What Rubin Saw Feb. 27Th

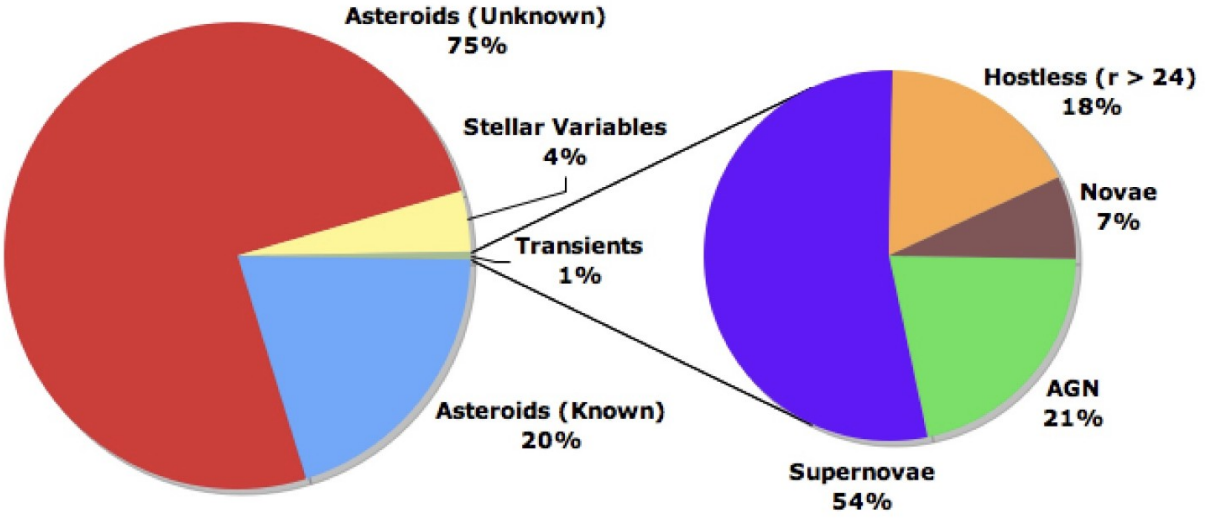
A presentation about Data Quality,
Flying rocks, and hopefully (?) transients

Work in Progress

Emmanuel Gangler - FINK@World 2026

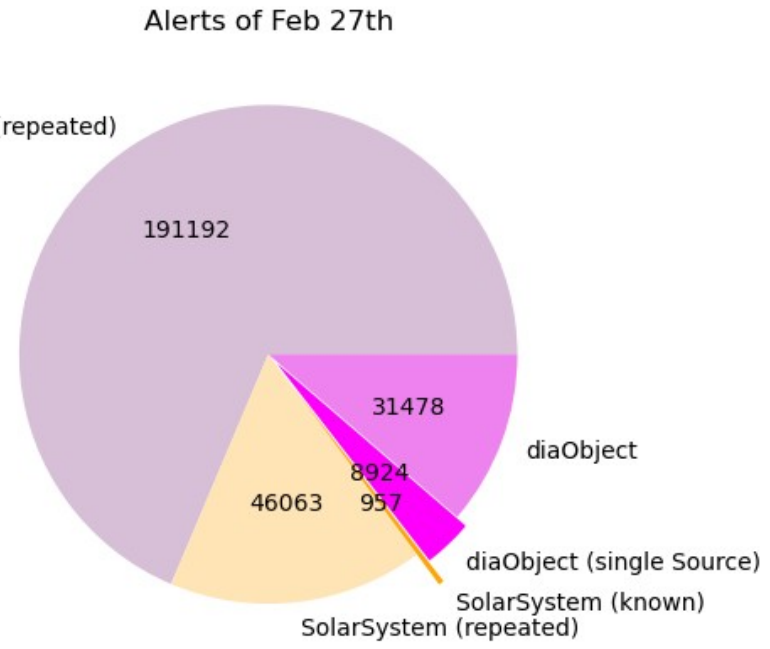
Driver of the talk :

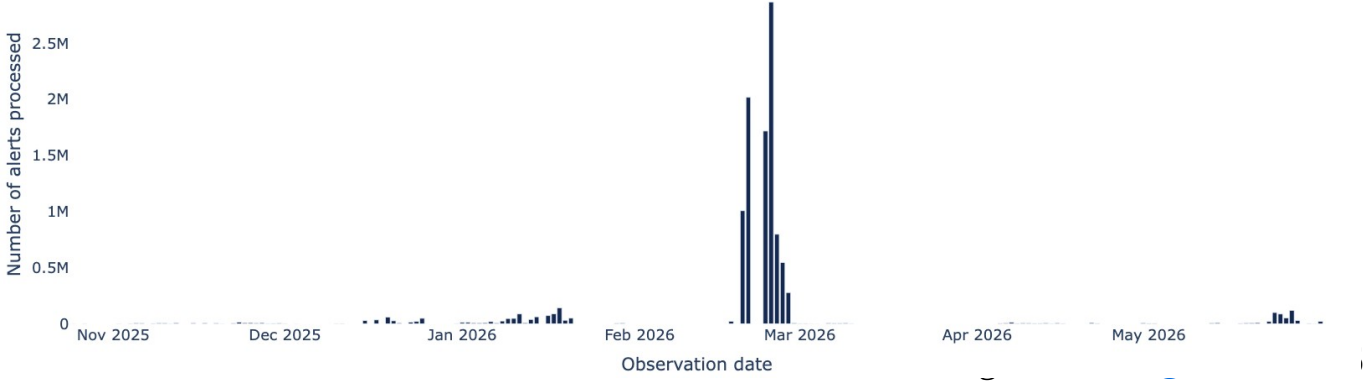
- Can we redo this graph from LSST science book (2009) ?



Data Selection : Feb 27th

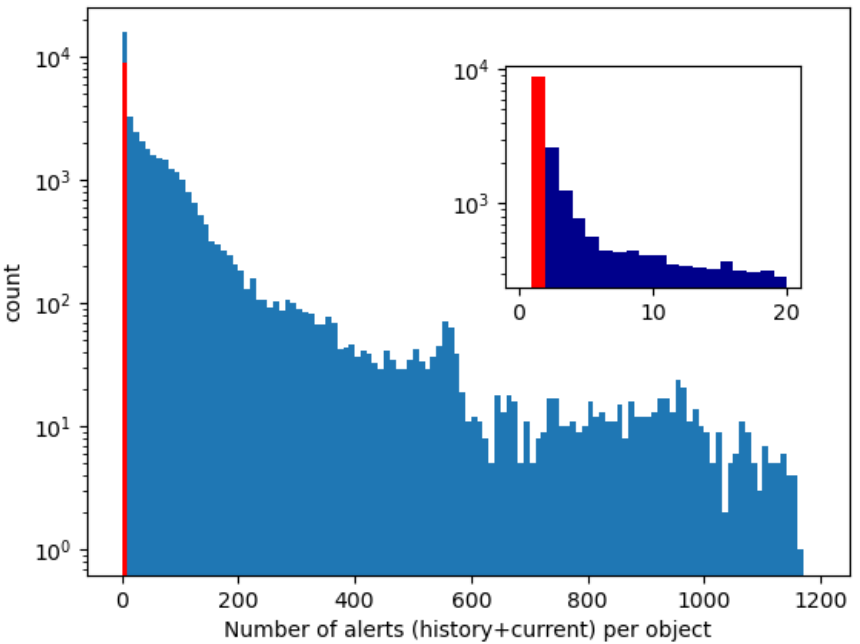
- 278614 alerts to download (26G)
  - 85 % are repetitions of same object
  - → Keep only last alert of each object
- Known Issue :
  - Some previous alerts not in alert history…





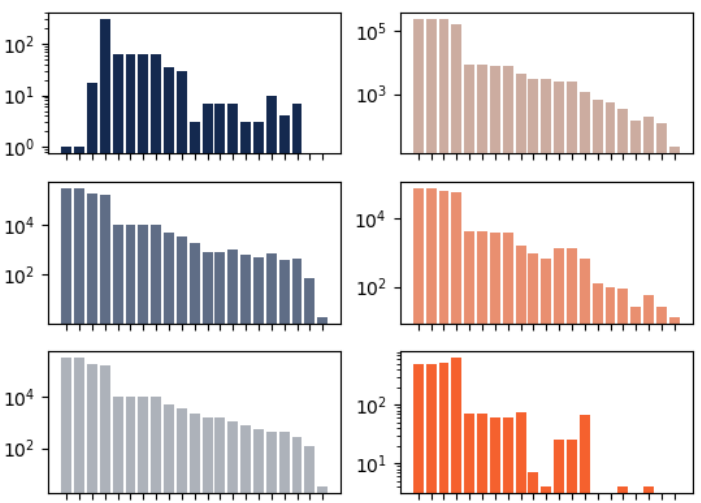
What now ?

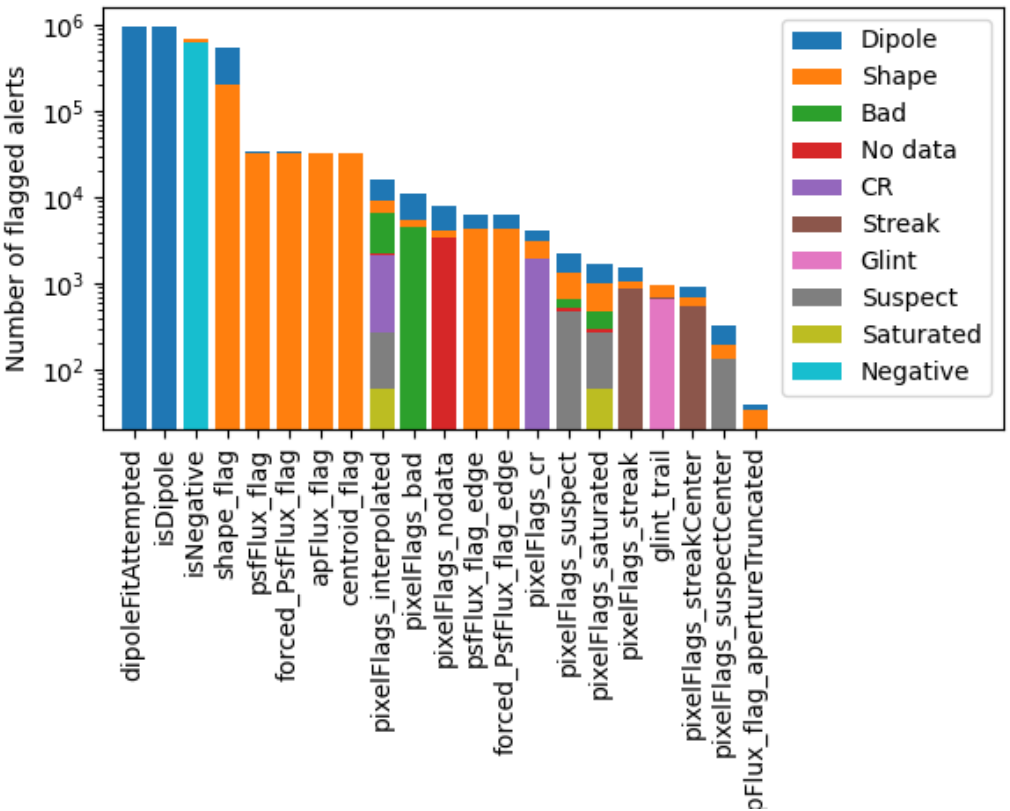
- 2,8 Million alerts for 40 k objects
- Alert-wise :
  - Study quality => Flags
  - Which photometry ?
- Object-wise :
  - Light-curves quality
  - Classify per object



diaSource Flags

- 38 flags | booleans
  - 17 are never set
  - 219 unique combinations





diaSource Flags

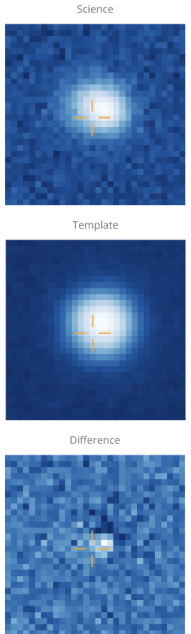
Subtraction

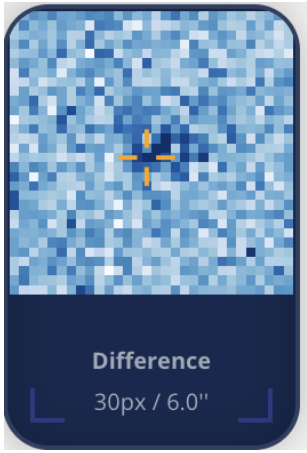
Dipole

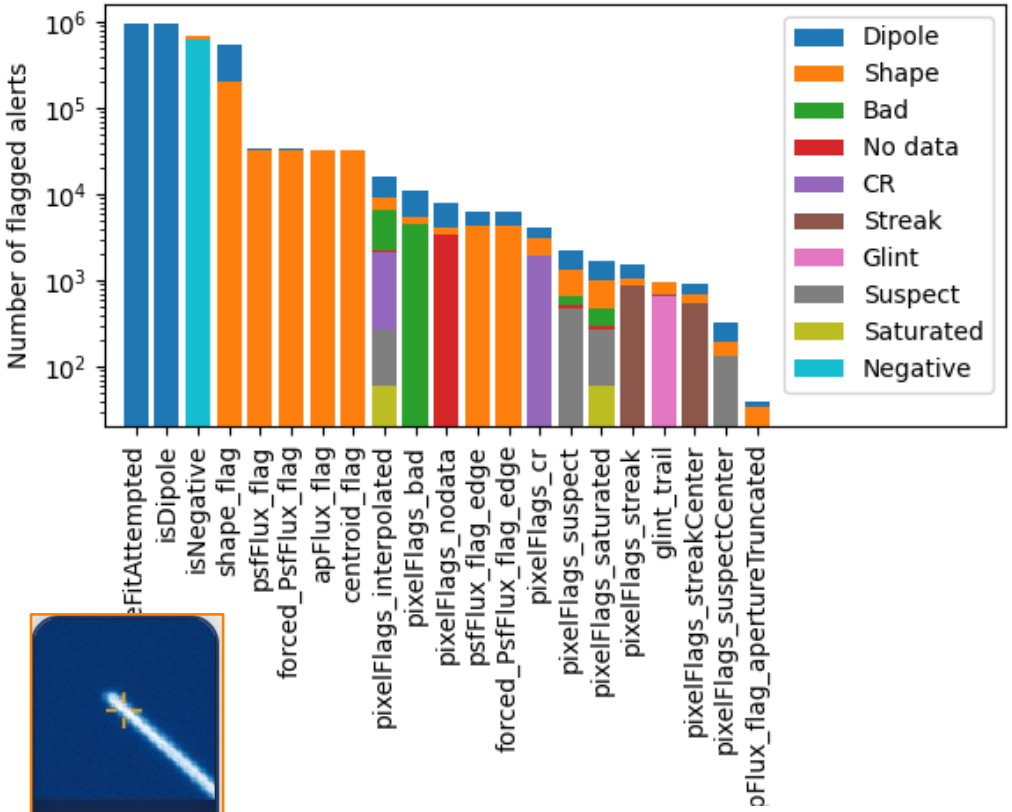
Shape

Negative

Negative







Another bad shape

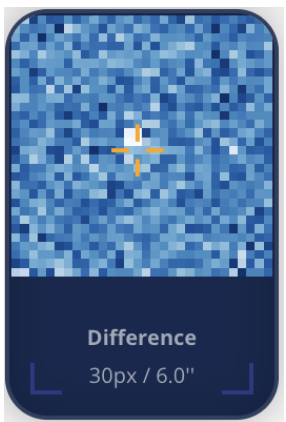
diaSource Flags

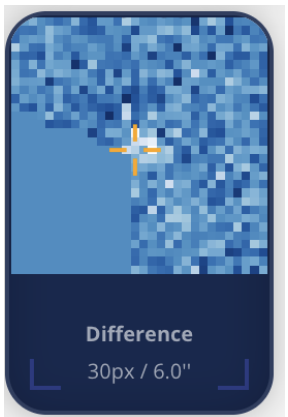
Artefacts

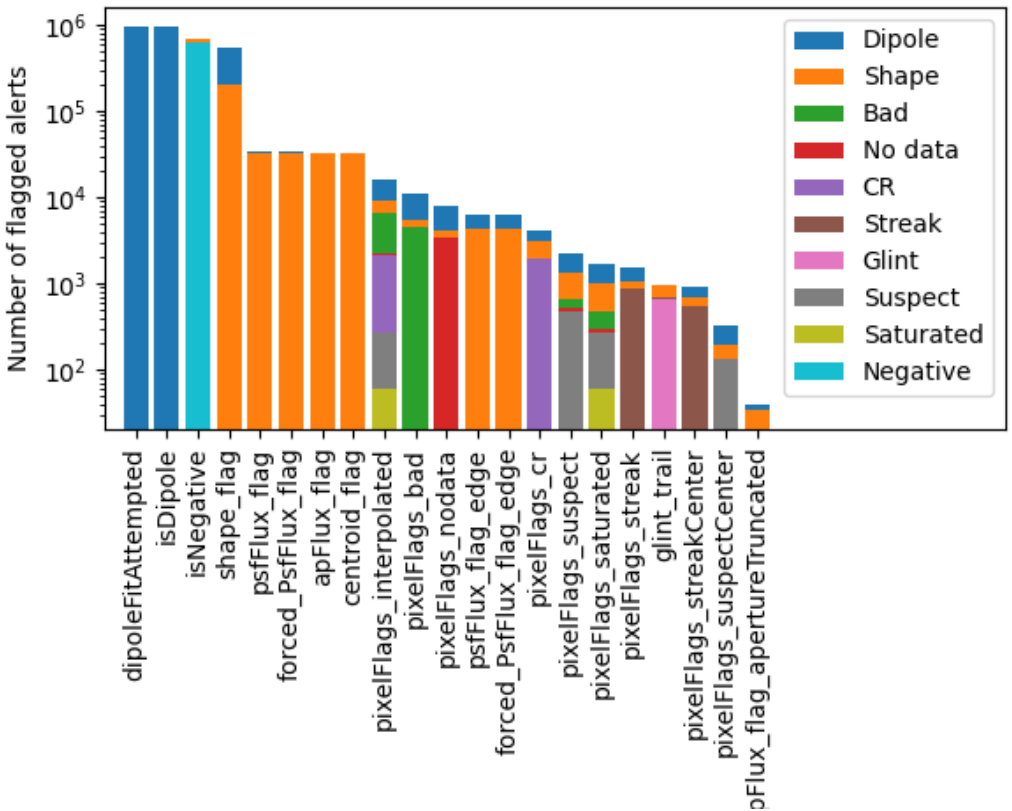
Bad

No data

Cosmic Rays







diaSource Flags

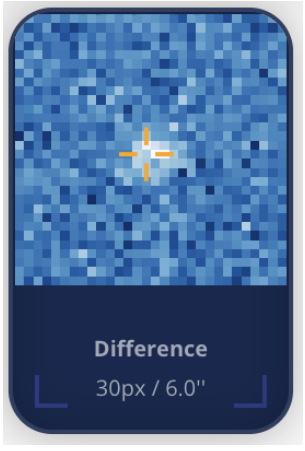
☒ Artefacts

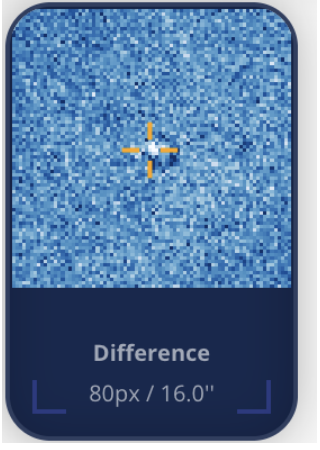
Streak

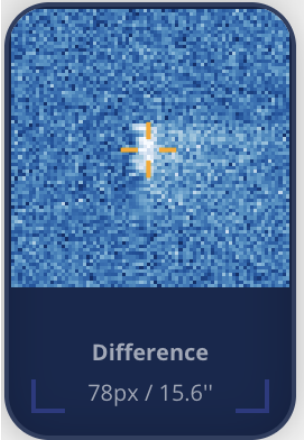
Glint

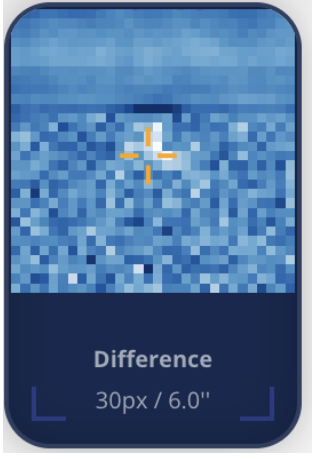
Suspect

Saturated

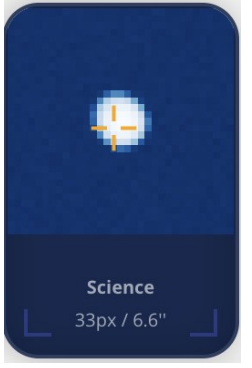








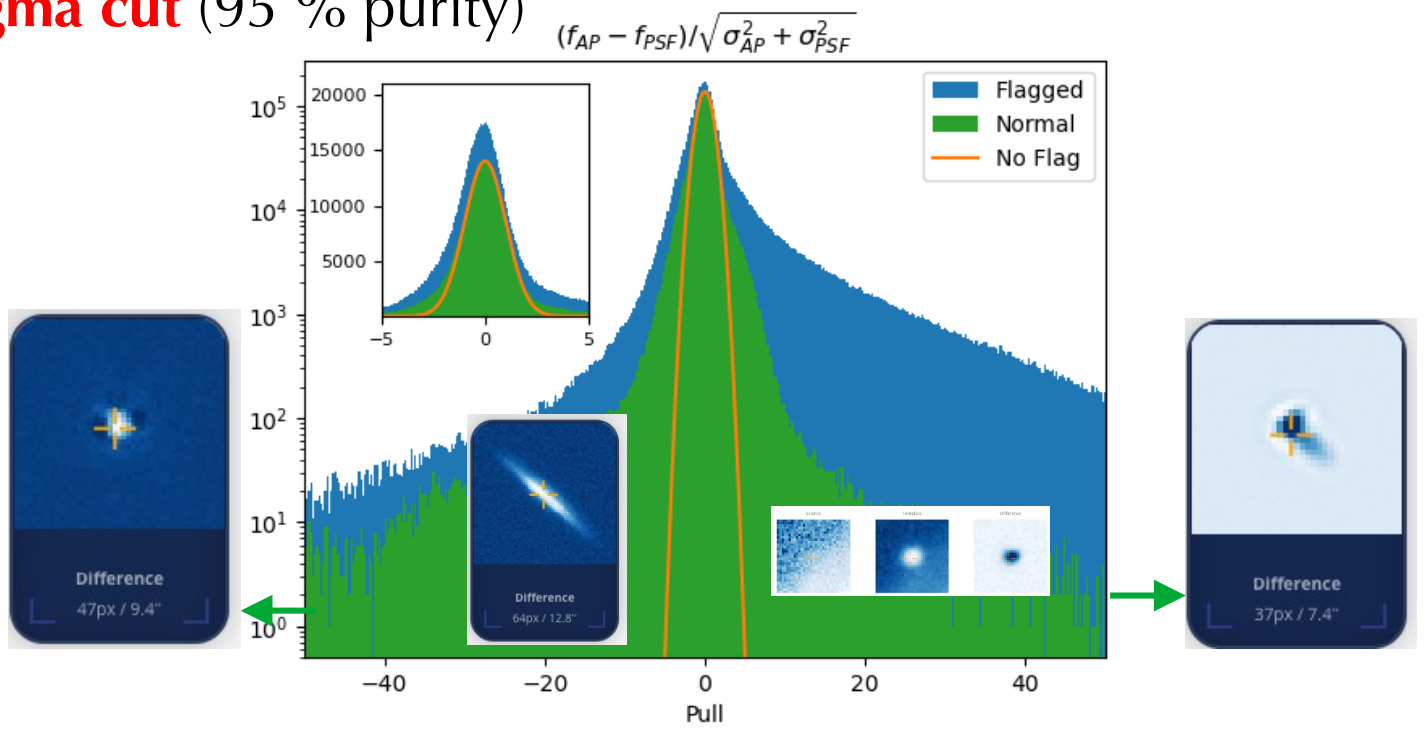
Not flagged !



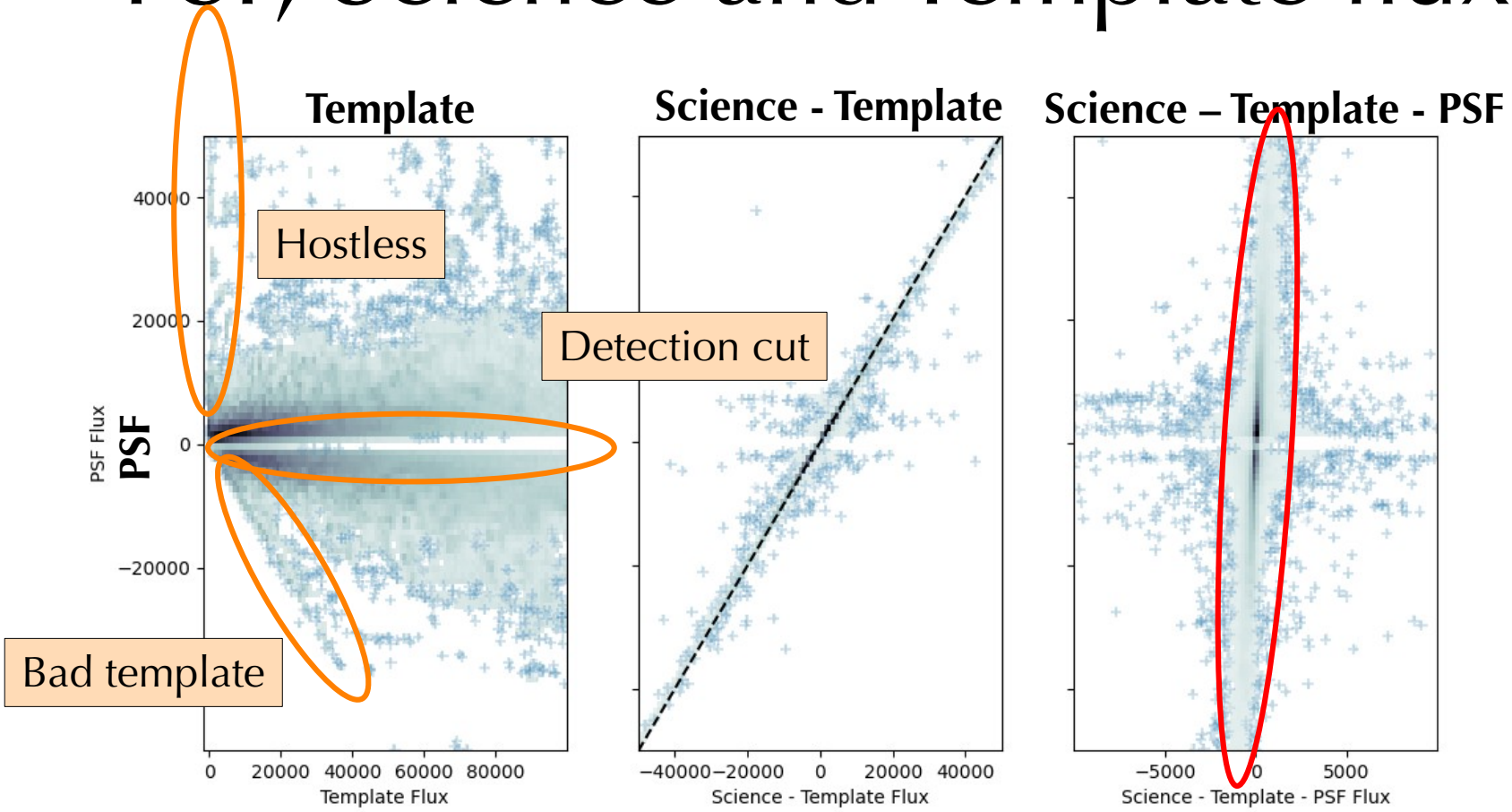
Aperture flux vs PSF flux

- AP and PSF should be compatible
- Large deviations : issue !

  → **2.5 sigma cut** (95 % purity)

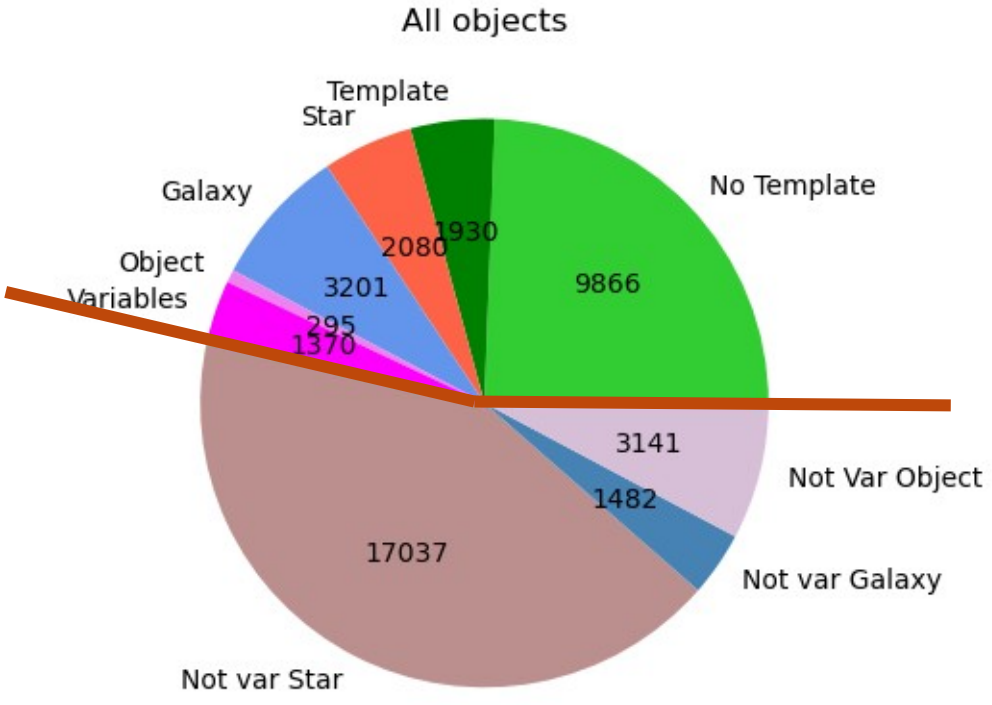


PSF, Science and Template flux



Quality of Light-Curves

- **No Template**
  - Tempalte SNR<5
- **A Template**
  - Unknown object
- **Star** : dr8 prob >0.8
- **Galaxy** : dr8 prob <0.2
- **Object** : others
  - dr8 or Gaia reference
- **Not variable**
  - Gaia object + Var flag = 0
- **Variable**
  - Gaia object + Var flag=1

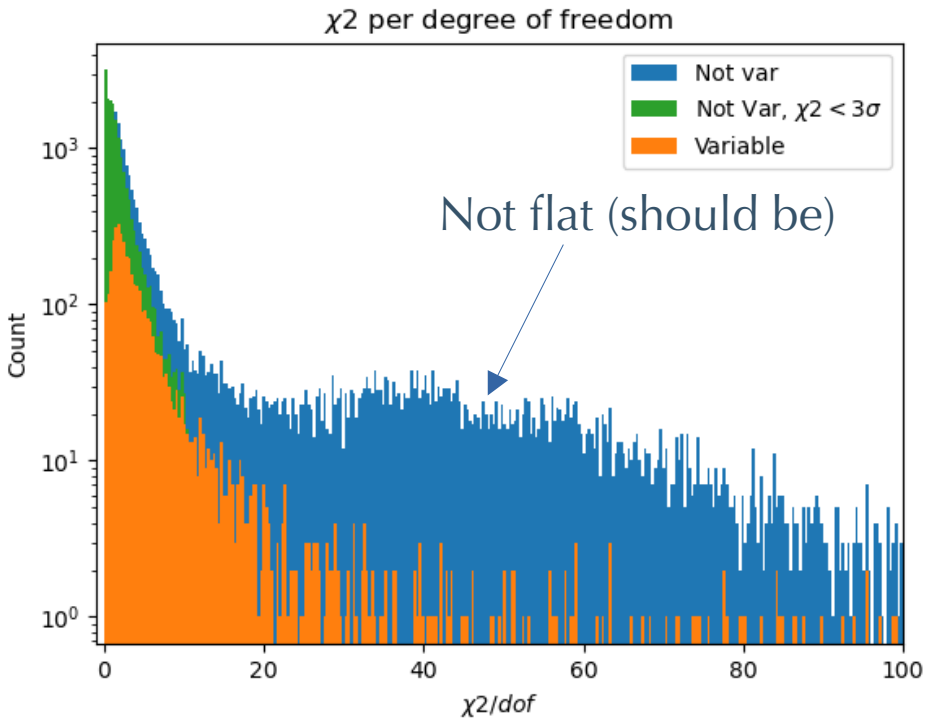


Why so many « flat » objects ?

Can we trust error bars ?

→ NO !

(37 % of non Var are not flat)

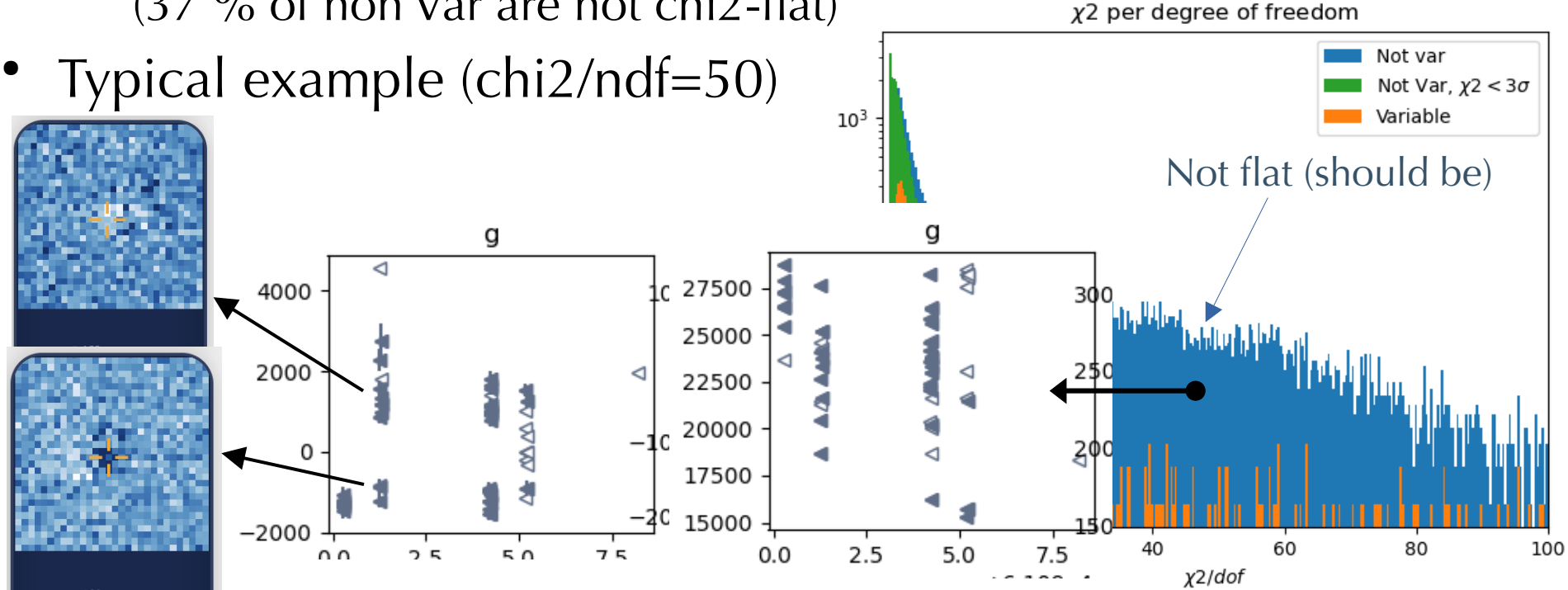


Can we trust error bars ?

→ **Not really...**

(37 % of non Var are not chi2-flat)

- Typical example (chi2/ndf=50)



PSF **Template**

How to tag these Light-Curves ?

- 1) PSF diff flux ~≤ RMS(Template) => bad quality
- 2) Many bad alerts
- 3) Fluctuations are random

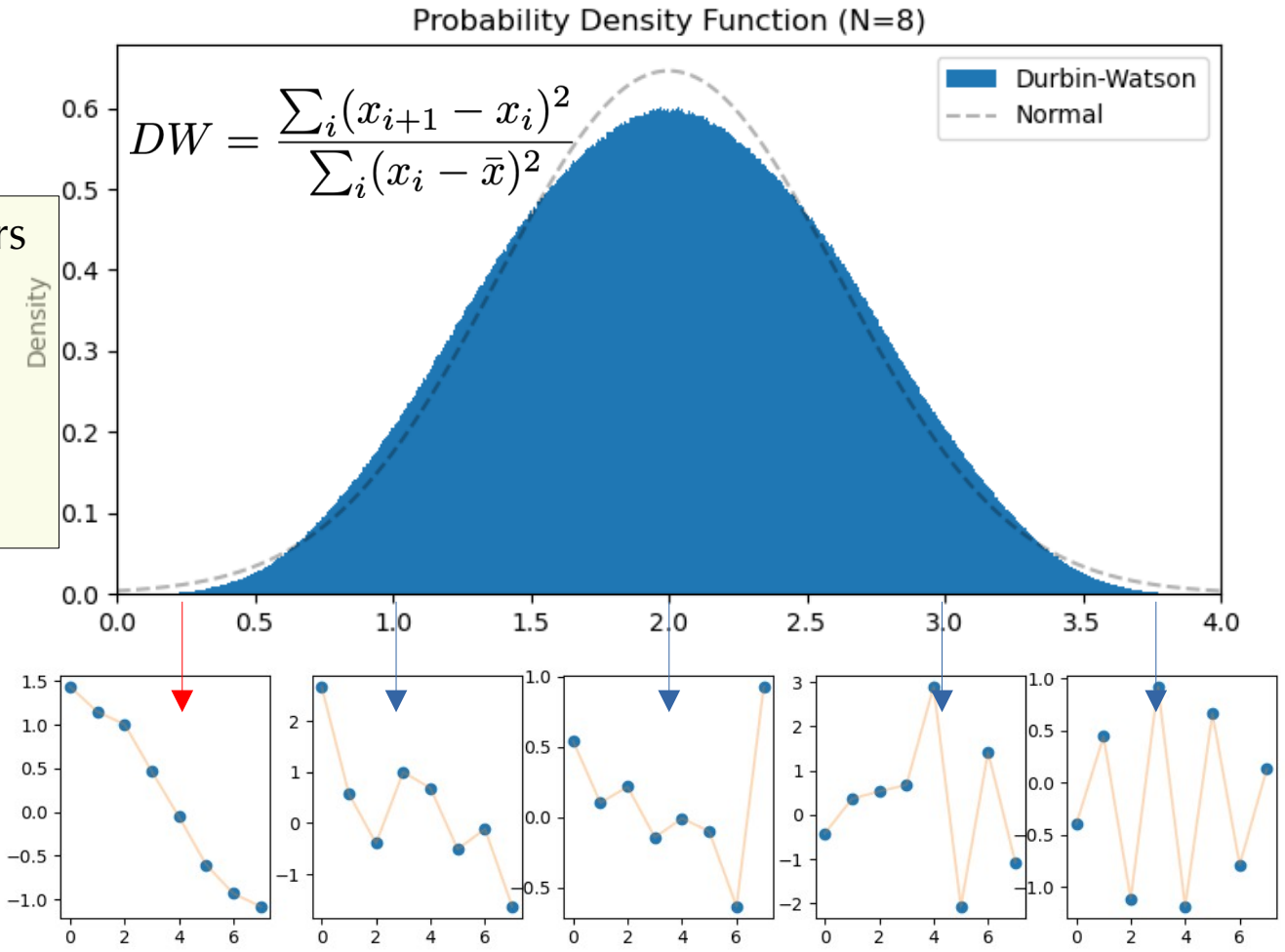
  → Statistical test of flatness when errors are unknown

The Durbin-Watson test

Compare variance between neigbours to sample variance estimator

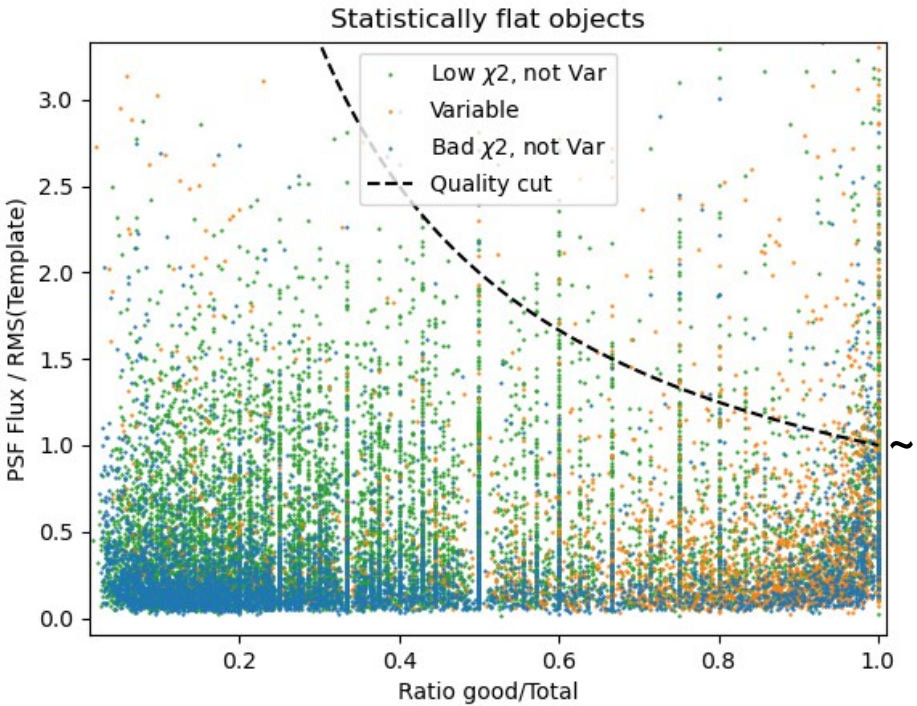
a.k.a : test autocorrelation at t+1

P-value are not trivial to compute...



Quality cut :

- Statistically « flat » → apply cut
  - Based on visual inspection
  - … and intituition …
    → Room for improvement
- Strong cut
  - Removes variables with small fluctuations
  - Keeps flat LC with offset



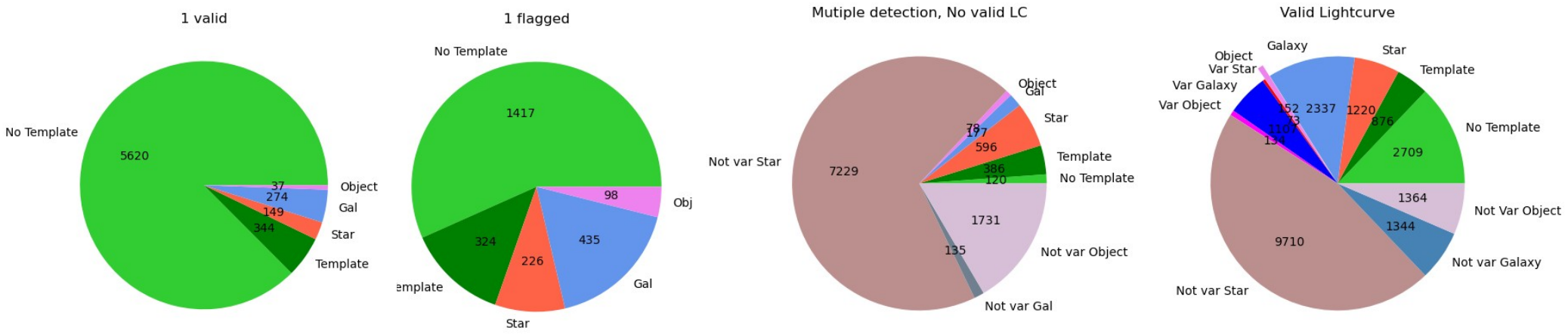
~0.08 mag

And now : Scan !

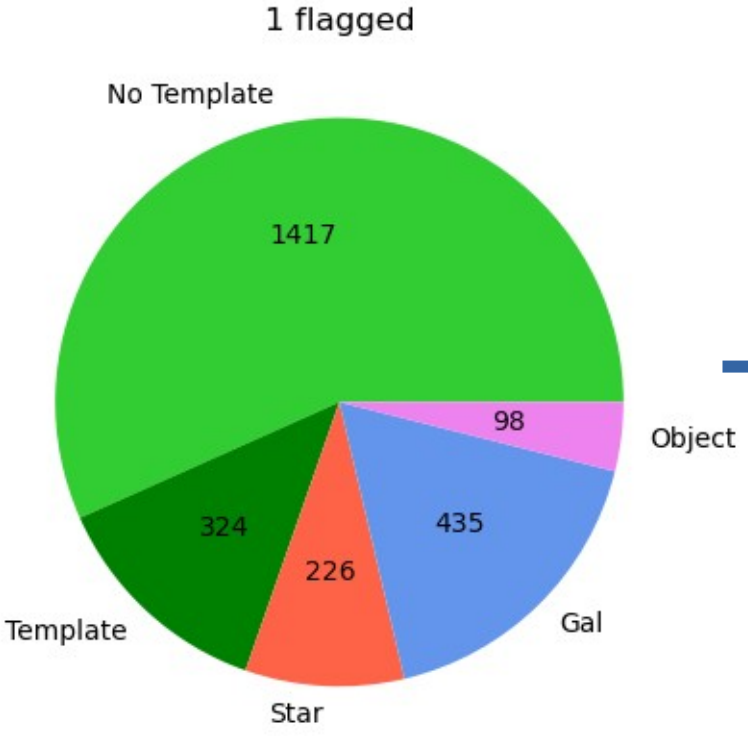
- Divide objects in subcategories (including chi2, flatness, slope, quality...)
- Inspect and Label (923 scanned)
- Extrapolate to the category

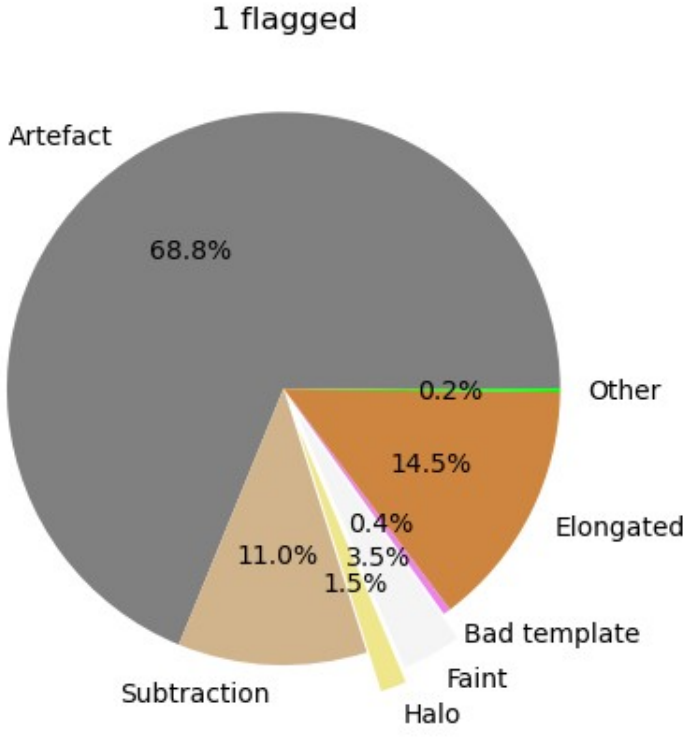
	Chi2 Low	Chi2 High
Flat, bad Q	6505	4330
Flat, ok Q	1601	95
Variable	592	2956

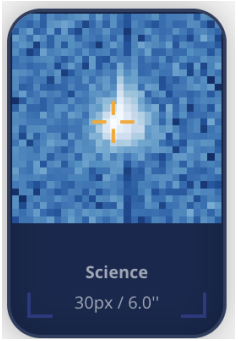
+

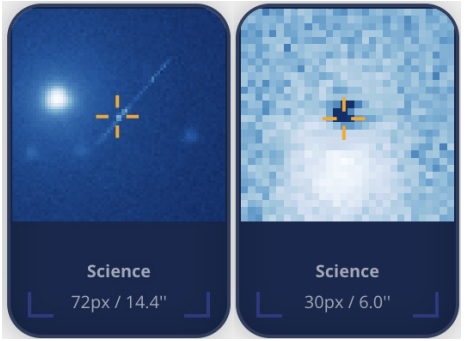


Single Source with Flag

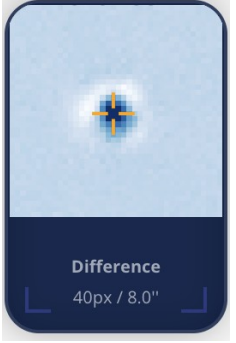




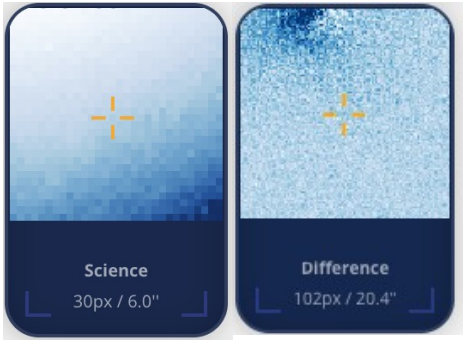




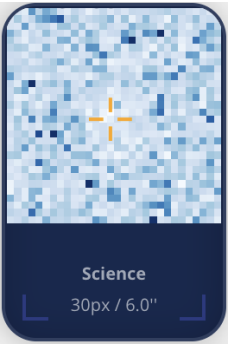
Artefact



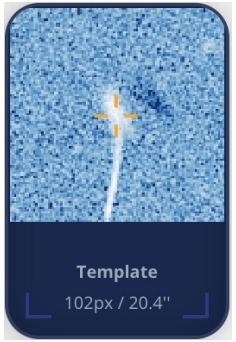
Subtraction



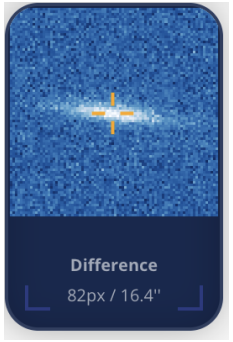
Halo



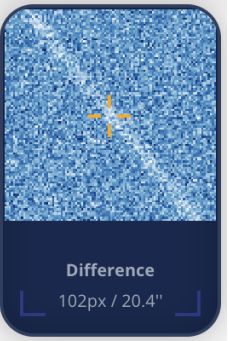
Faint

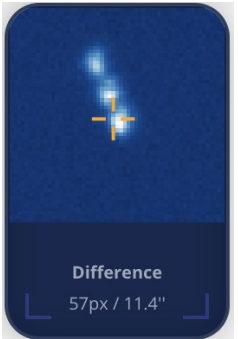


Bad template



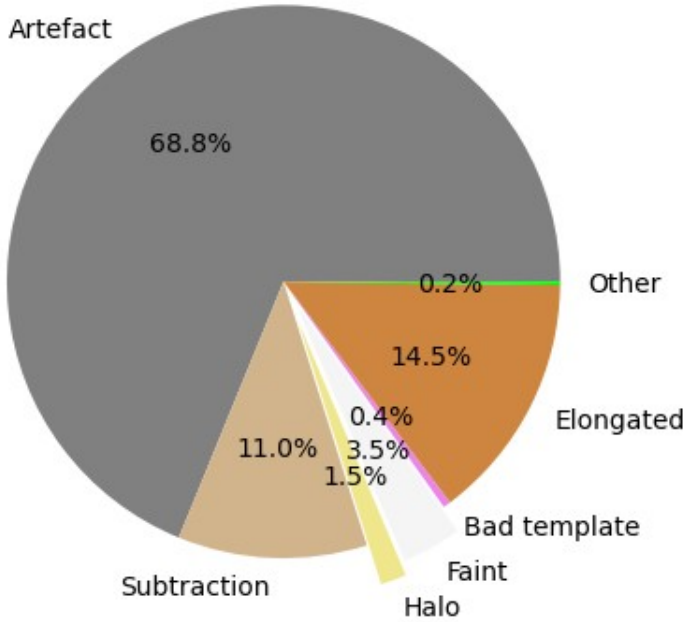
Elongated



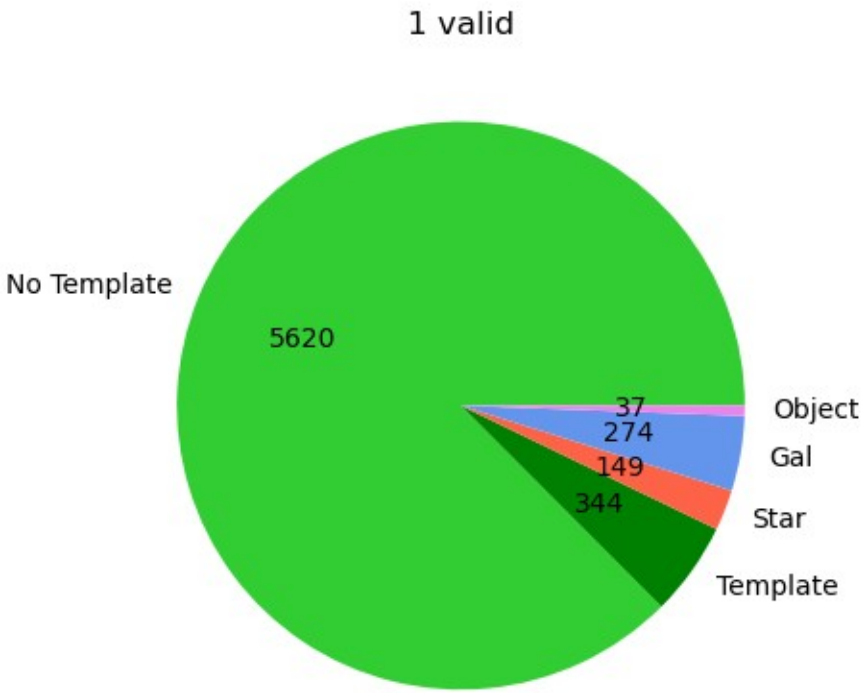


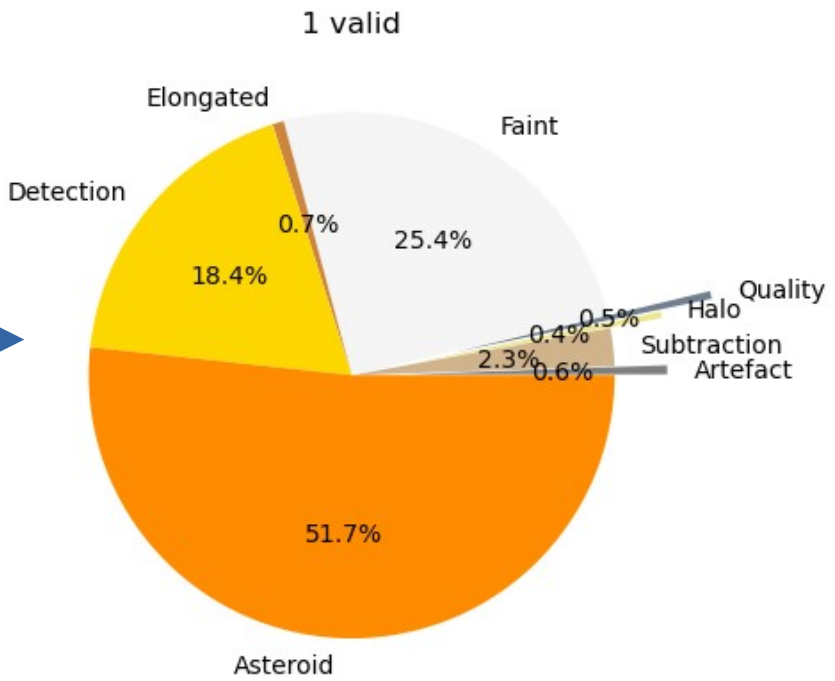
Others

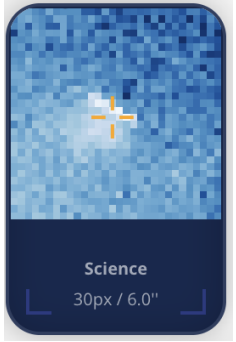
1 flagged

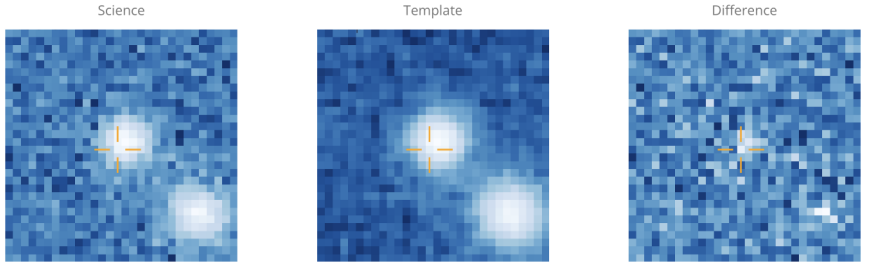


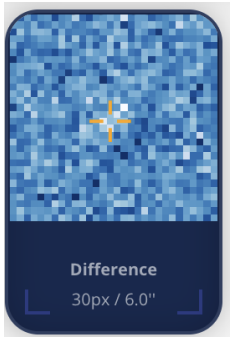
Single detection (valid)





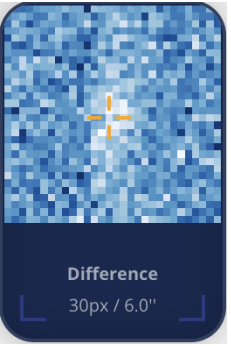


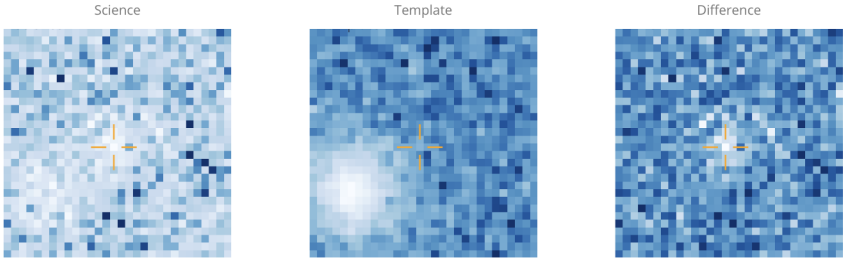




Artefact

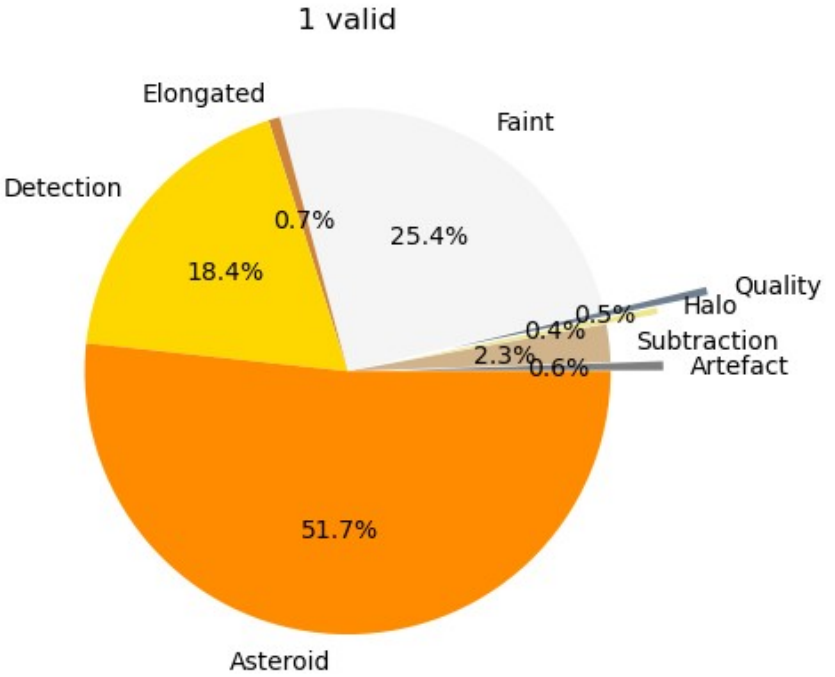
Subtraction

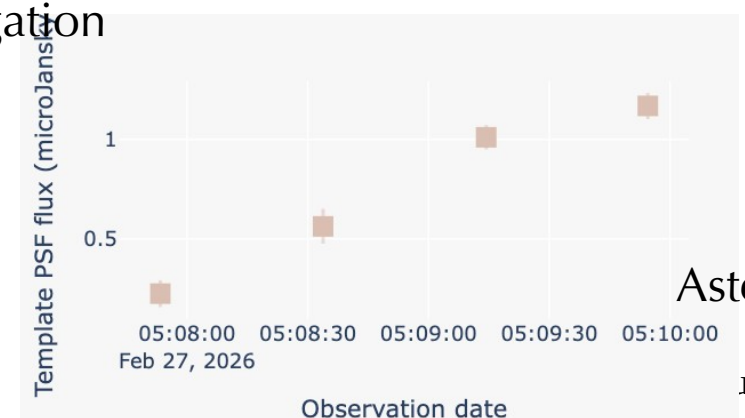




Detection

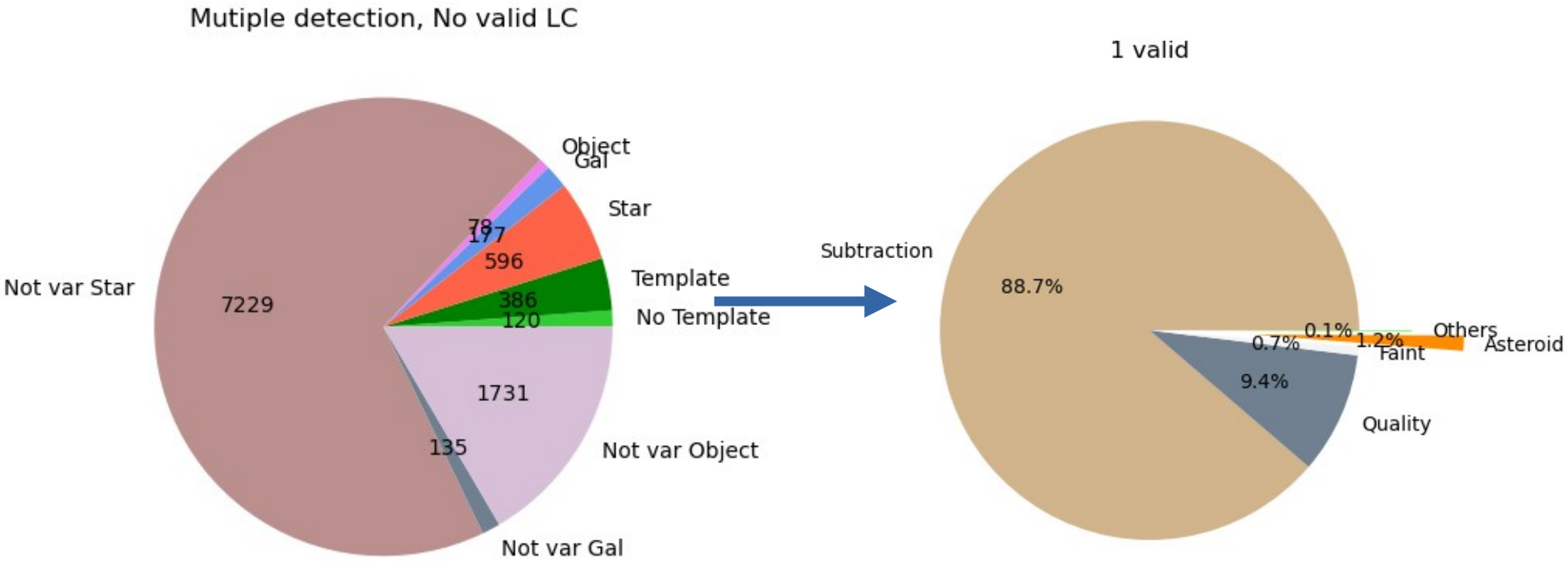
Elongation

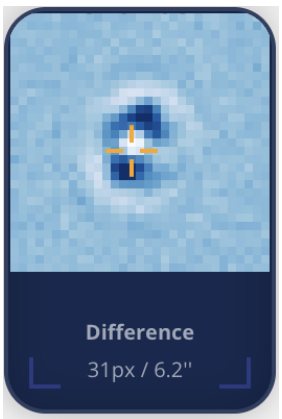




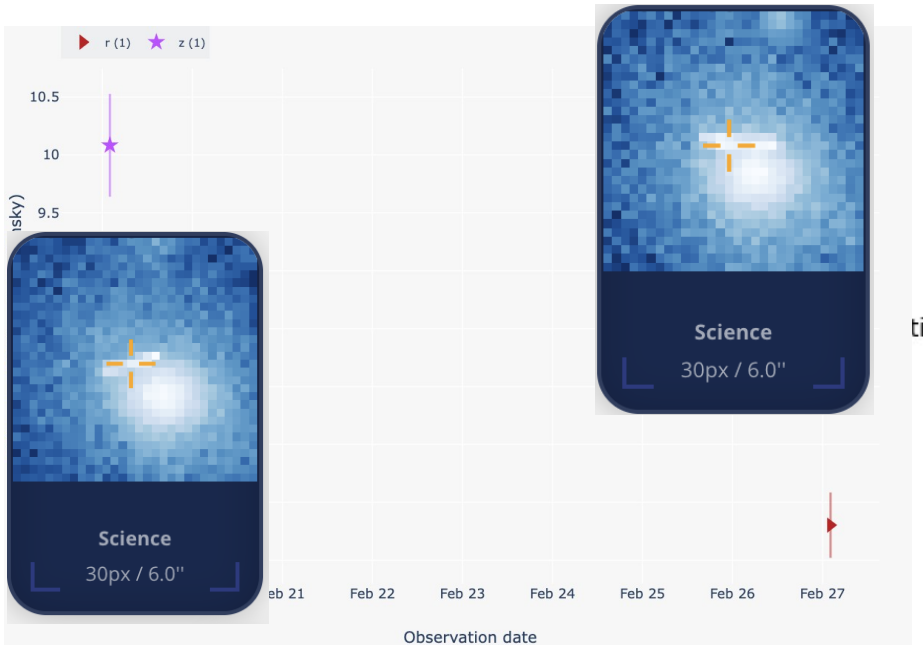
Asteroid

Multiple detection, no LC

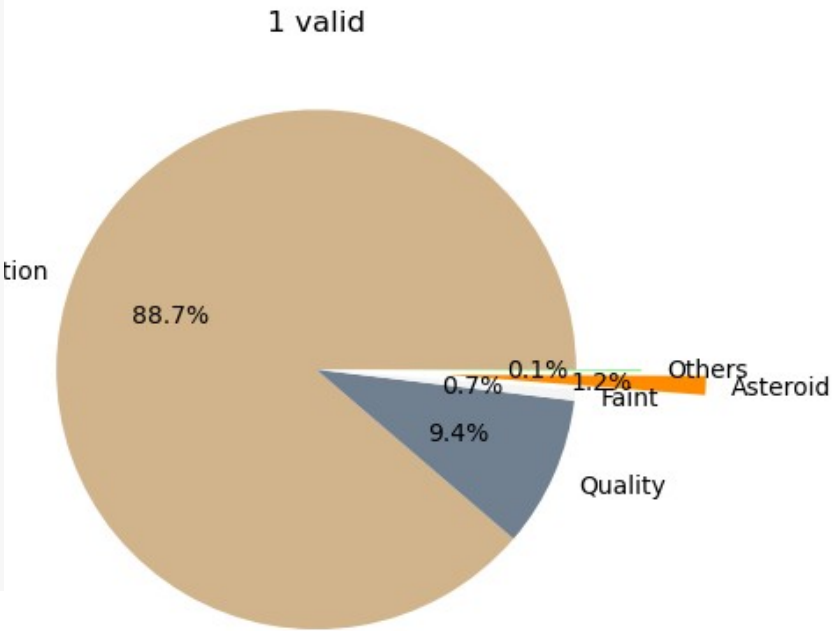




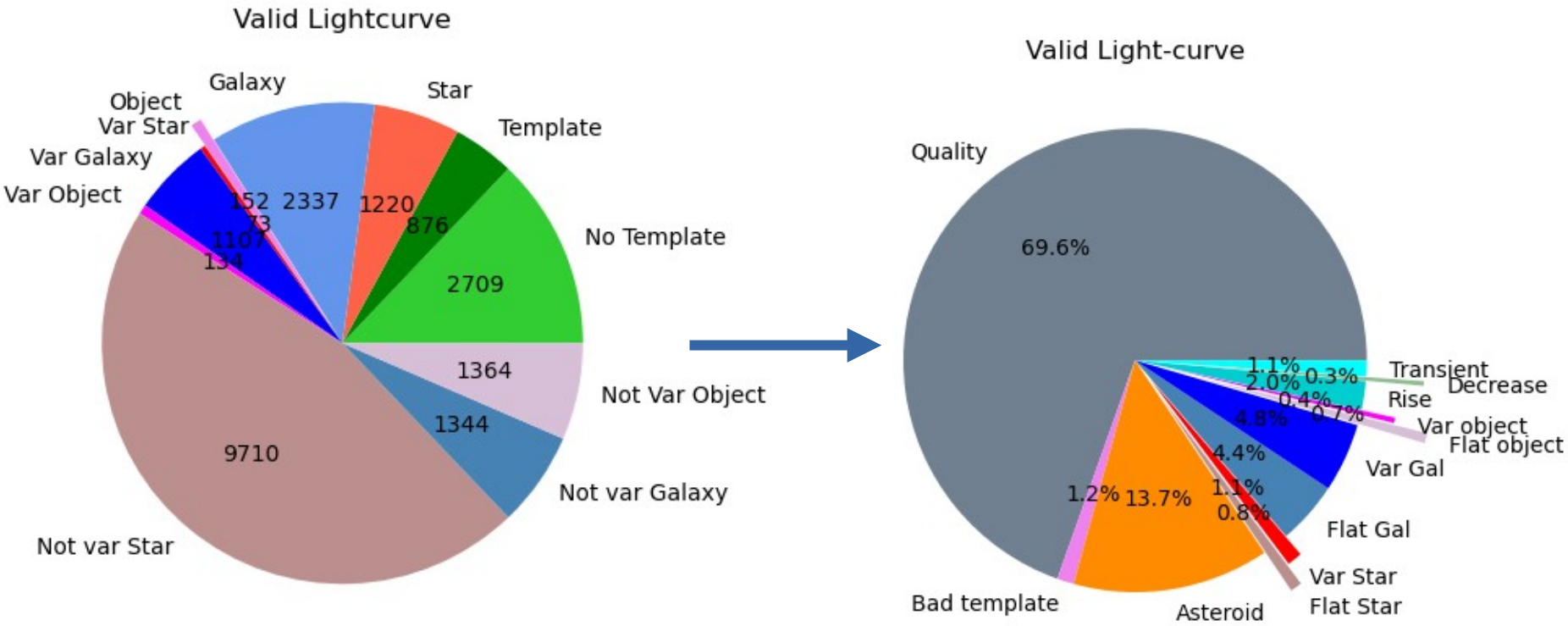
Subtraction

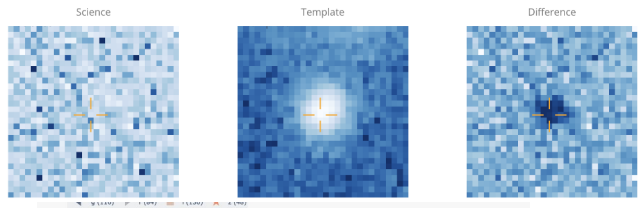


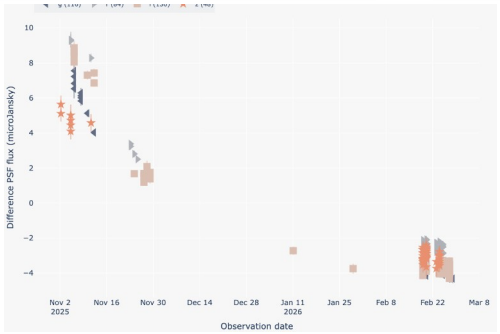
Double cosmic-ray !



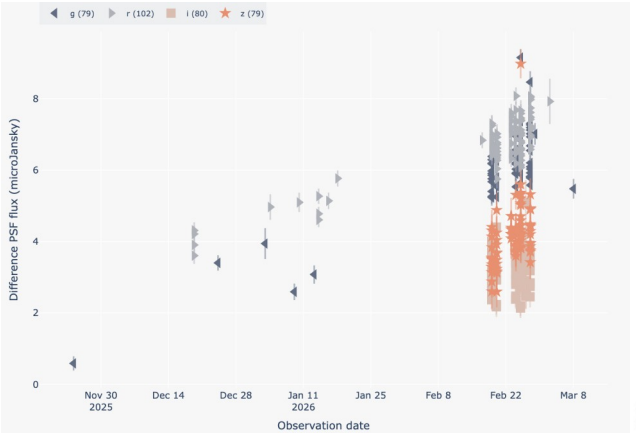
Finally : there is a LC !



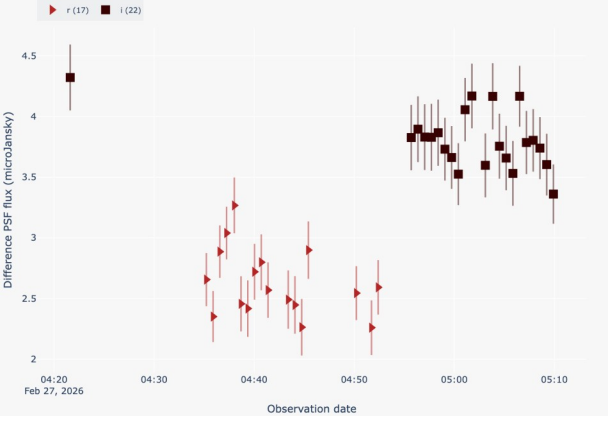




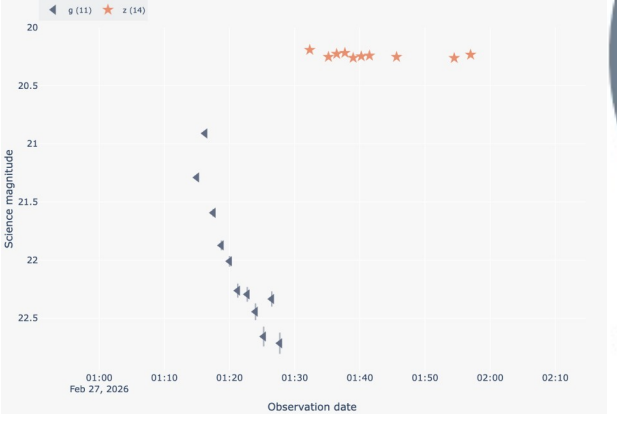
Transient in template



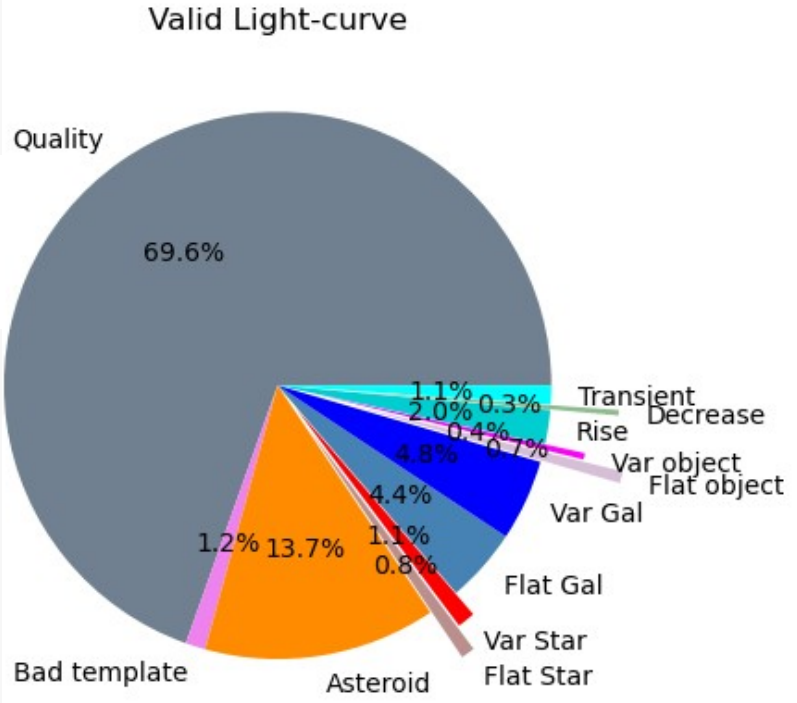
Rise

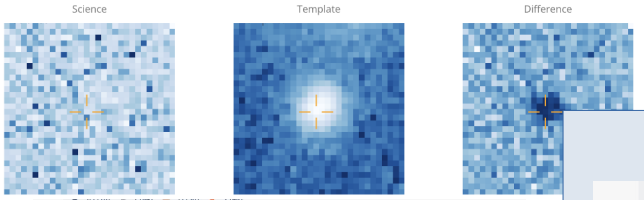


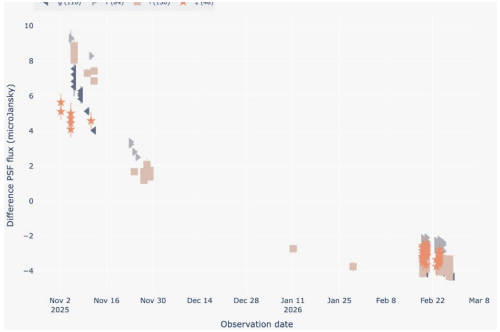
Asteroid



Var star (flare?)

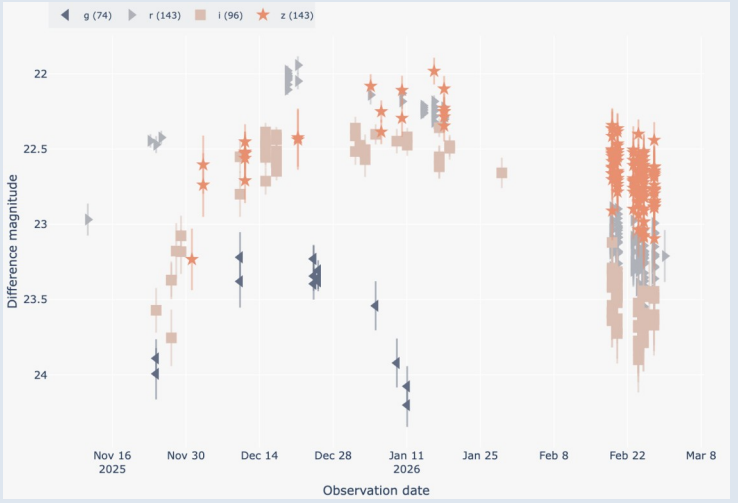




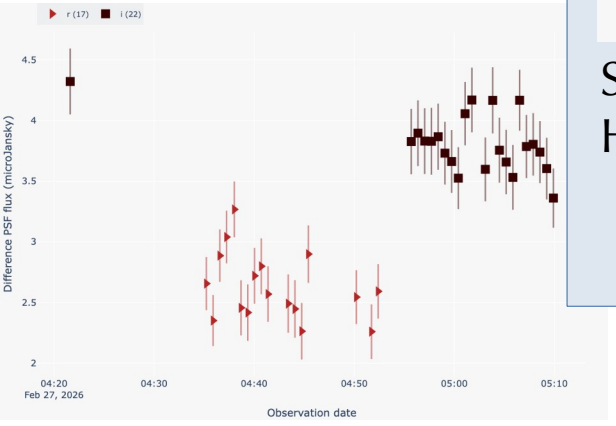


Transient in template

Transient candidate

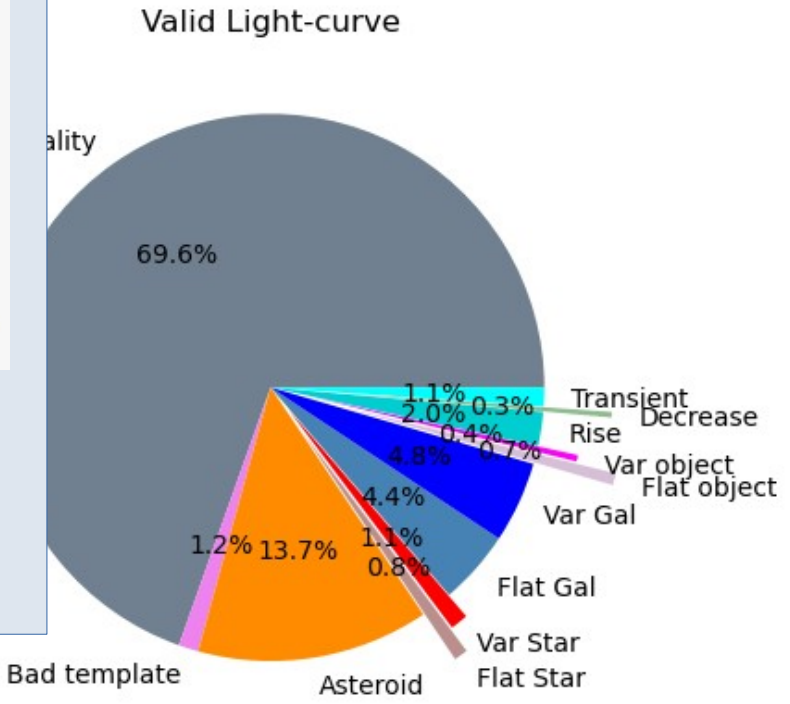


Star probability 91 %
Host photo-z : 0.475
→ SLSN ?

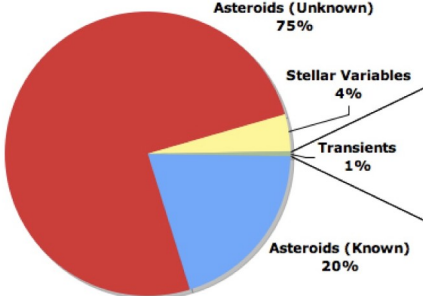


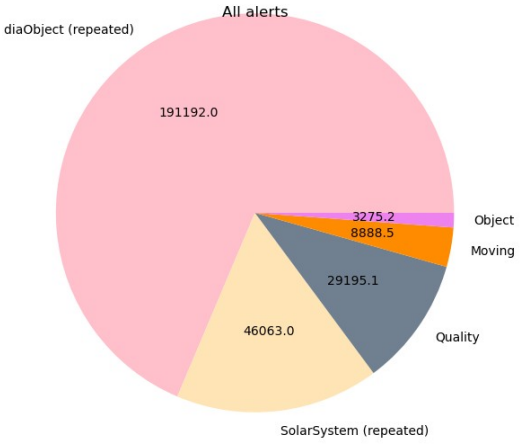
Asteroid

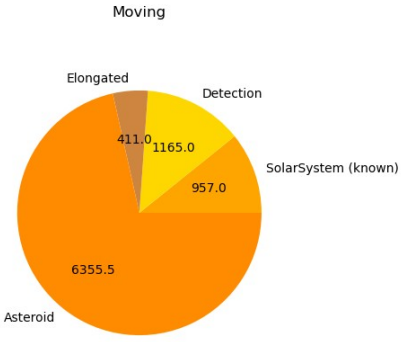
Var star (flare?)

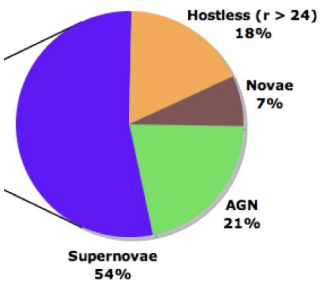


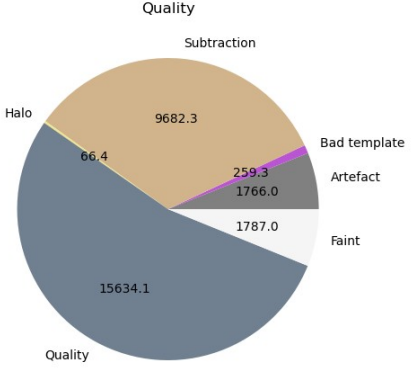
Mission accomplished !

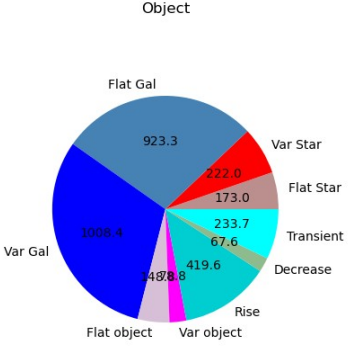












Conclusion and take away

- Dealing with « only » 280 k alerts is already a lot
- **Data quality is non trivial !**
  - Delivered subtraction can be improved
  - Home-made recipee for the moment : room for improvement
    - Better recipee
    - Add image stamp analysis (unseen dipoles or bad subtraction…)
    - Study band correlation as an indicator
- **Durbin-Watson test is powerful** (there is also a slope test)

  → in production ?
- **923 human-labeled objects** : available for further trainings
- **19 categories of alerts**

  → can be used to pre-tag alerts

PSF, Science and Template flux