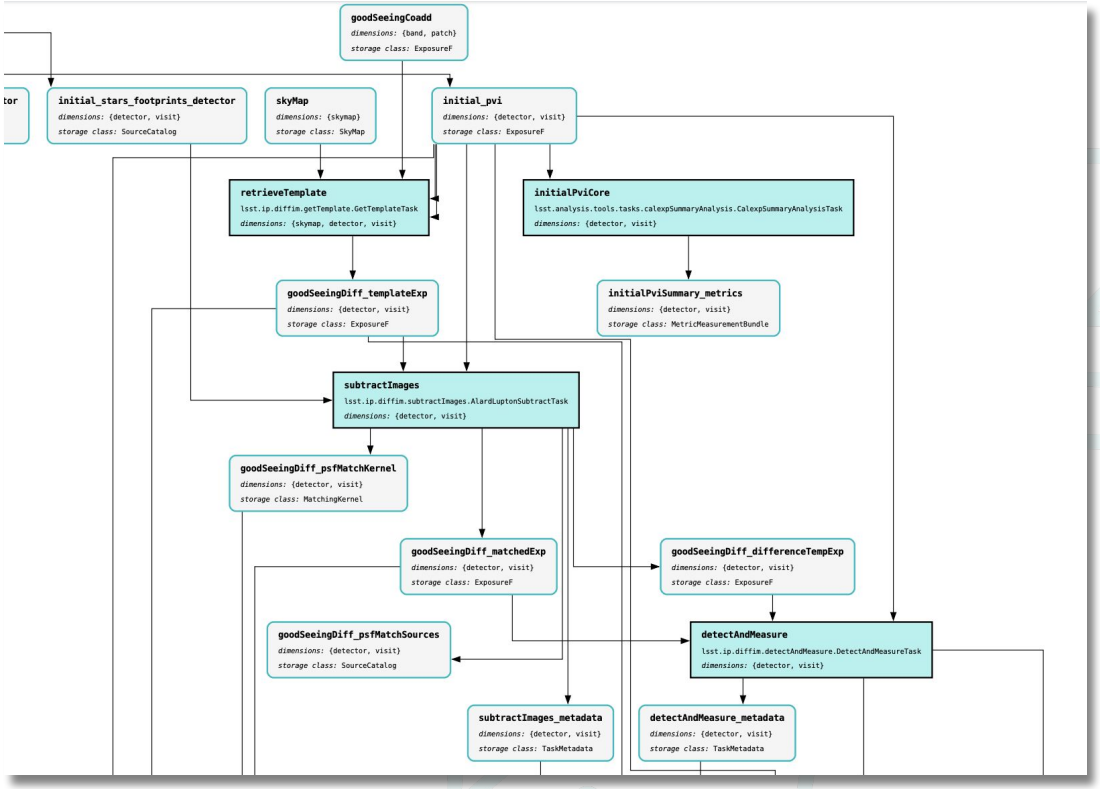
  $v * D = v * N - k * R$

Difference Image Analysis

In project we work with **pipelines**

This is best represented as a **Direct Acyclic Graph (DAG)** that represents the *flow of data through inter-dependent tasks with individual responsibilities*.

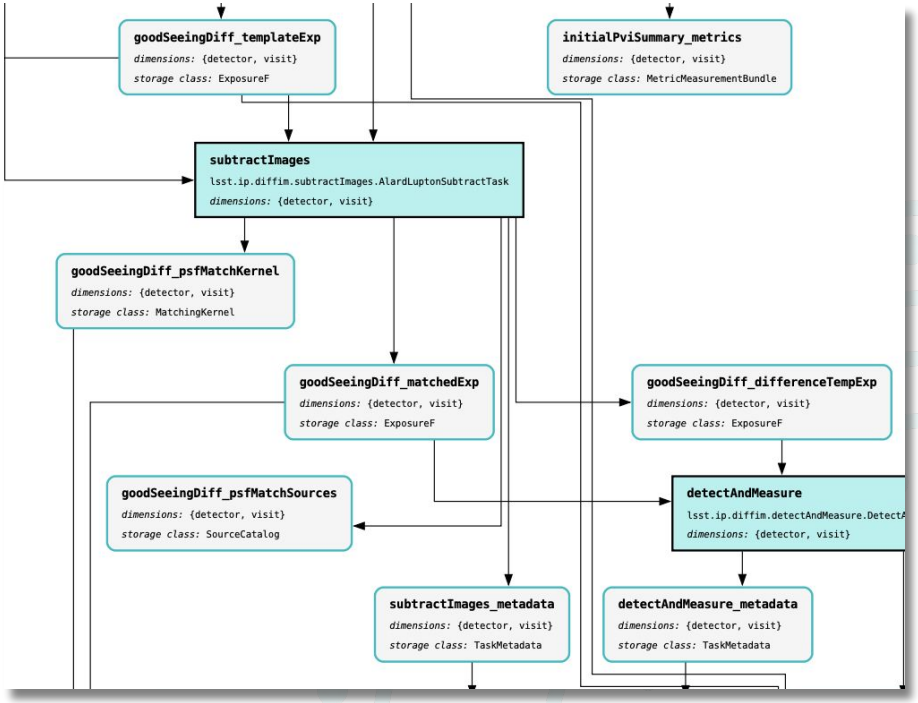
In the case of Alert Production this is the portion that includes the key ***subtractImages task***



The SubtractImages task

The outcome of this task is:

- The image difference: `differenceExp`
- The matching kernel: `psfMatchKernel`
- The template transformed to match the science exposure: `matchedExp`
- Relevant (and valuable) information: `metadata`
- The list of used sources: `psfMatchSources`

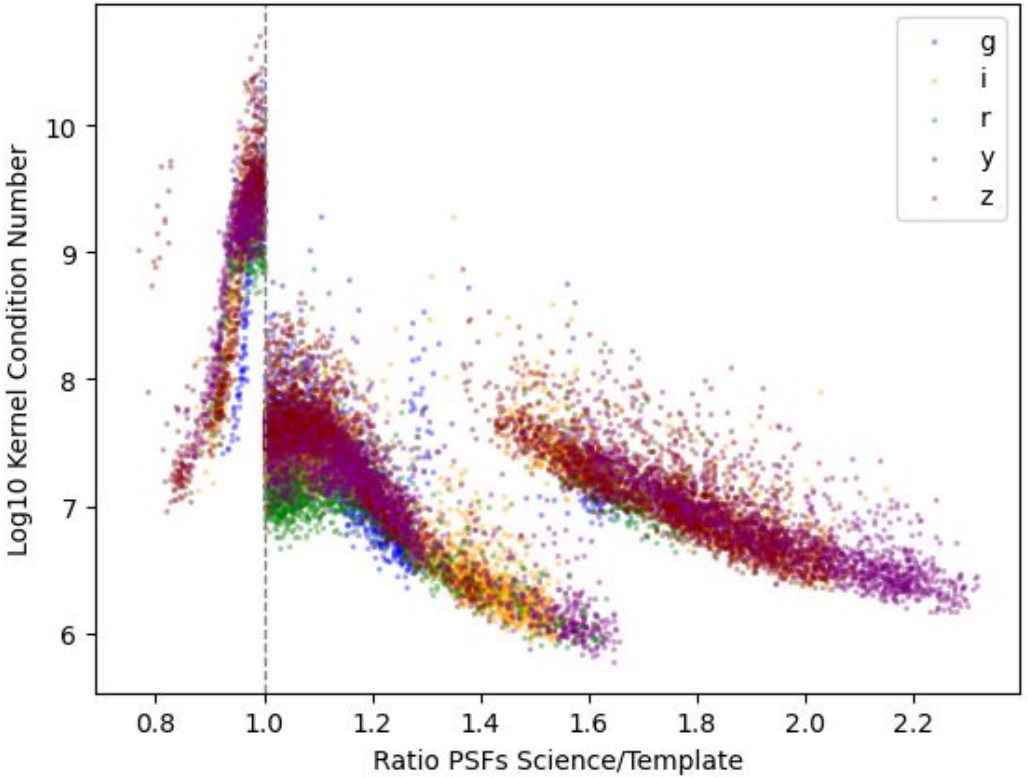


Inspection of the metadata

Metrics or metadata on the kernel is crucial for understanding our performance!

We inspect the dependency of two key parameters with the ratio of PSF sizes:

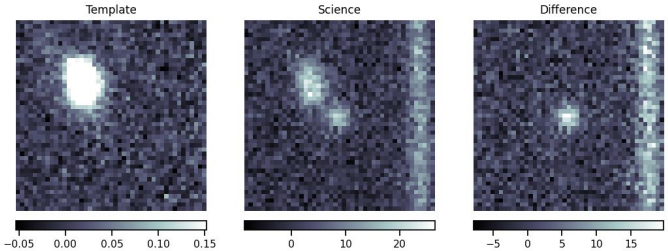
- ratioPsfs = $PSF_{Science} / PSF_{Template}$
- Condition number: the power ratio of kernel top coefficients
- SigGauss: the smallest size of the kernel basis components

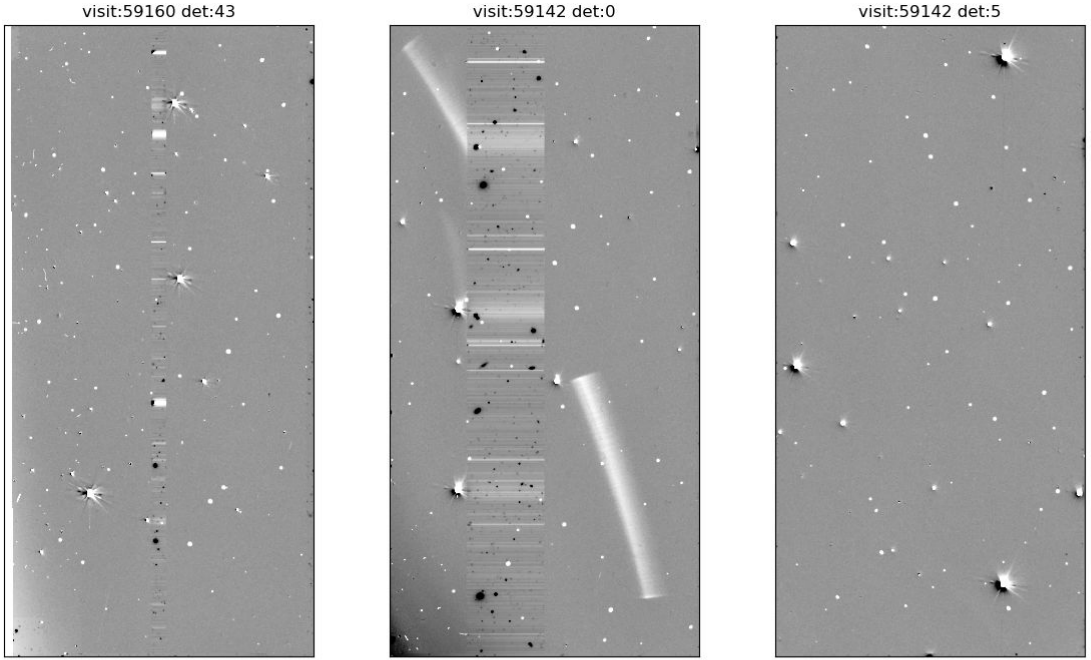


Inspecting the `differenceExp`

We check for visual artifacts, and also **inject synthetic sources**, to test recovery metrics for example

CI dataset from HSC ap_verify_cosmos_pdr2

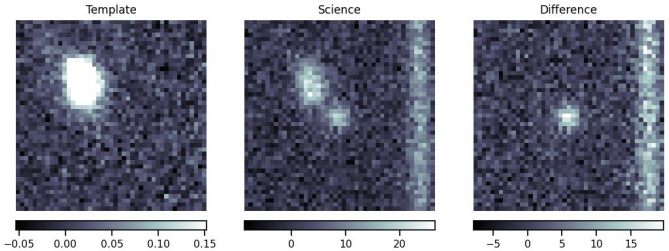


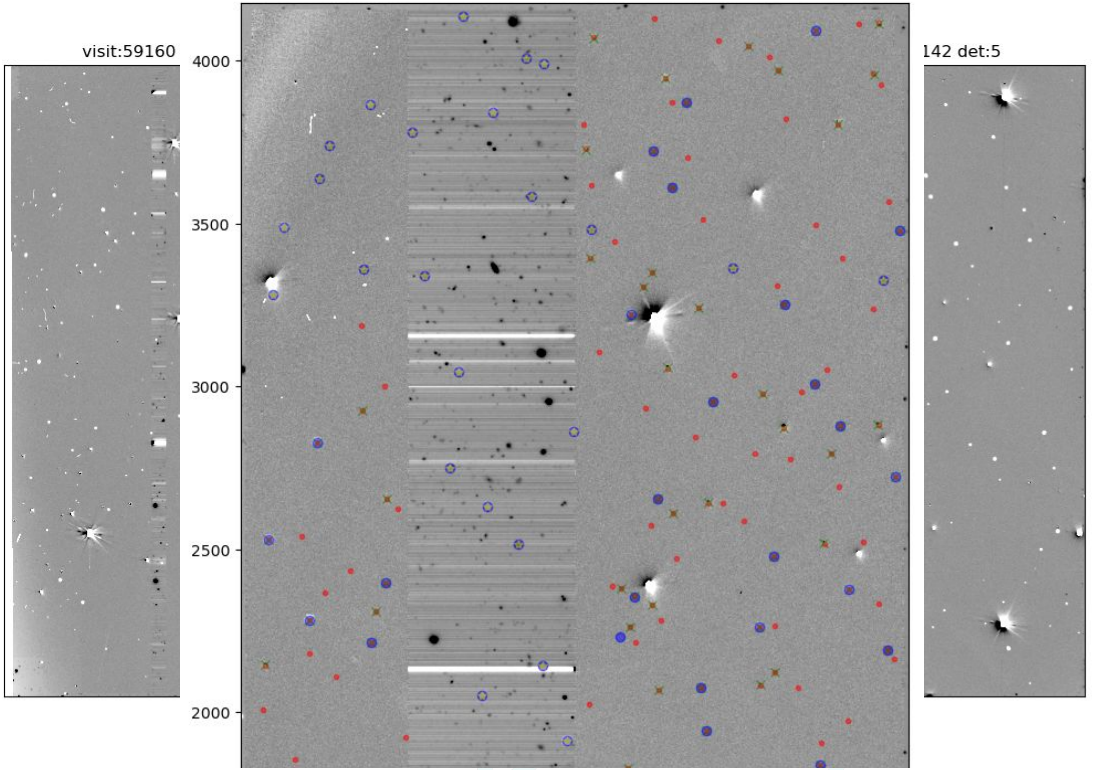


Inspecting the `differenceExp`

We check for visual artifacts, and also **inject synthetic sources**, to test recovery metrics for example

CI dataset from HSC ap_verify_cosmos_pdr2

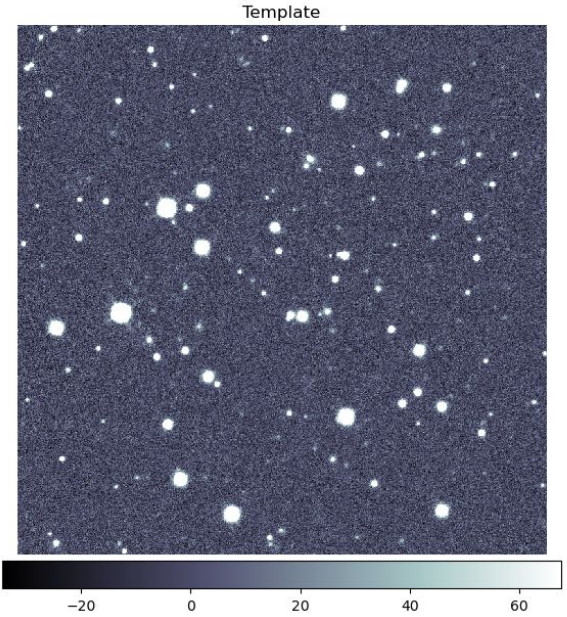


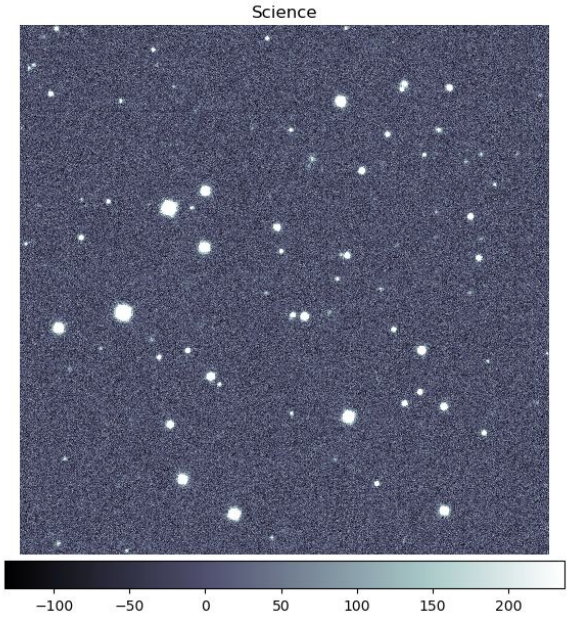


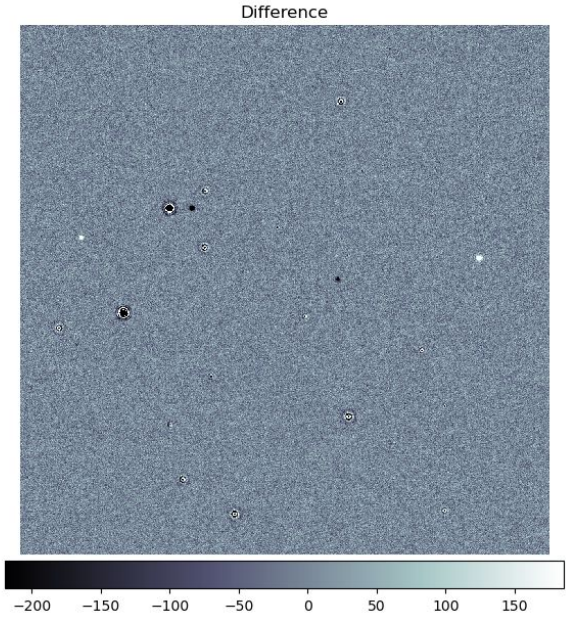
Inspecting the `differenceExp`

- Artifact zoo!
  - Deconvolution (Template PSF is worse than Science PSF)
  - Dipoles (Astrometric registering solution is imperfect, Differential Chromatic Refraction effects)
  - Saturation effects
  - Spikes! LSST will have diffraction spikes
  - Background subtraction effects
  - Correlated noise (wave ripples and granularity effects)
  - Trailed sources (not a DIA problem per se!)
  - …

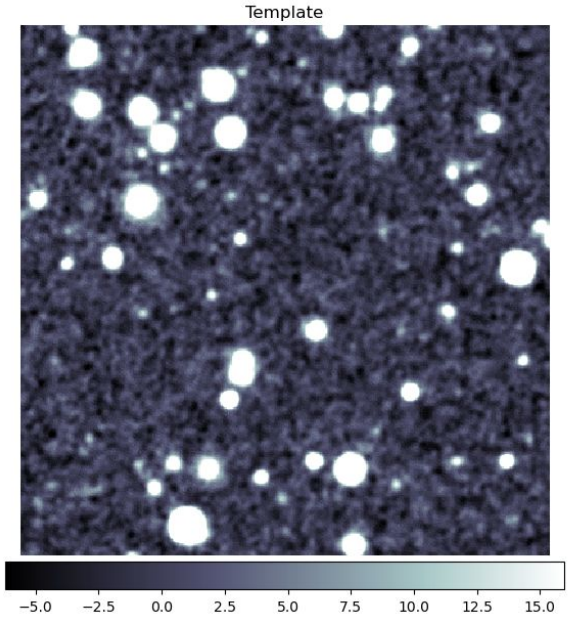
OR4 Dataset inspection

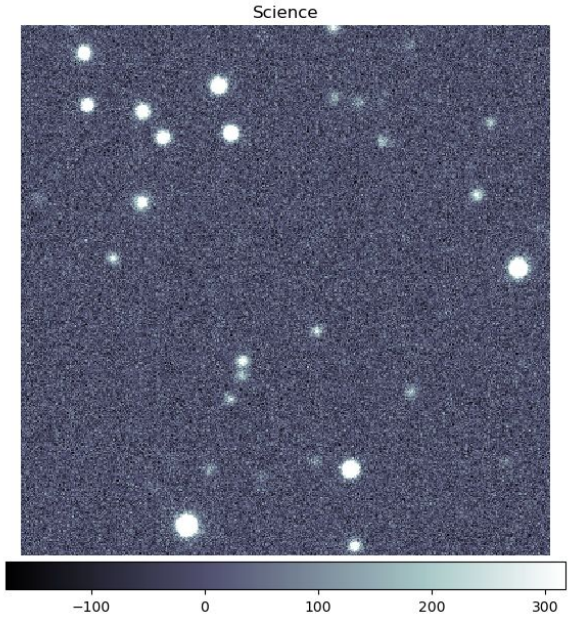


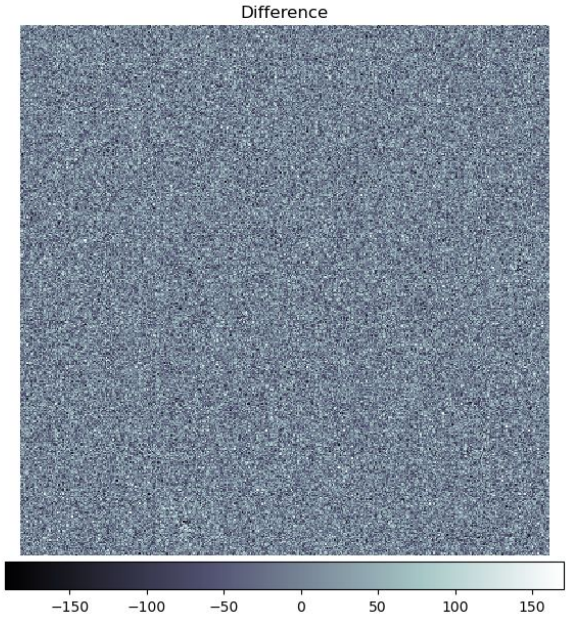




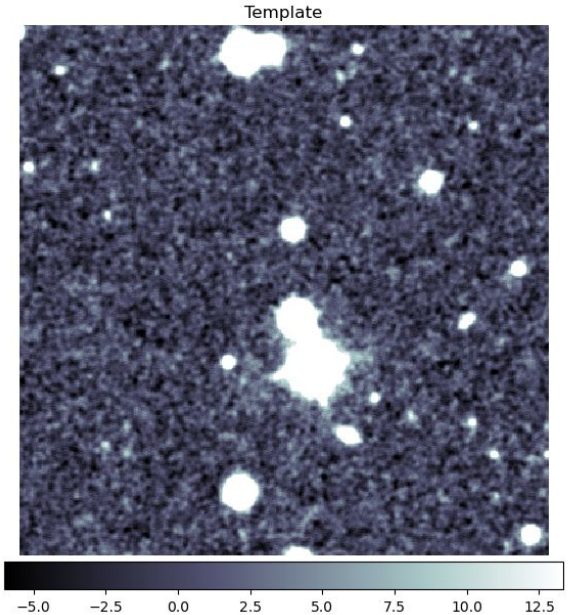
OR4 Dataset inspection

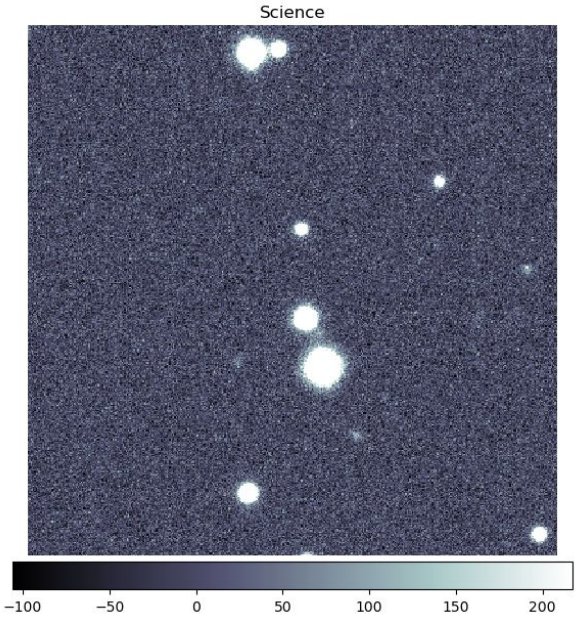


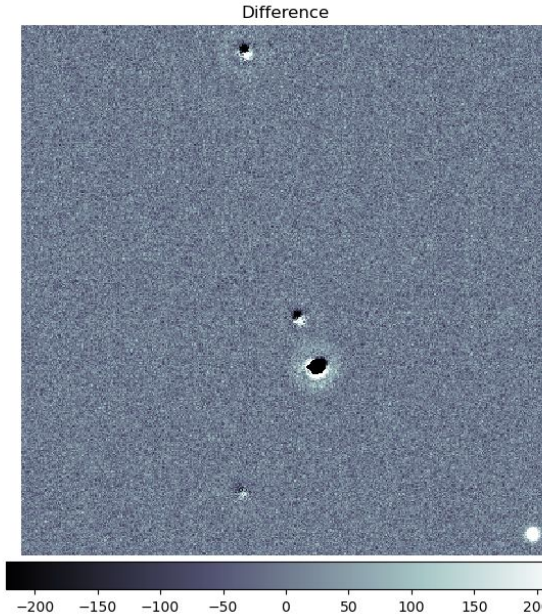




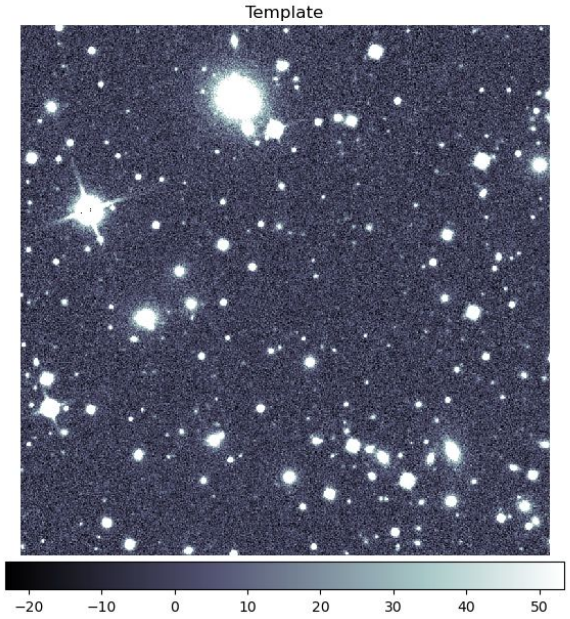
OR4 Dataset inspection

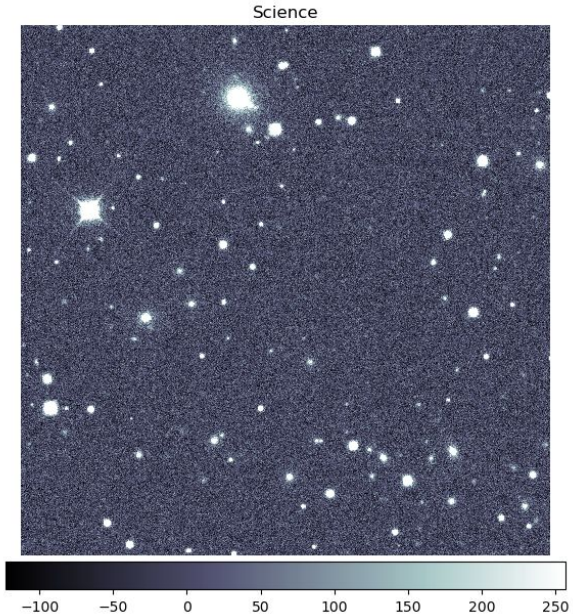


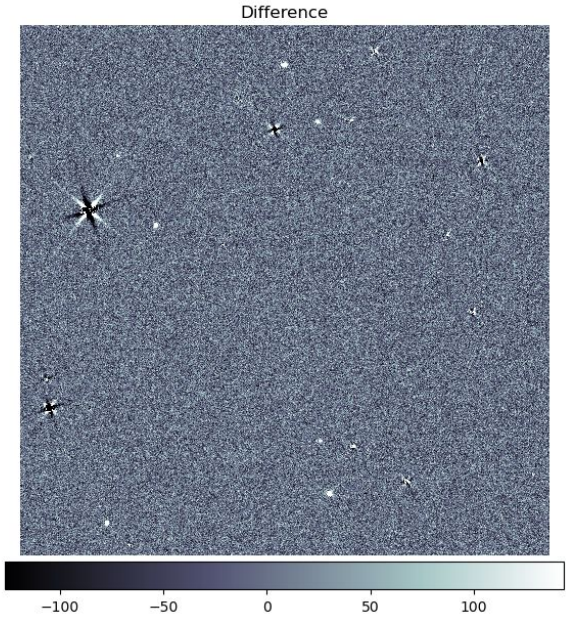




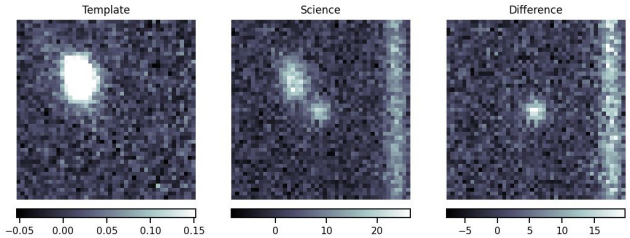
OR4 Dataset inspection

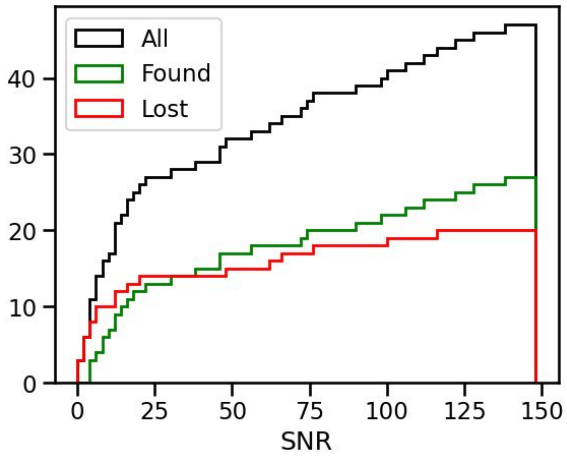


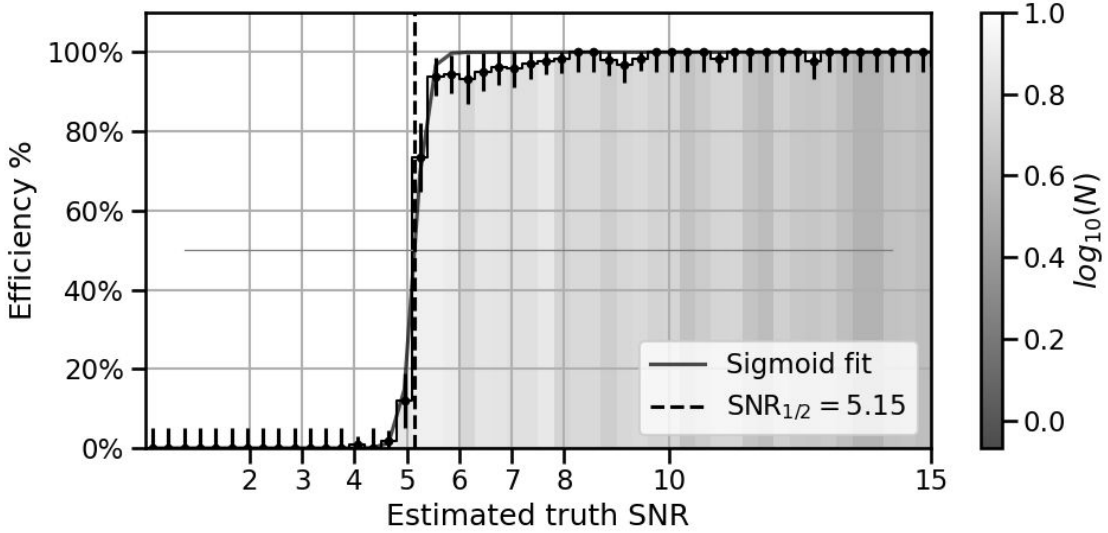




Synthetic sources: HSC example



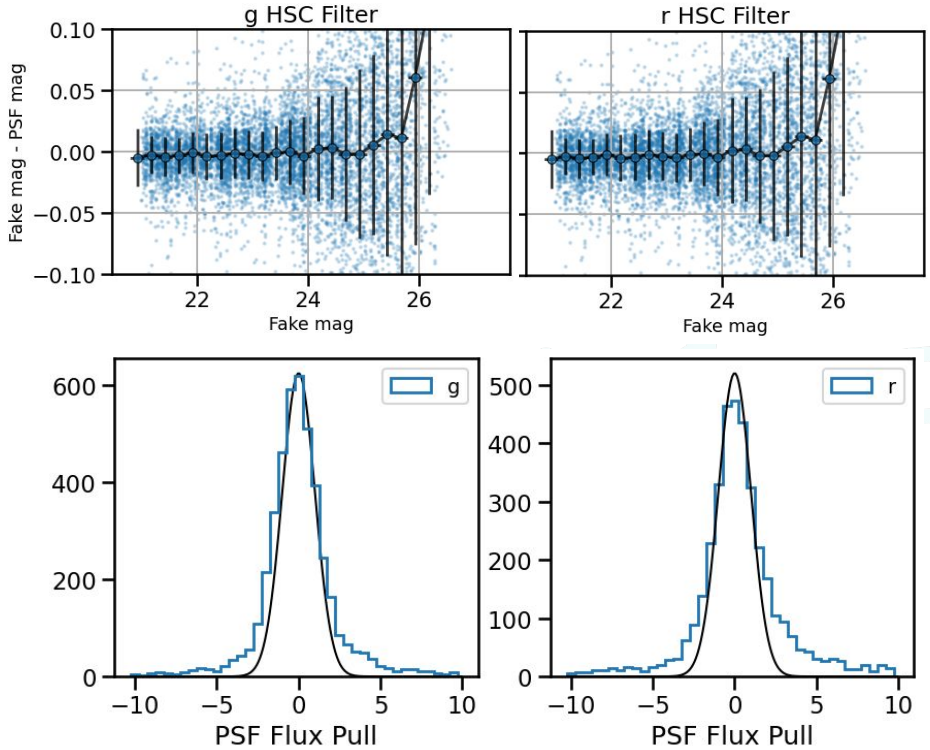




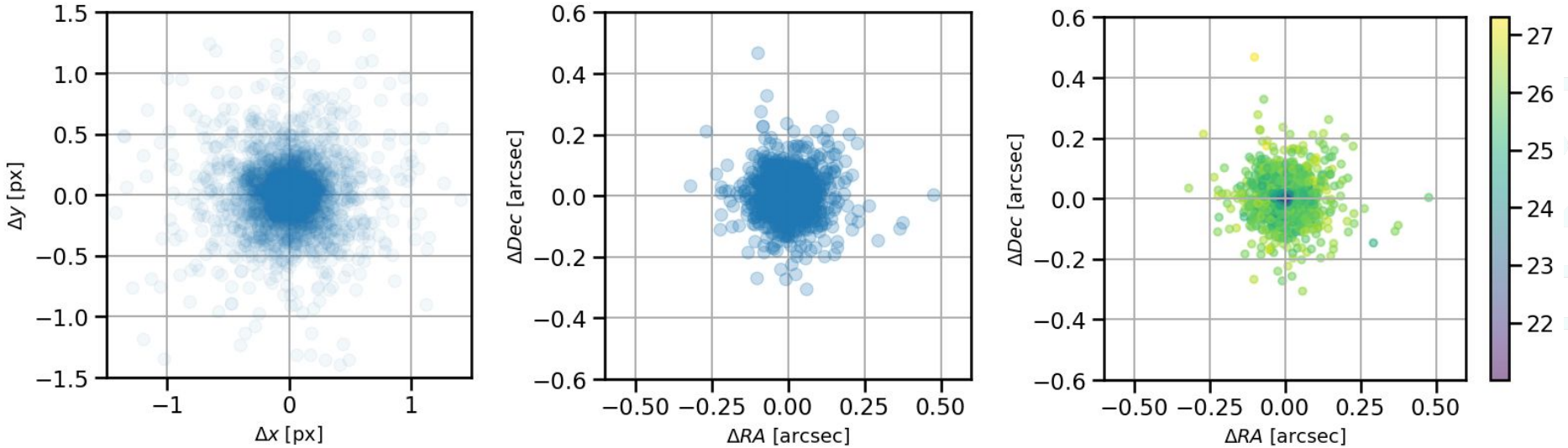
We define the value of
$S/N_{1/2}$:= SNR at which
Recall = 50%.
(Lower $SNR_{/2}$ is more sensitive)

Synthetic sources: HSC example

With fakes we can also test measurement results, like for example photometric measurements.
PSF photometry as well as Aperture
As function of fake magnitudes or even astrometric centroid distance to true centroid.



Synthetic sources: HSC example

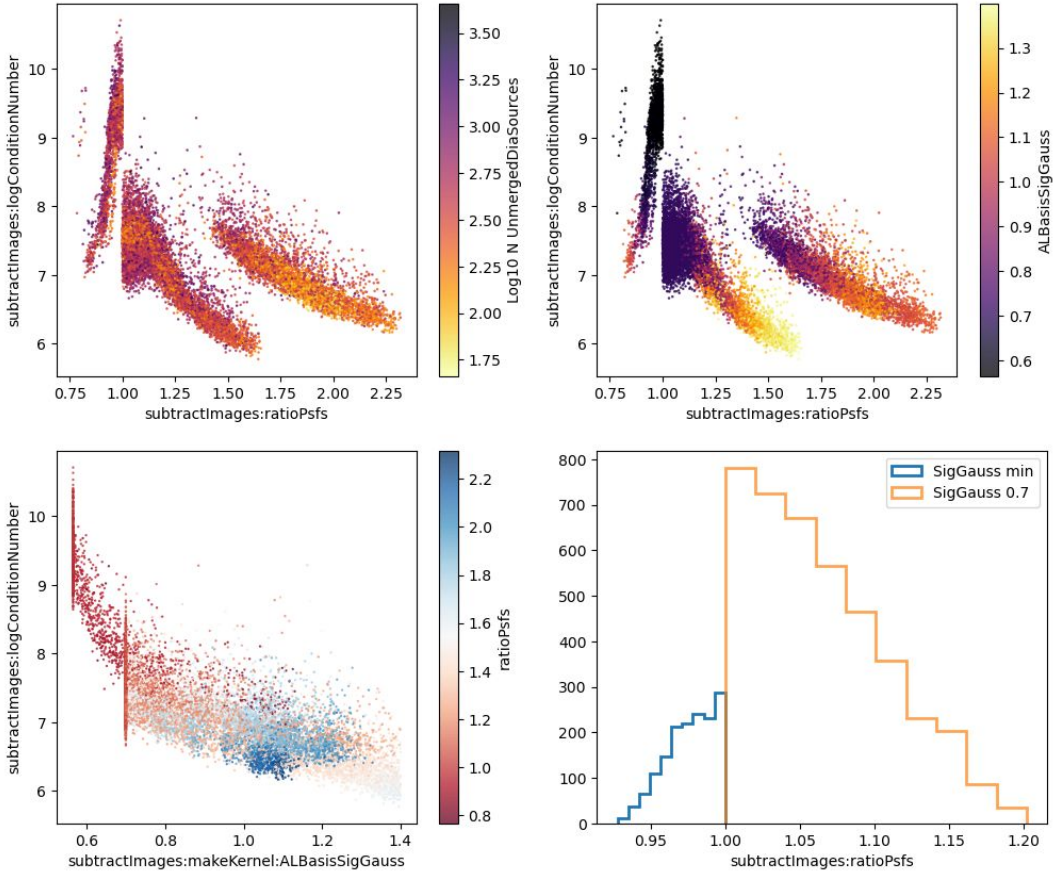


Astrometric measurement of our Fakes position. In pixels, sky coordinates, and in combination with the brightness of the source

Inspection of the metadata

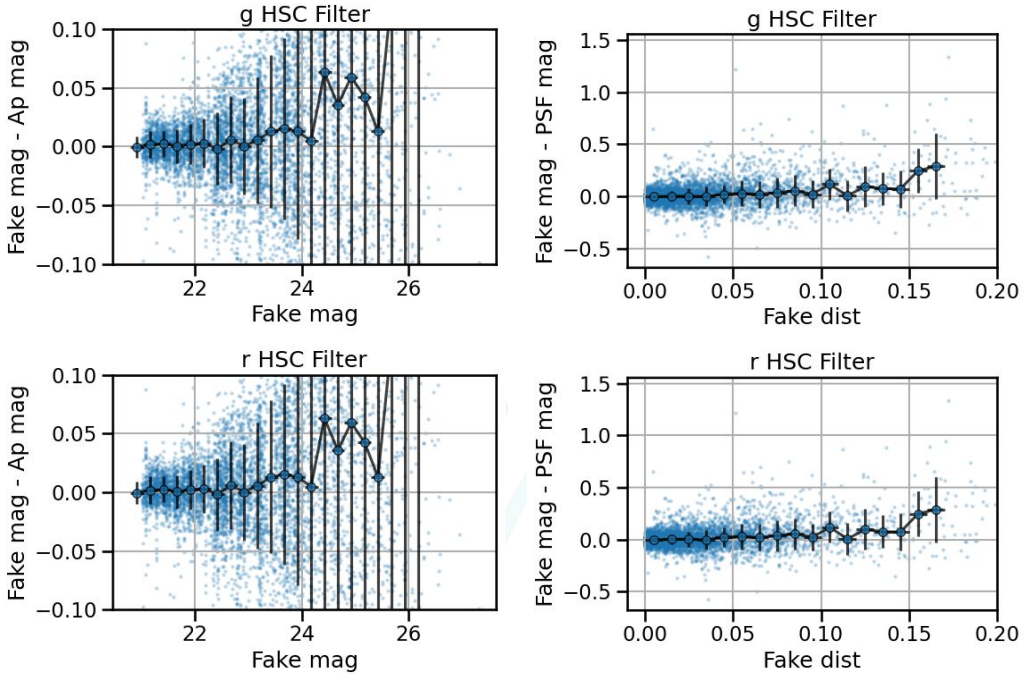
Metrics or metadata on the kernel is crucial for understanding our performance!
We inspect the dependency of two key parameters with the ratio of PSF sizes:

- ratioPsfs = $PSF_{Science}$ / $PSF_{Template}$
- Condition number: the power ratio of kernel top coefficients
- SigGauss: the smallest size of the kernel basis components

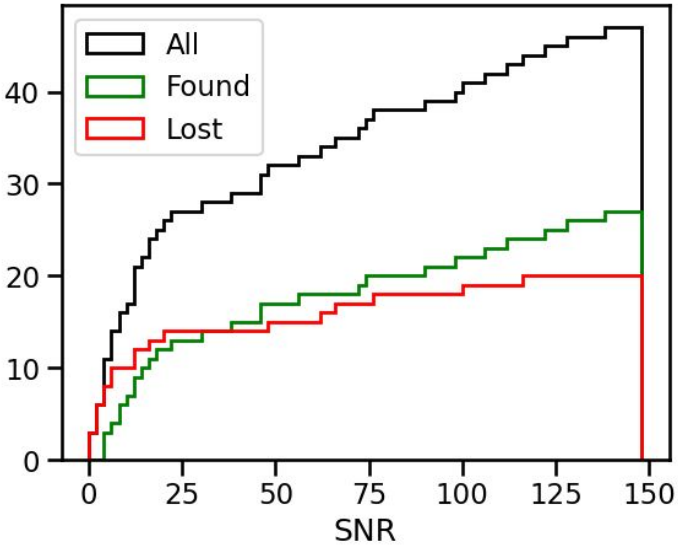


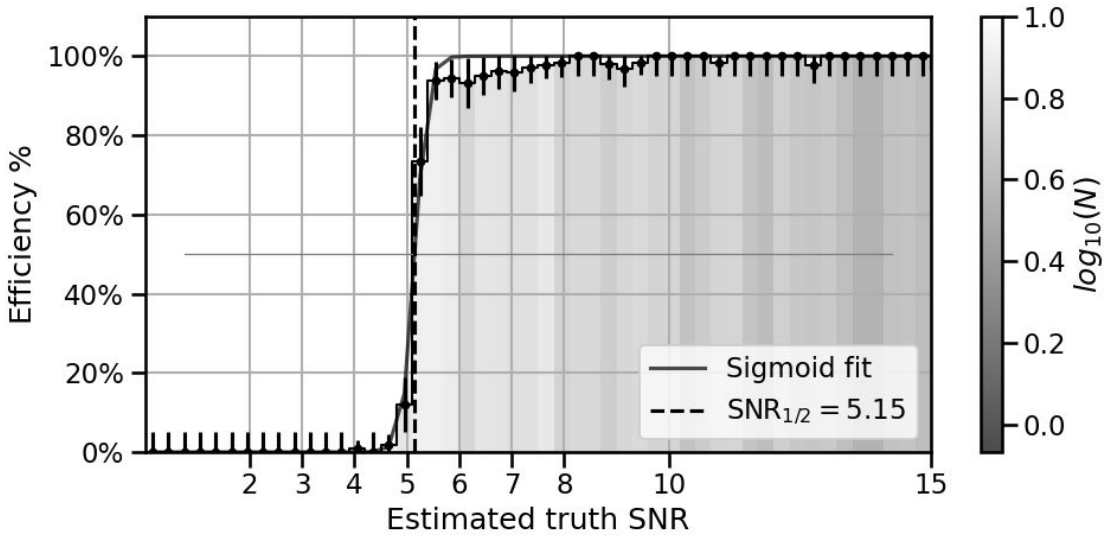
Synthetic sources: HSC example

With fakes we can also test measurement results, like for example photometric measurements.
PSF photometry as well as Aperture
As function of fake magnitudes or even astrometric centroid distance to true centroid.



HSC Processing of several images





Recovery fraction per S/N bin

lp_diffim Subtraction tasks

Things to keep in mind about broad perspective:

- The **difference image** subtraction task **is a link in a chain** inside a **very complex system**
-

lp_diffim Subtraction tasks

Things to keep in mind about broad perspective:

- The **difference image** subtraction task **is a link in a chain** inside a **very complex system**
- The main objectives of this system are always the detection of transients but **many more things depend on it**
  - The subtraction results yield products that are used for **measurements**
  - The **subtraction metadata** is also valuable piece of information – each implementation might handle this differently
-

lp_diffim Subtraction tasks

Things to keep in mind about broad perspective:

- The **difference image** subtraction task **is a link in a chain** inside a **very complex system**
- The main objectives of this system are always the detection of transients but **many more things depend on it**
  - The subtraction results yield products that are used for **measurements**
  - The **subtraction metadata** is also valuable piece of information – each implementation might handle this differently
- The core algorithm is to find a transformation kernel that is best suited for finding pixel flux changes, but implementation has to **deal** with many more **things such as masks** – this means **extra development** (such as Zogy case)