The iPTF IPAC Pipelines: what works and what doesn't (*optimally*)

Frank Masci & the iPTF / ZTF Team

ZTF-Photometry Workshop, September 2015

http://web.ipac.caltech.edu/staff/fmasci/home/miscscience/masci_ZTFmeeting_Sep2015.pdf

Data path / pipeline summary

- Data flows through multiple pipelines, creating a variety of science products tailored for different purposes. These pipelines run asynchronously on different timescales.

- **Photometric (or frame processing) pipeline:** daily (end-of-night) processing to produce high quality instrumentally-calibrated images and source catalogs

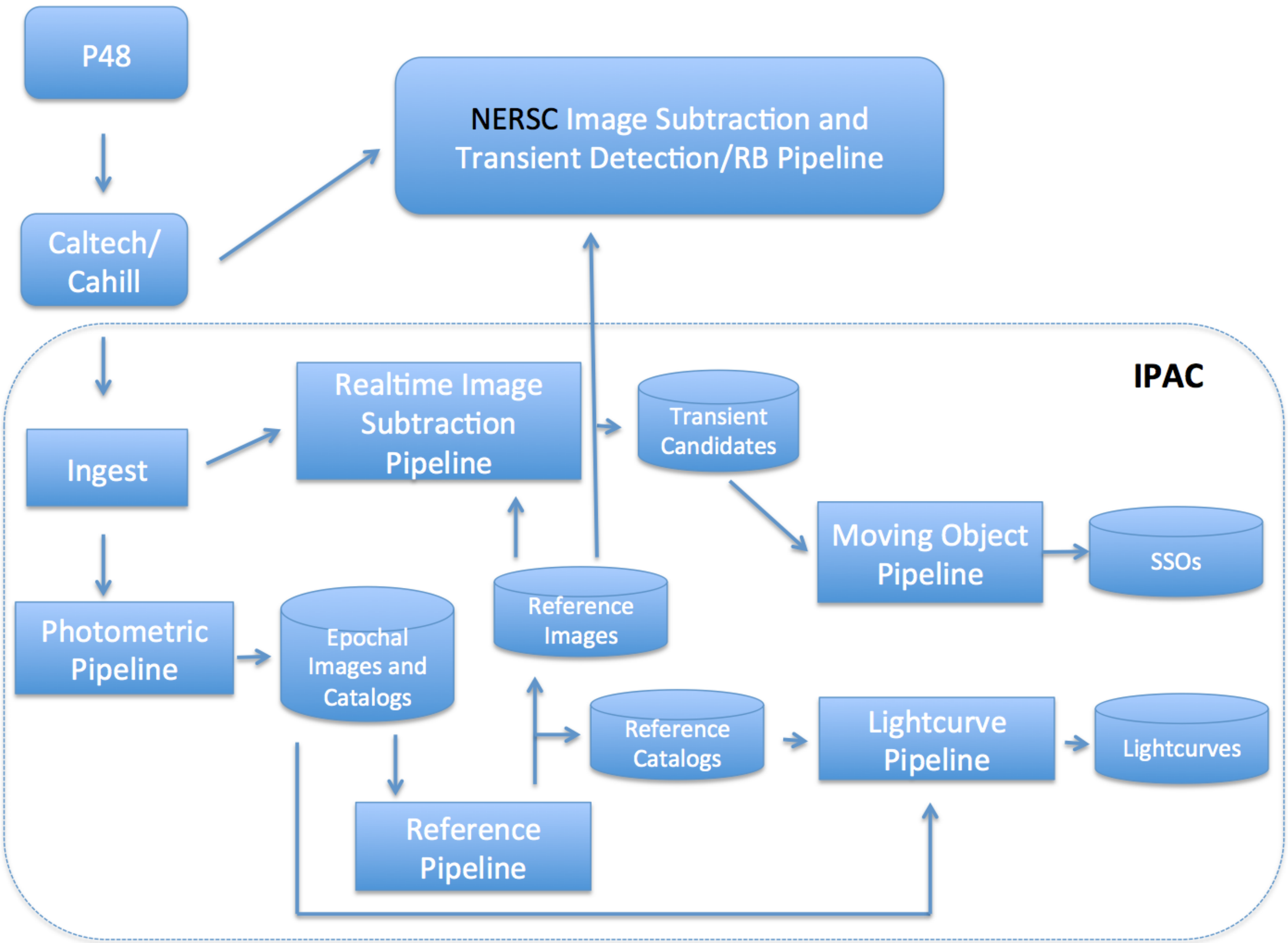
- **Reference image pipeline:** combines high quality frames into deeper images (coadds) – products are used in the real-time and lightcurve pipelines. Reference images are periodically made, depending on availability of good data for a given field/chip (more later).

- **Lightcurve (or relative-photometry) pipeline:** uses source catalogs from the photometric pipeline to create high precision lightcurves. Also periodically made.

- **Real-time pipeline:** runs throughout a night to support transient-discovery via image-differencing (PTFIDE). Outputs feed into various science marshals, including solar system object discovery and/or recovery (PTFMOPS)

- Interfacing with the above: an advanced data archive with exploratory tools to support long-term data curation and public distribution – storage of raw data, processed images, and source catalogs

Data Flow



P48
Caltech/ Cahill
NERSC Image Subtraction and Transient Detection/RB Pipeline
IPAC
Ingest
Realtime Image Subtraction Pipeline
Transient Candidates
Moving Object Pipeline
SSOs
Photometric Pipeline
Epochal Images and Catalogs
Reference Images
Reference Catalogs
Lightcurve Pipeline
Lightcurves
Reference Pipeline

Infrared Processing and Analysis Center

- Multi-mission Science Center (IRAS, ISO, Spitzer WISE, Herschel, Planck, 2MASS, etc)
- Maintains several data rooms
- iPTF generates ~1TB of data every 4-5 days
- iPTF compute cluster consists of 24 machines with 240 cores
- Roughly 0.5 PB of spinning disk
- Associated network equipment
- Database and file servers
- Archive servers
- Tape backup
- These will increase by a factor of 10 for ZTF

Disclaimers

- All shortcomings (and peculiarities) in the initial PTF pipeline design result from an expediency in getting the software working rapidly on a tight budget
- We have learned a few lessons and are still learning…
- All opinions mentioned herein are my very own
- Masci joined the project in mid 2012 – so don't shoot the messenger!

Photometric Pipeline

- Triggered at the end of the night, after all the data has been received
- Instrumental calibrations are derived from an entire night's worth of data. Specifically, the bias corrections and flat-fields are derived from the on-sky data
- Photometric calibration is from a nightly model-fit using the SDSS overlap region (more later)
- Astrometric (and distortion) calibration is done at the individual CCD-image level against a combined SDSS and UCAC4 catalog. Typically good to 0.15'' in unconfused regions.
- Outputs are calibrated single-CCD FITS images with bit-masks and accompanying source catalogs in FITS binary table format – both aperture and psf-fit photometry is provided.

Reliance on SExtractor Photometry

- Primary output photometry in catalogs is aperture-based from SExtractor

- To account for the variable seeing (per frame), the project adopted Kron-like aperture magnitudes: aperture size is dynamically derived from the 2nd order moments of the light distribution per source
  - also known as the *mag_auto* measure in Sextractor
  - traditionally used for extended sources (galaxies)

- Why?
  - goal was to obtain science products ASAP
  - project knew how to this better in early 2009: e.g., derive a model PSF per frame
  - infrastructure is now in place to perform PSF-fit photometry

- Currently, *mag_auto* measurements are the only magnitudes that are *absolutely* calibrated
- These are tied to the SDSS photometric system

Calibration of *mag_auto* using SDSS

- Described in Ofek et al., 2012, PASP, 124, 62
- Uses frames that overlap with SDSS footprint to fit a global linear model for nightly data
- Enables calibration of all CCD images observed during a night
- Absolute precision (with respect to SDSS) is ~ 0.02 – 0.04 mag.
- Primary outputs: a global ZP value per image and a spatially-binned ZP residuals map (ZPVM)
- These ZP estimates are only applicable to *mag_auto* instrumental magnitudes
- R^{inst} and g^{inst} below are *mag_auto* (SExtractor) instrumental magnitudes

details). For observations taken using the R-band[18] filter, we fit the following model:

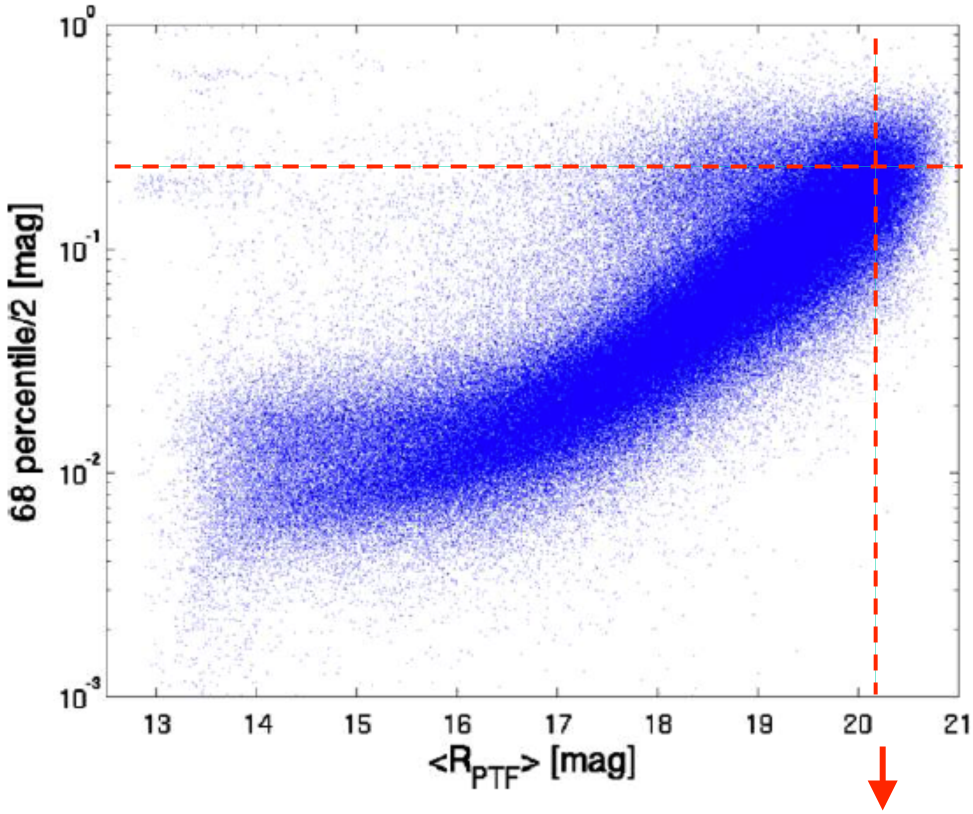
$$\begin{aligned} r_{\mathrm{SDSS}} - R^{\mathrm{inst}}_{\mathrm{PTF}} = {} & \mathrm{ZP}_R + \alpha_{c,R}(r_{\mathrm{SDSS}} - i_{\mathrm{SDSS}}) + \alpha_{a,R}\mathrm{AM} \\ & + \alpha_{ac,R}\mathrm{AM}(r_{\mathrm{SDSS}} - i_{\mathrm{SDSS}}) + \alpha_{t,R}(t - t_m) \\ & + \alpha_{t2,R}(t - t_m)^2 - 2.5\log_{10}(\delta t), \end{aligned} \tag{1}$$

while for g-band observations we fit

$$\begin{aligned} g_{\mathrm{SDSS}} - g^{\mathrm{inst}}_{\mathrm{PTF}} = {} & \mathrm{ZP}_g + \alpha_{c,g}(g_{\mathrm{SDSS}} - r_{\mathrm{SDSS}}) + \alpha_{a,g}\mathrm{AM} \\ & + \alpha_{ac,g}\mathrm{AM}(g_{\mathrm{SDSS}} - r_{\mathrm{SDSS}}) + \alpha_{t,g}(t - t_m) \\ & + \alpha_{t2,g}(t - t_m)^2 - 2.5\log_{10}(\delta t). \end{aligned} \tag{2}$$

Aperture photometry (*mag_auto*) repeatability (R band)

- Multi-epoch stack-RMS of calibrated photometry (Ofek et al., 2012, PASP, 124, 62)
- Uses stars from 100 PTF fields and all CCDs observed over > 3 photometric nights
- Analysis here is not the same as in light-curve pipeline that uses refined gain-correction factors



5σ limit ~ 20.2 mag

PSF-fit photometry

- Implemented and deployed in early 2013 to support iPTF (for single CCD images, reference images, and difference images)

- Uses a re-optimized version of DAOPhot adapted to iPTF stellar densities, seeing and pixel-noise distributions, and detector dynamic-ranges for each filter

- PSF-template is derived using a linearly-varying spatial model assuming a Gaussian basis with correction-residuals stored in a look-up table (DAOPhot format)

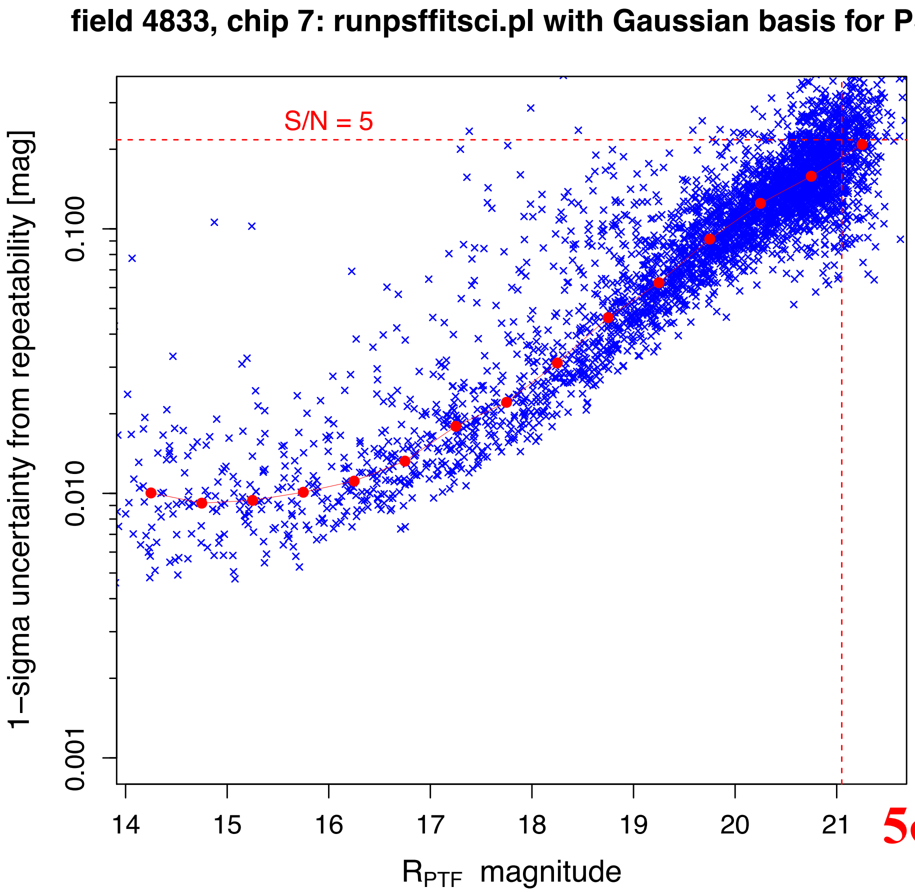
- Products per CCD image: a map of the spatially variable PSF; catalog of instrumental photometry with accompanying (fixed) aperture photometry; astrometrically-calibrated positions; goodness-of-fit metrics per source; DS9 region files to support analysis

- Optimized for point-sources only! Extended source photometry will be biased.
  - ➢ crucial for transient detection since most of these are point sources
- **Huge benefit:** de-blending ability and photometric accuracy at faint fluxes

- Not yet *absolutely* calibrated; existing *mag_auto*-based image ZPs get you within 5 – 15%

Performance of PSF-fit photometry (repeatability RMS)

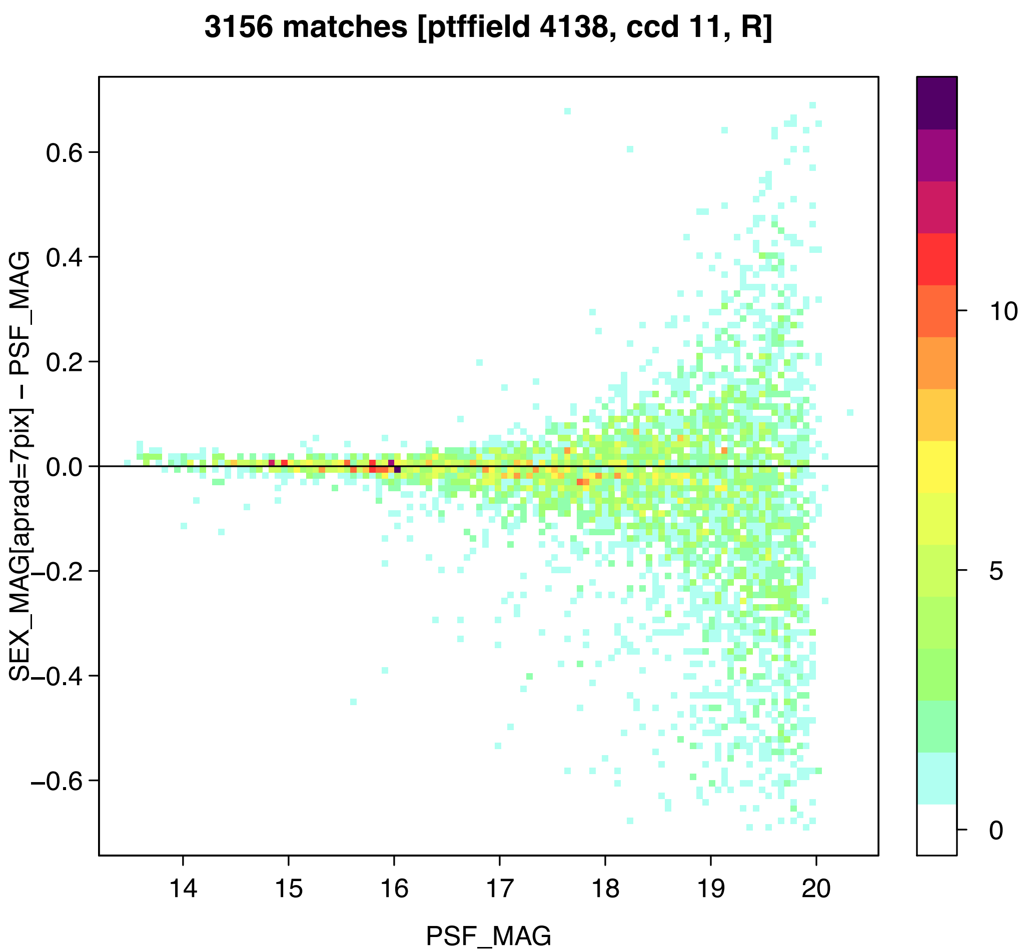
- PSF-fit photometry is not as good as aperture photometry at the bright end
  - ➢ knowledge of underlying PSF is more critical; systematics inflated by centroiding error
- But sensitivity limit is fainter compared to aperture photometry



5σ limit ~ 21 mag

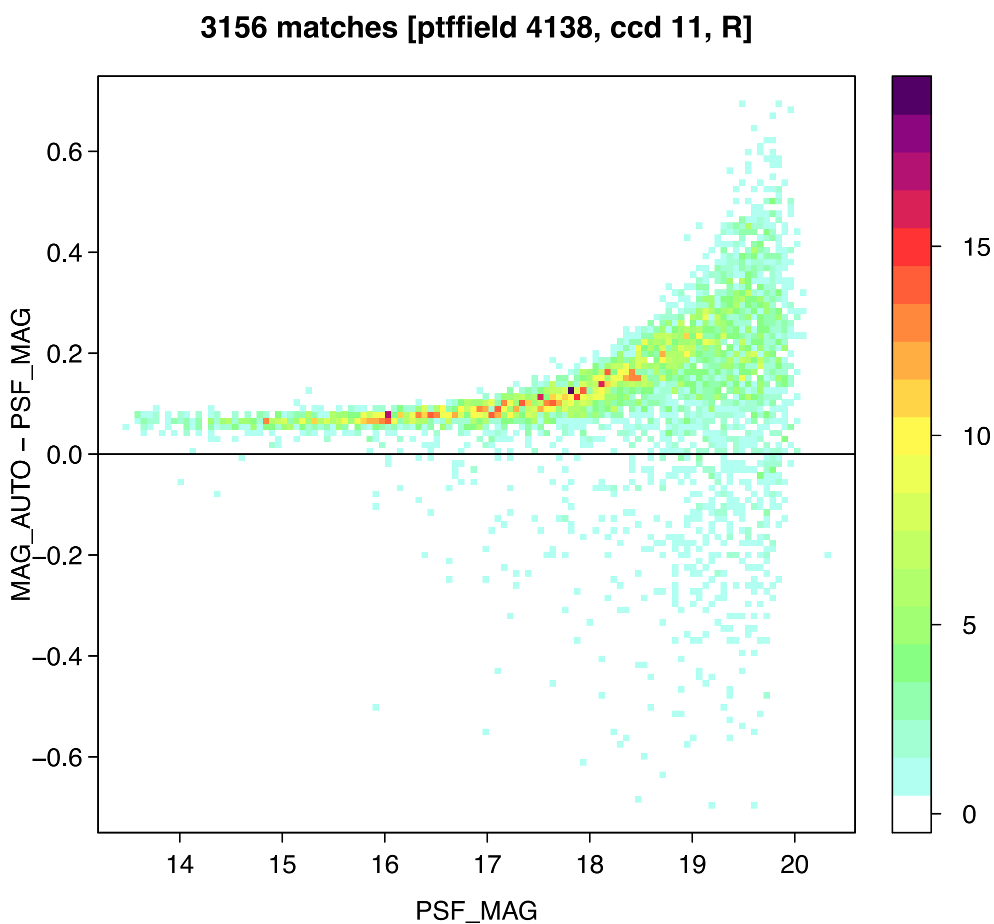
PSF-fitting vs fixed **big-aperture** SExtractor photometry for single frame

Instrumental magnitudes agree within measurement error (i.e., all flux is captured for given seeing)



PSF-fitting vs ***mag_auto*** SExtractor photometry for single frame

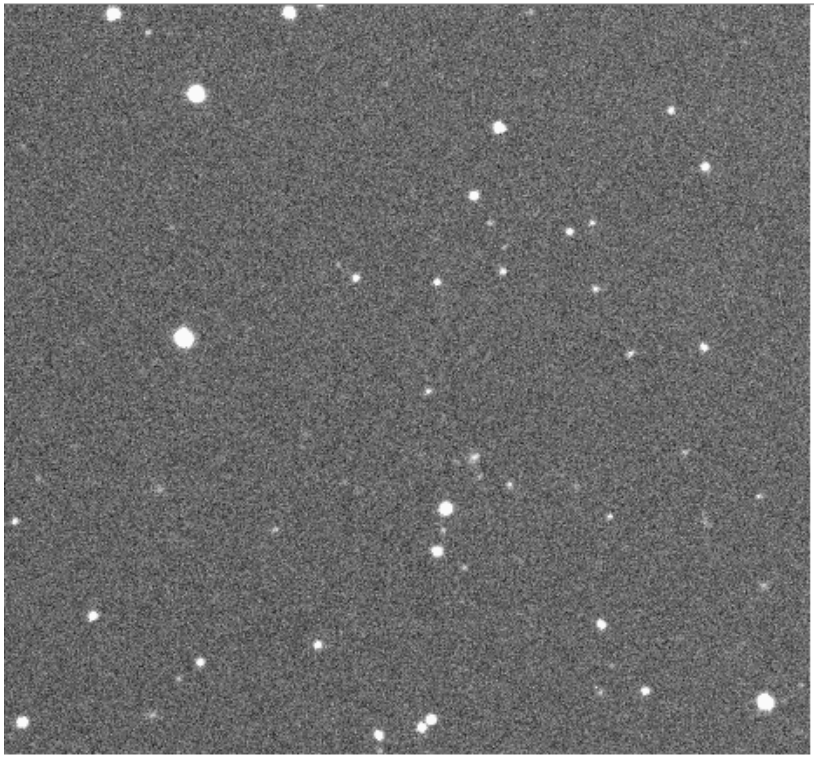
- *Mag_auto* instrumental fluxes underestimated by ~ 5 – 15% relative to PSF-fitting or big-apertures
- Mean bias (per frame) is calibrated-out after absolute calibration since ZP ~ <*m_sdss – mag_auto*>
- **Remaining problem:** this bias is magnitude-dependent! Not intuitive; will bias lightcurve shapes



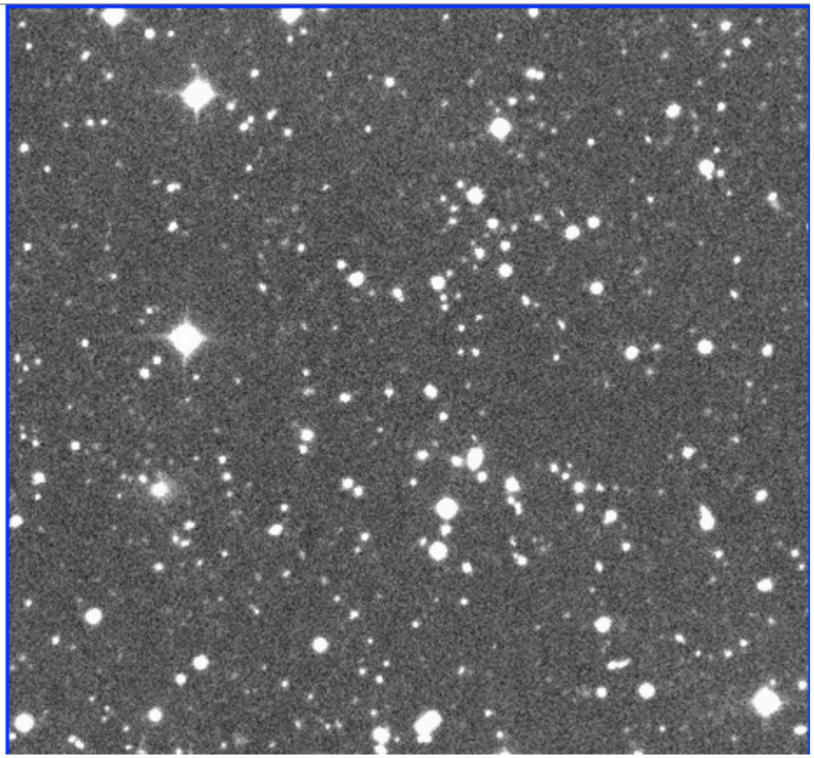
Reference Image Pipeline

- When enough individual exposures accumulate, the “reference image” pipeline is triggered
- This pipeline coadds the “best” image data for a given CCD, field, and filter: e.g., with best seeing, photometric conditions, astrometry, etc.
- Frames are coadded using an outlier-trimmed weighted-average after resampling them using Lanczos interpolation
- The coadds are images of the “static” sky as represented by the state of the input exposures used
  - deeper than the individual exposures: currently $5 < N_{frames} < 50$
- Source catalogs are also generated from these images: both PSF-fitting and aperture (SExtractor)
- Output products support the real-time (image subtraction) and light-curve (relative-photometry) pipelines

Reference Image Example



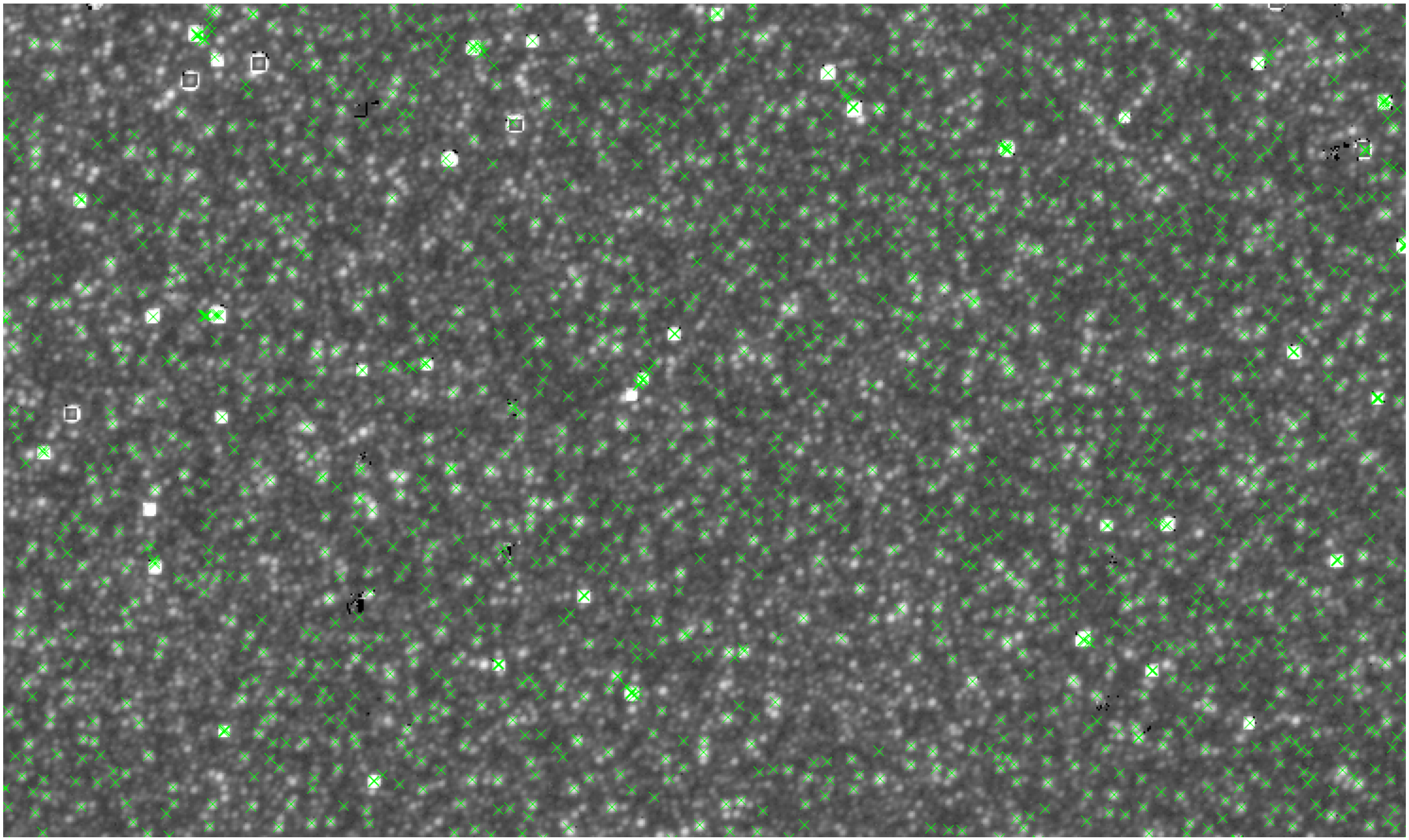
Single image 60 sec in R



Field 5257, Chip 7, Stack of 34

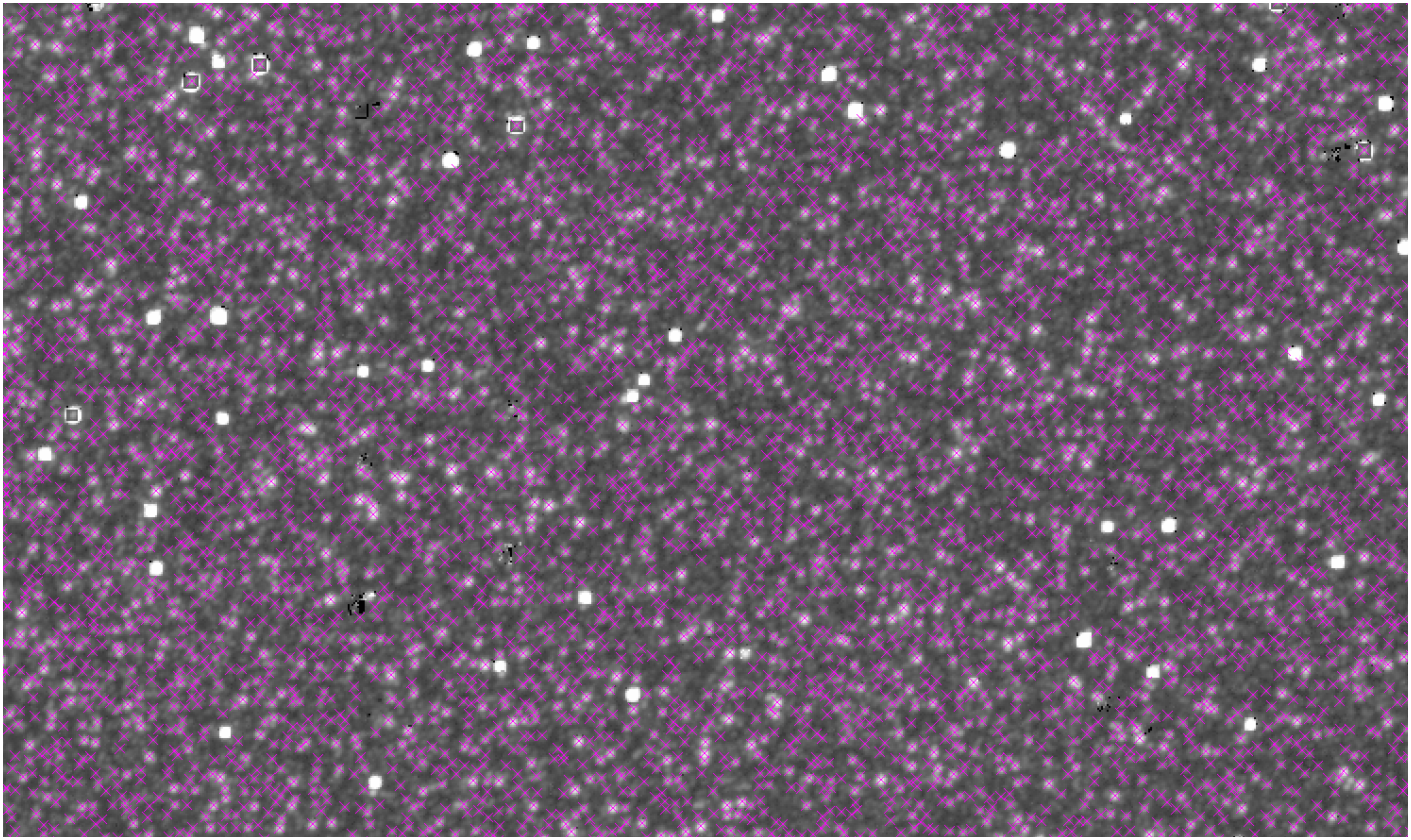
SExtractor extractions on a galactic-plane reference image

Field 1549, Chip 5: 14’ x 7.5’ (*l, b* ~ 7.4°, -6.1°)



PSF-fit (DAOPhot) extractions on the same galactic-plane reference image

Field 1549, Chip 5: 14’ x 7.5’ (*l, b* ~ 7.4°, -6.1°)



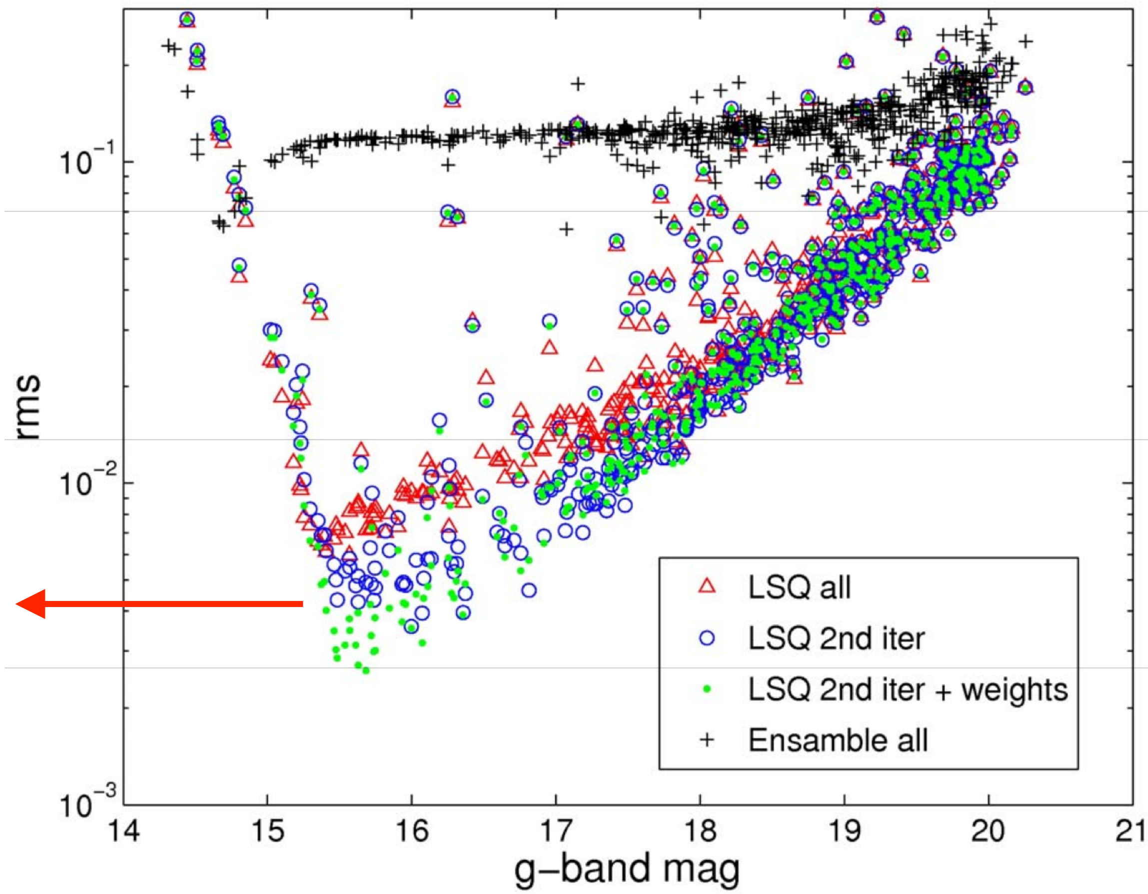
Lightcurve (or relative-photometry) pipeline

- At the end of each night, all SExtractor-detected sources from the photometric pipeline are matched against the reference-image SExtractor catalog for a given CCD, field, and filter
- Uses an optimal, relatively fast matching method; caveats may exist in dense (galactic plane) fields – not yet fully characterized
- The “cleanest” least variable sources are used as anchors for the relative photometric calibration
- Individual image gain-correction factors are computed using an optimal least-squares fitting method; these corrections are stored in a look-up table
- *mag_auto* measurements are currently used throughout this process
- Application of these refined gain-correction factors improves the overall relative calibration to a few millimag for bright sources
- This pipeline is triggered on timescales of typically 1 to 2 weeks

Performance of relative-photometry pipeline (repeatability RMS)

- Plot courtesy of Eran Ofek: based on the method used in the relative photometry pipeline
  - ➢ goal is to minimize and homogenize gain/throughput variations across epochs
- RMS will reach limit of photon-noise + read-noise and centroiding error

~ 4 mmag



Real-time pipeline overview

- Uses image-differencing against the reference-image library to extract transient candidates. Candidates are then automatically "scored" using machine learning.

- Data is processed in near real-time as it's received; turnaround is 10-25 minutes from telescope to vetted transient-candidates in the database

- Outputs are used for same-night follow-up
  - pushed to an external gateway for pickup by the science marshals: galactic, extragalactic, solar-system, and generic ToO alerts

- Difference images and transient-source catalogs are astrometrically and photometrically calibrated
  - both aperture and PSF-fit photometry is performed
  - for the last 2+ years, *mag_auto*-based ZP from reference-image was used to calibrate PSF-fits
  - knew this was wrong!
  - recently, an interim fix was implemented to calibrate the PSF-fit photometry: refined ZPs are derived per CCD-image by matching to a filtered set of pre-calibrated reference image sources

- Outputs from this pipeline also feed a "streak detection" module to find fast-moving objects and a moving-object pipeline (PTFMOPS) to construct moving-object tracklets

Realtime pipeline at IPAC/Caltech

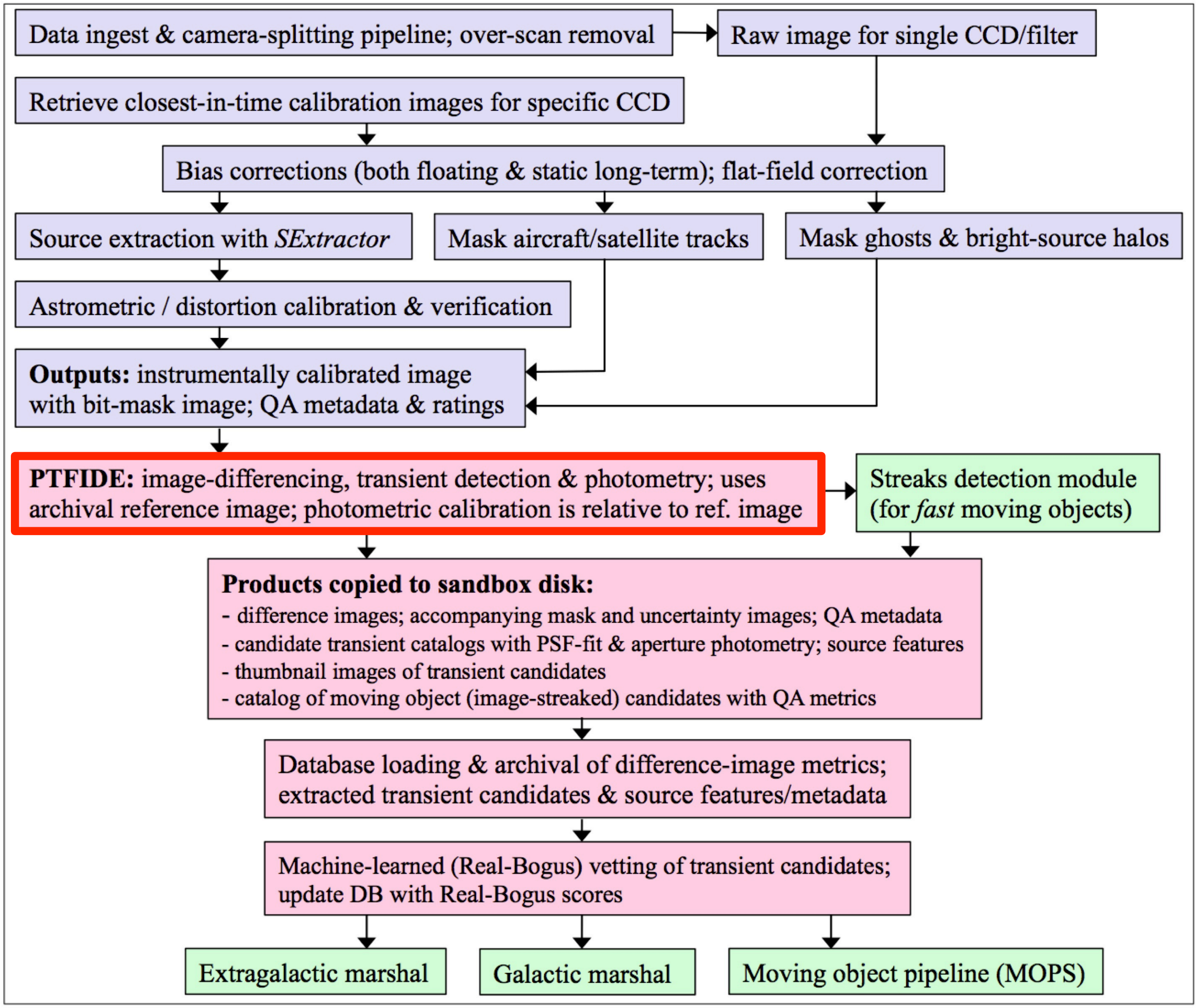
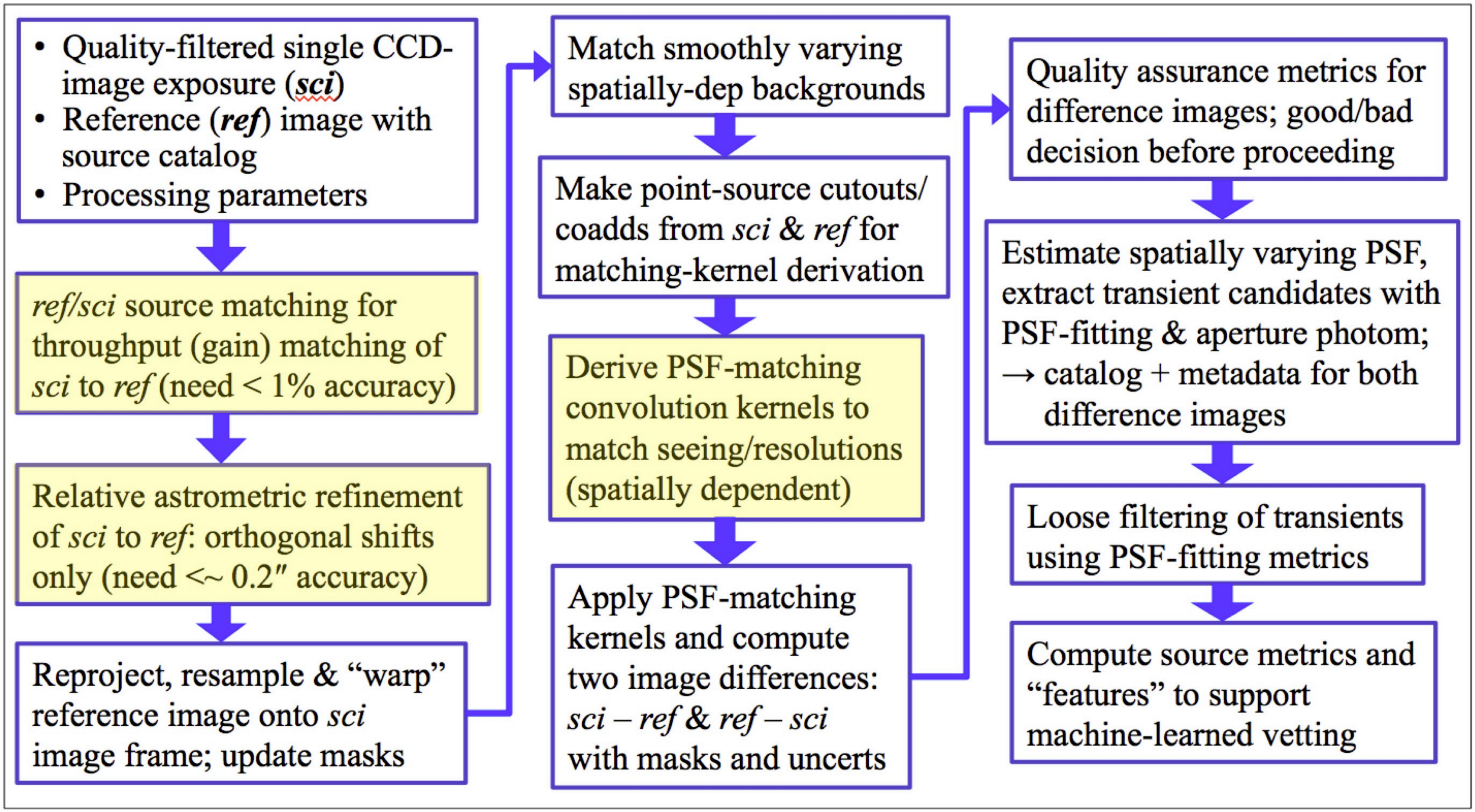


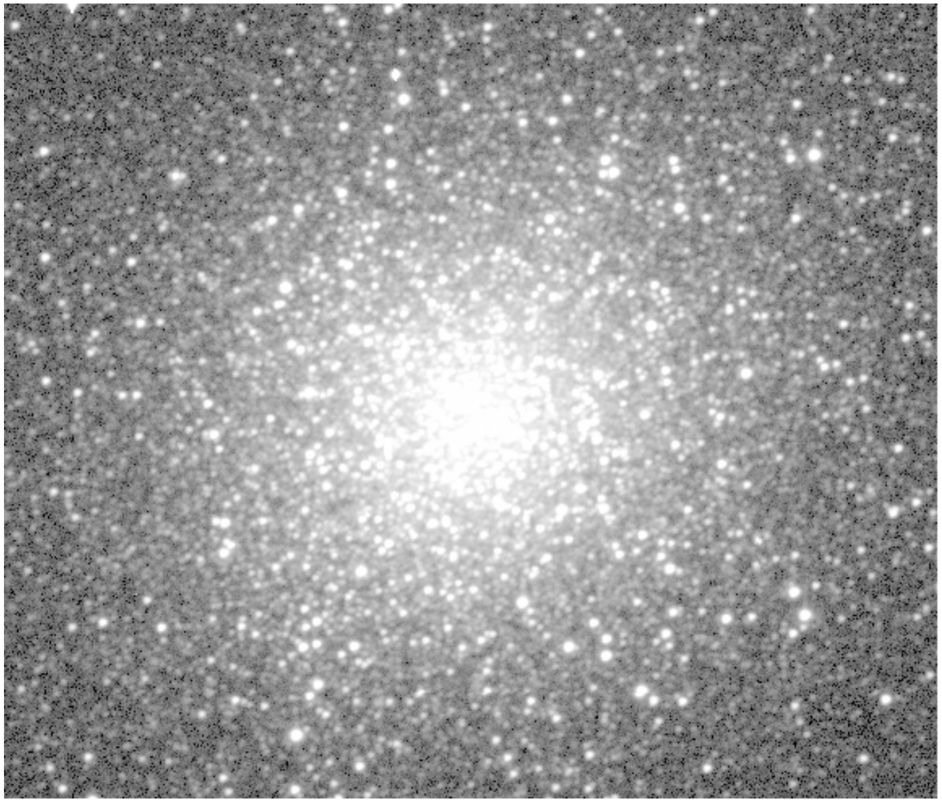
Image-Differencing & Extraction (PTFIDE) processing flow



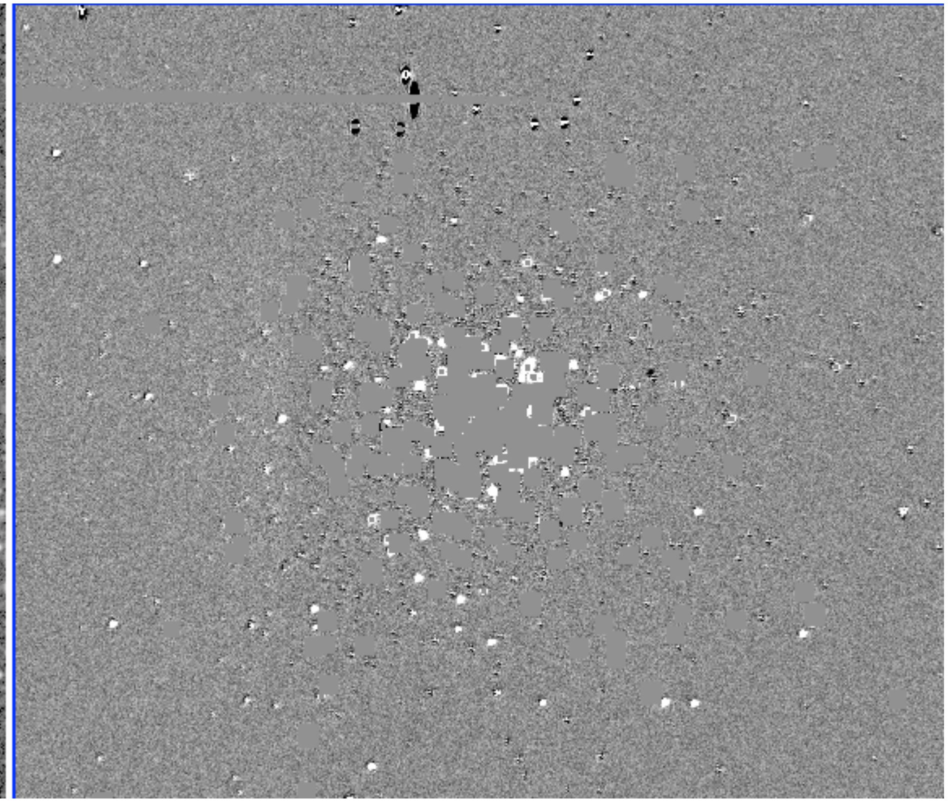
Difference image: zoom on M13 globular cluster

bad/saturated pixels in difference replaced by zero in difference

science image exposure (~ 9’ x 9’ *zoom*)



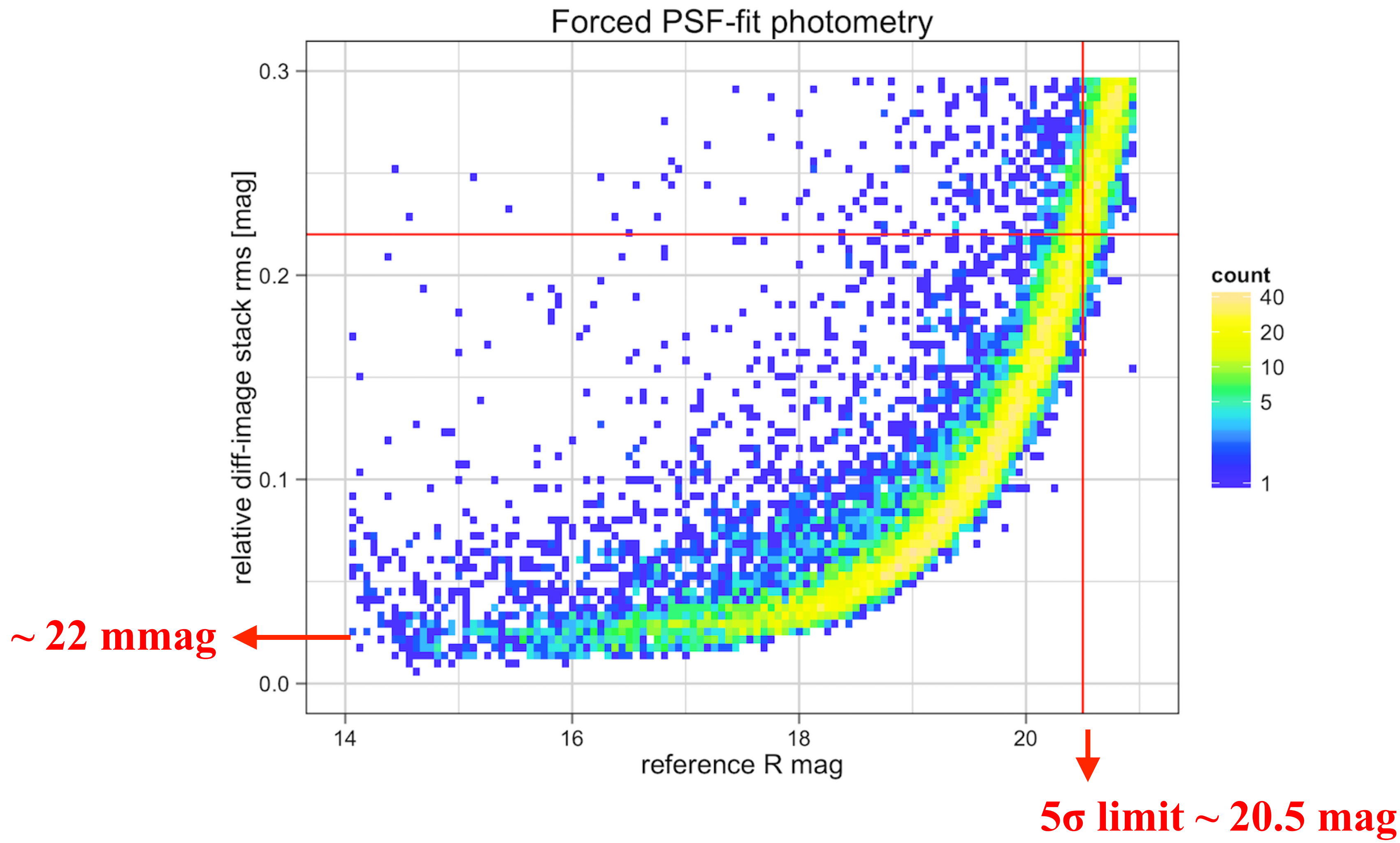
sci – *K* (x) *ref difference image*



Lots of RR-Lyrae!

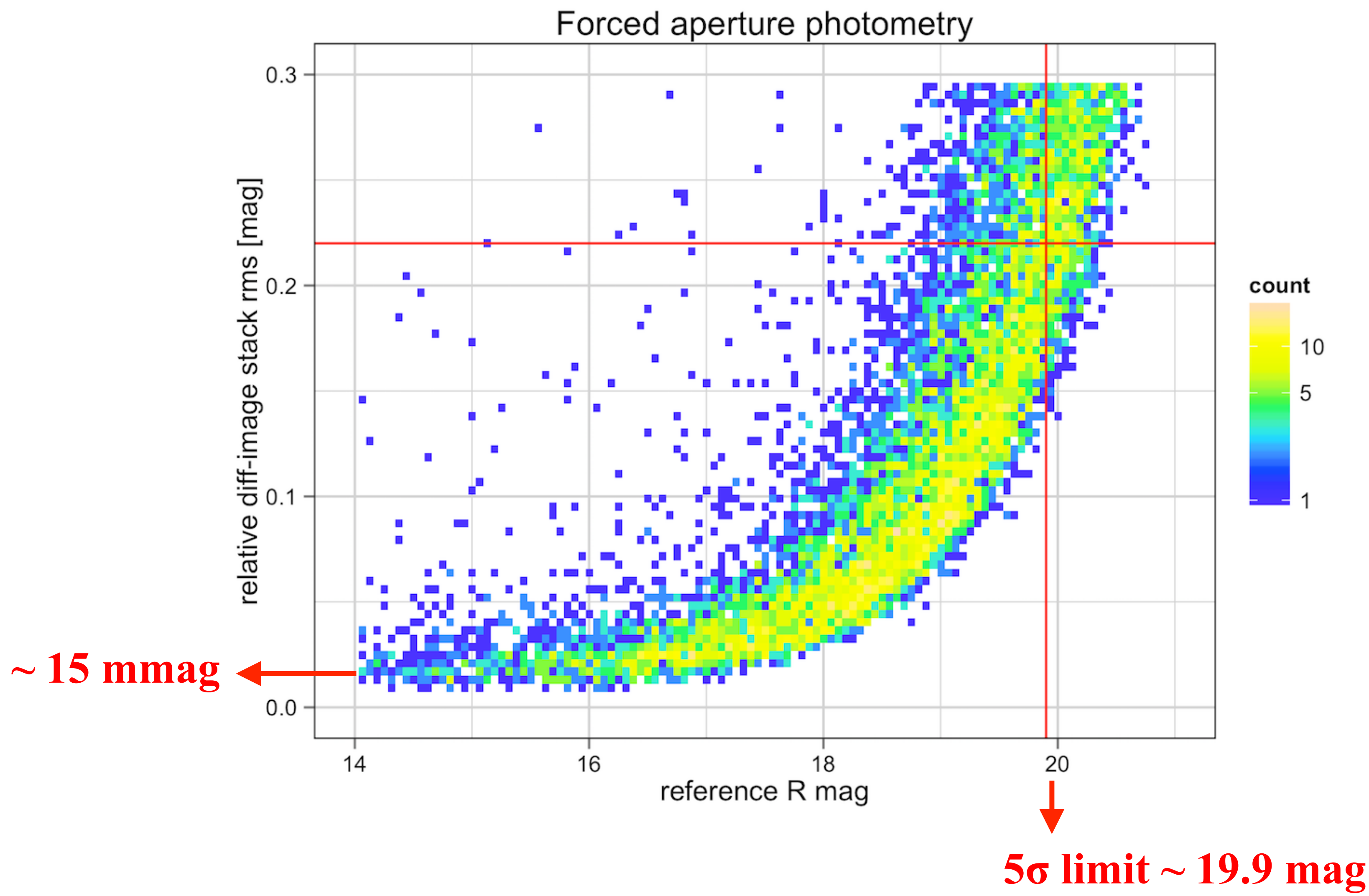
Photometric sensitivity in PTFIDE difference images from **PSF-fitting**

North American Nebula field



Photometric sensitivity in PTFIDE difference images from **Aperture-phot.**

North American Nebula field

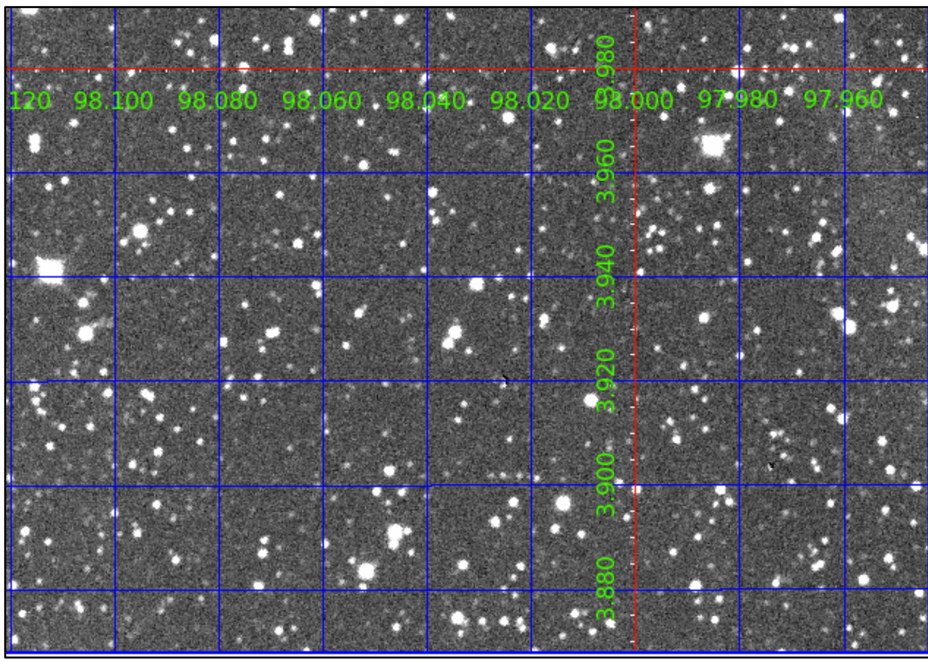


“Good” difference in Galactic Plane

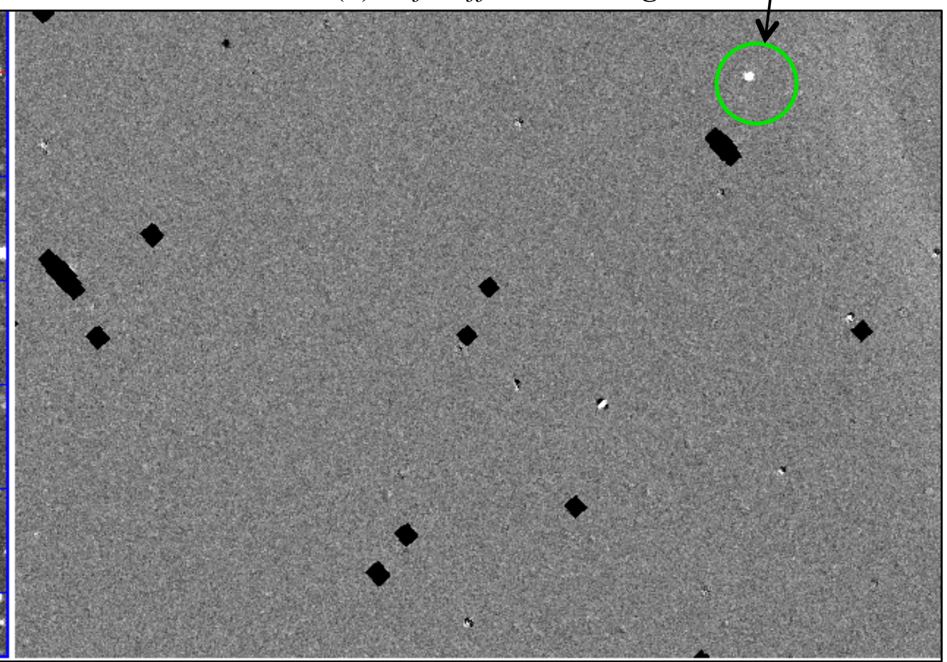
When upstream astrometric/distortion calibration is near perfect, it works!

known variable

science image exposure (~ 10’ x 7’ *zoom*)



Sci – K (x) *Ref difference image*

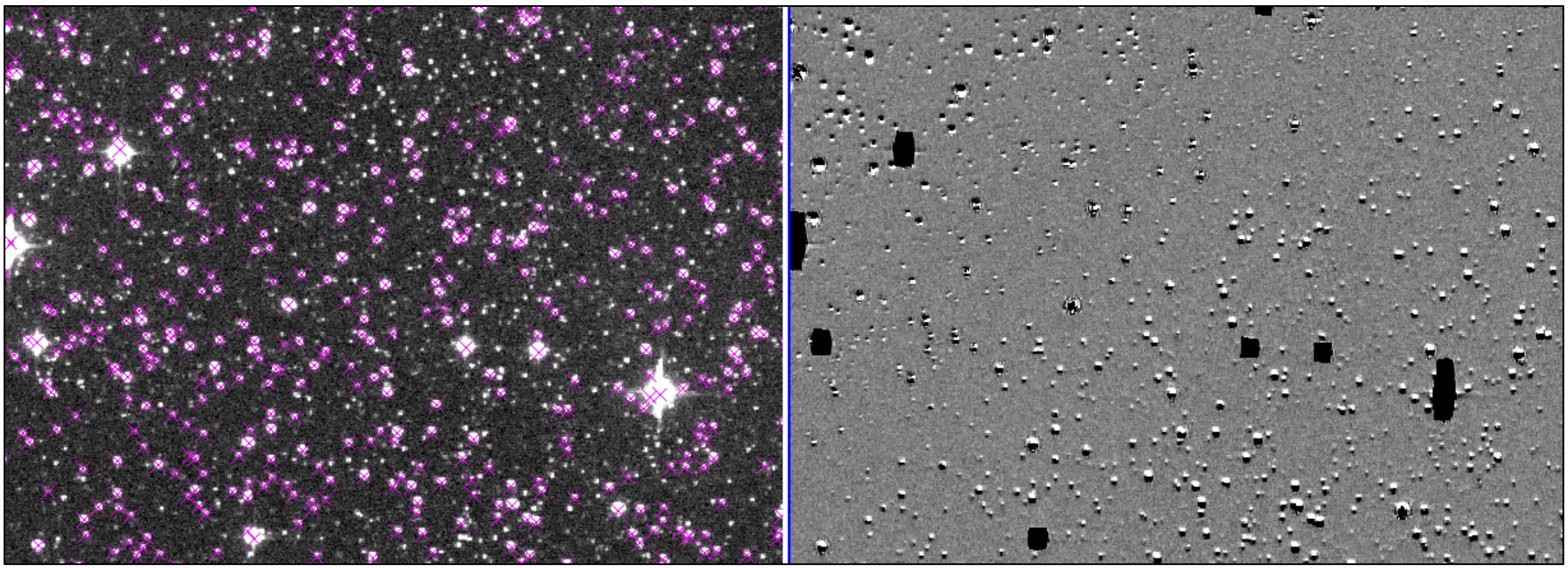


coordinate grid is galactic

When things go wrong:
e.g. “bad” difference in Galactic Plane

When upstream astrometric / distortion calibration is slightly wrong (>~ 0.2 pixel across image)

science image exposure (~ 12’ x 8’ *zoom*) *Sci – K* (x) *Ref difference image*



magenta crosses: 2MASS positions

Crowded-field conundrum(s)

- Crowded fields are a challenge (e.g., galactic plane)! Large science program is planned for ZTF.

- Instrumental calibration is more difficult:
  - astrometric and photometric calibrations require source matching of some sort

- Source-matching is ambiguous and messy in crowded fields
  - naïve nearest neighbor matching using some radial tolerance is not robust
  - use of aperture photometry for photometric calibration (relative & absolute) is not optimal

- E.g., current success rate for good (usable) difference images in galactic plane is ~ 50%
  - bad subtractions strain the candidate extraction and ML vetting steps downstream
  - currently, no transients are extracted/stored from *really* bad subtractions: this impacts survey completeness

Summary: what works and what doesn't (optimally at least)

- The iPTF pipelines work (reasonably) well with product quality depending critically on the quality of upstream instrumental calibrations (e.g., astrometry, flat-fielding, bias corrections, image ZPs)

- What isn't optimal and needs work; or where accuracy is difficult to assess:
  - absolute photometric calibration using *mag_auto* instrumental photometry: seeing and signal-to-noise dependent systematics are hard to characterize
  - calibration performance for fields outside SDSS footprint (model-prediction accuracy?)
  - use of aperture measurements for photometric calibration (both absolute & relative)
  - astrometric / distortion calibration in crowded fields
  - source-matching algorithms that feed the astrometric and photometric calibration steps – want to improve (and salvage) more products in the galactic plane

- Software, infrastructure, and most algorithms are generic and robust enough to use for ZTF
  - most of the effort will be optimizing and tuning the end-to-end system
  - adapting to the new detectors, survey design, cadence, data rates, etc.

Closing thoughts / discussion points

- With PSF-fit photometry currently in the iPTF pipelines, can now move towards using this as the primary *calibrated* photometric product
  - can derive instrumental zero-points per-image using Pan-STARRS in the near future
  - can also use these same zero-points for “big-aperture” instrumental photometry (with caution)
  - also use in light-curve pipeline: PSF-fitting mitigates some of the problems in crowded-fields

- Discussion points:
  - *calibrated* extended-source photometry?
  - software to support new calibration infrastructure is needed
  - careful selection of well-characterized (non-variable) calibrators
  - magnitude zero-points to physical-flux conversion => color corrections, spectral modeling?
  - astrometric reference catalog to use?
  - requirements on absolute (and relative) photometric accuracy
  - requirements on what exactly the (legacy / archival) products will be

Further reading

- IPAC-PTF pipelines and data archiving (as of May, 2014); Laher, Surace et al.
  http://www.jstor.org/stable/pdfplus/10.1086/677351.pdf

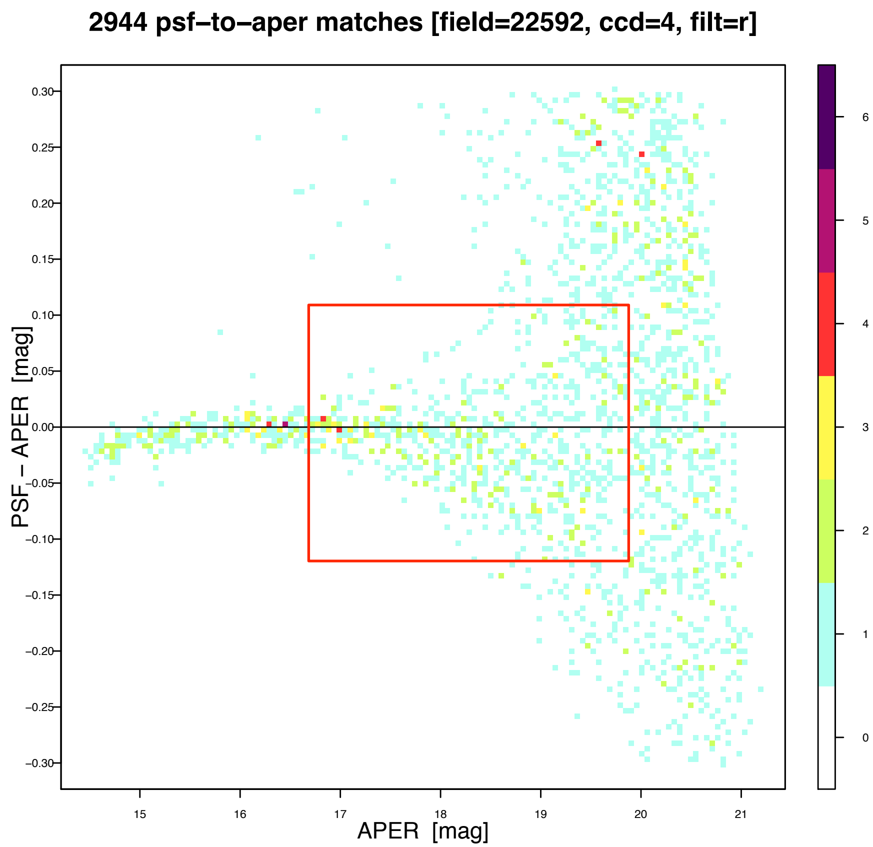
- More detailed presentation on image-differencing – PTFIDE (Masci, 2014)
  http://web.ipac.caltech.edu/staff/fmasci/home/miscscience/masci_lsst_ztf_Nov2014.pdf

- (Forced) photometry on difference-image products (Masci et al., 2015)
  http://web.ipac.caltech.edu/staff/fmasci/home/miscscience/forcedphot.pdf

Back up slides

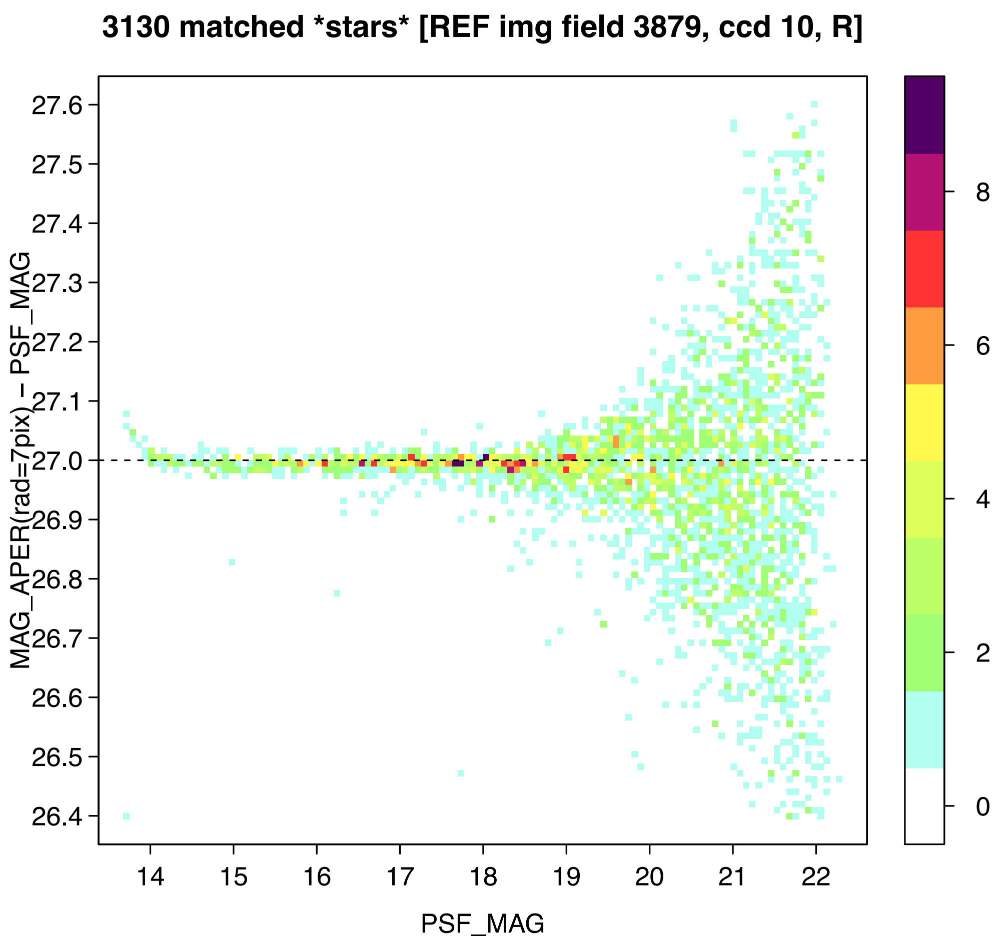
The “Bigger-Fatter” effect in PSF-fitting

- When seeing is really good (=> undersampled PSF) biases can creep in.
- If calibrate the PSF at the bright (fat) end, could overestimate the PSF-fitted fluxes of faint stars.



PSF-fitting vs fixed **big-aperture**
Sextractor photometry for reference image

Instrumental magnitudes agree within measurement error (i.e., all flux is captured)



PSF-fitting vs ***mag_auto*** Sextractor photometry for reference image

- *Mag_auto* instrumental fluxes underestimated by ~ 5 – 15% relative to PSF-fitting or big-apertures
- Mean bias (per image) is calibrated-out after absolute calibration since ZP ~ <*m_sdss* – *mag_auto*>
- **Remaining problem:** this bias is magnitude-dependent! Not intuitive; will bias lightcurve shapes