ZTF Data System, Deliverables & Plans

Frank Masci & the IPAC-Caltech ZTF Team

ZTF Working Group Meeting, August 2017

Outline

- Overview and summary of data processing
- Status of ZTF Pipelines
- Performance (real-time processing)
- Primary products and deliverables
- Data access/retrieval portals, user-interfaces and tools
- What remains for v1 (baseline system)
- Priorities and schedule

Processing details, all metrics, products, and access methods are further described in:
http://web.ipac.caltech.edu/staff/fmasci/home/junk/ztf-archive/ztf_pipelines_deliverables.pdf

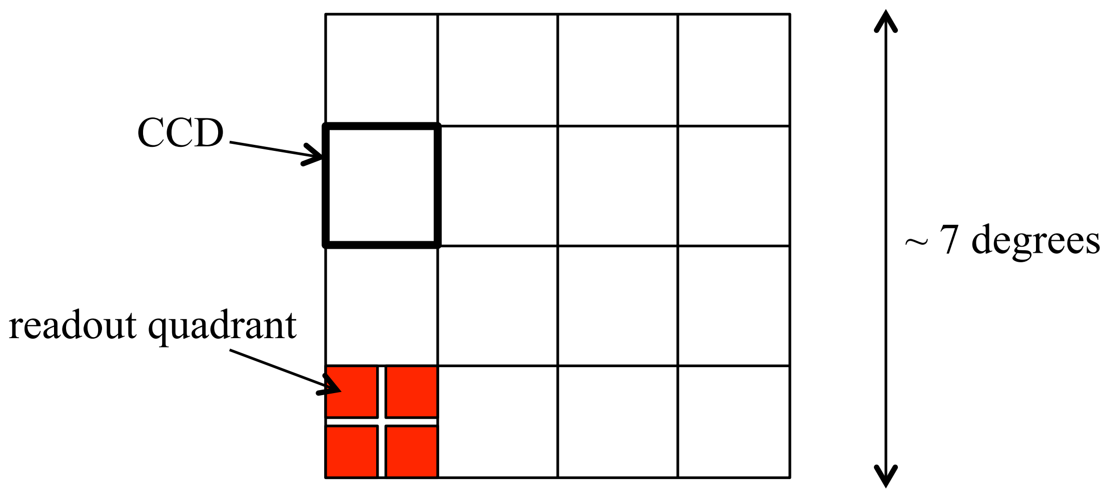
The ZTF Science Data System (ZSDS): staff

- **Ben Rusholme:** data link from P48 to IPAC; pipeline job scheduling/executive; optimization; software/configuration management; hardware config.; alert distribution infrastructure (Kafka)
- **David Shupe:** astrometric calibration; source-matching and relative photometry pipeline
- **Russ Laher:** pipeline infrastructure; integration and testing; data ingest; pipeline executive; database schemas and stored procedures; bias- and flat-generation pipelines;
- **Steven Groom (and staff; IRSA Lead):** pipeline/archive interface design; system engineering; hardware shopping/costing and provisioning.
- **Frank Masci (ZSDS Lead):** instrumental and photometric calibration; reference-image generation; image-subtraction; extraction; moving-objects; algorithms; analysis; documentation…
- **David Flynn (and staff; ISG Lead):** system-engineering and hardware
- **Ed Jackson:** database management
- **Jason Surace:** image simulation; data analysis
- **Ron Beck:** pipeline operations
- **David Imel (IPAC manager):** budgeting and personnel
- **George Helou (IPAC director)**

The ZSDS is not a clone of PTF. Developed from scratch & optimized to handle high data rate/volume

Raw Camera Image Data

- One camera exposure: 16 CCDs; each ~ 6k x 6k pixels
- Image data packet transmitted is per CCD (four readout-quadrant images & overscans)
- 16 CCD-based image files are acquired every 45 sec (30s exposures).
- Full camera exposure: ~ 1.3GB uncompressed
- Require some compression to accommodate transfer bandwidth (TBD)



Basic image-unit for pipeline processing from which all products are derived is a ~3k x 3k readout quadrant image

- Approximately 4 TB of data products are generated on average per (~ 8-hour) observing night

ZSDS development: guided (primarily) by this document

Zwicky Transient Facility Science Requirements Document

Eric Bellm

December 16, 2015

http://www.oir.caltech.edu/twiki_ptf/bin/viewfile/ZTF/RequirementsDocuments?filename=ZTF_SRD_1.0.pdf

ZTF Pipelines and run frequency

Overall, there are 9 interdependent pipelines, grouped into four categories:

Raw data ingestion and initial processing:

1. Raw data ingest, archival of raw images and storage of metadata in database [*realtime*]
2. Raw-image uncompression, splitting into readout-quadrant images, floating bias correction, QA metrics [*realtime*]

Calibration-image generation:

3. Bias-image derivation from stacking calibration images acquired in afternoon [*before/after on-sky operations*]
4. High-ν flat (pixel-to-pixel responsivity) from stacking illum. flat-screen exposures [*before/after on-sky operations*]

Real-time science-level processing:

5. Instrumental calibration of readout-quadrant images: includes astrometric and photometric calibration [*realtime*]
6. Image-subtraction with transient-event extraction (point sources & streaks), alert packets & distribution [*realtime*]

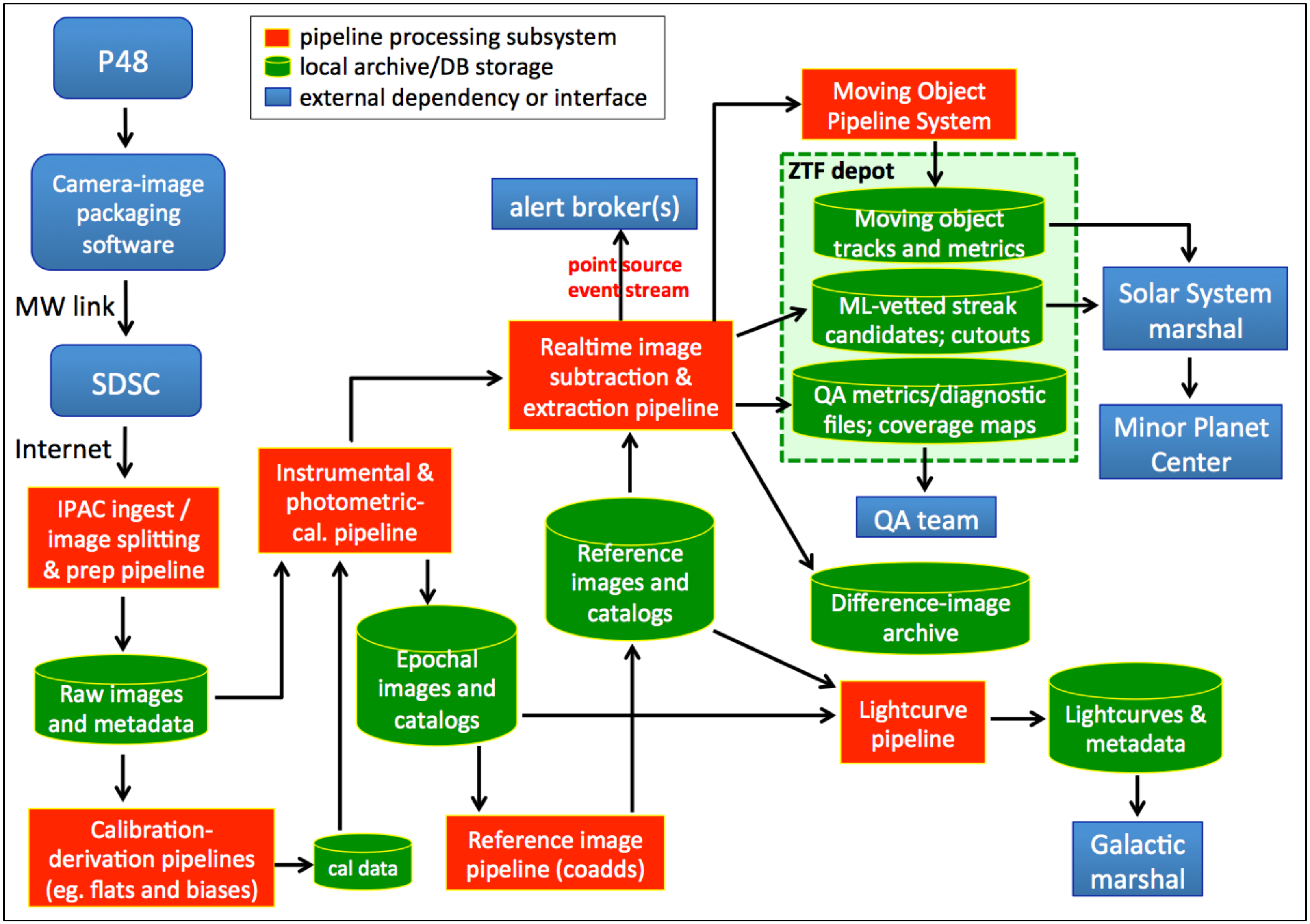
Ensemble-based (collective-image/catalog) processing:

7. Reference-image generation (co-addition of epochal images from 5) [*when sufficient good quality data available*]
8. Source-matching/lightcurves with relative photometric refinement; inputs from 5 & 7 [*every month, TBD*]

In progress:

9. Moving object tracks, orbit-fitting, QA; from linking point-source events from 6 [*end of night, 3 day data window*]

Overall processing & data flow



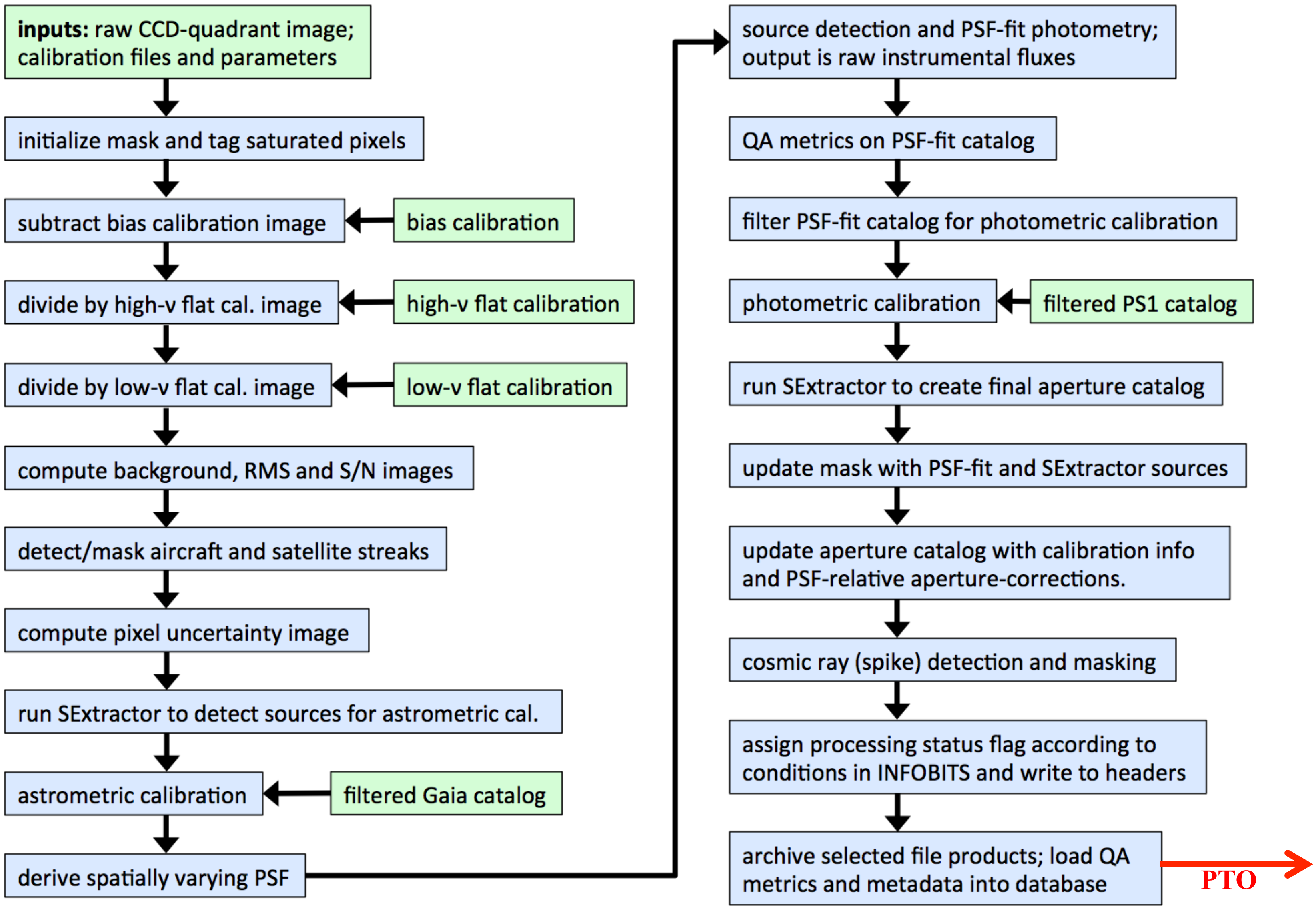
ZTF Real-time pipeline

- Primarily to support fast-response science
- **Timing requirements (Dec 2015):**
  - > 95% of the images acquired at P48 need to arrive at IPAC within 10 min (goal: 5 min)
  - > 95% of the images received at IPAC must be processed with alerts in < 10 min (goal: 5 min)

- Real-time pipeline consists of two phases:
  1. **Instrumental calibration** (bias-corrections, flat-fielding, astrometry, photometric calibration, masking of bad pixels, …): generates single-epoch image and catalog products for archive.
  2. **Image subtraction & extraction** of transient events (point-sources & streaks), QA & source features, filtering, ML-vetting, cutouts, point-source alert packet generation …

- Currently tested using a camera-image simulator:
  - Takes as input a "schedule" of camera pointings from Eric's survey simulator.
  - Point sources are injected with same photometric properties and positions as in the PS1 catalog; appropriate noise is injected.
  - Random fake point-source and streaking transients are also added.
  - Image headers adhere to camera software specifications

ZTF Real-time pipeline (phase 1): instrumental calibration



ZTF Real-time pipeline (phase 2): image subtraction & extraction

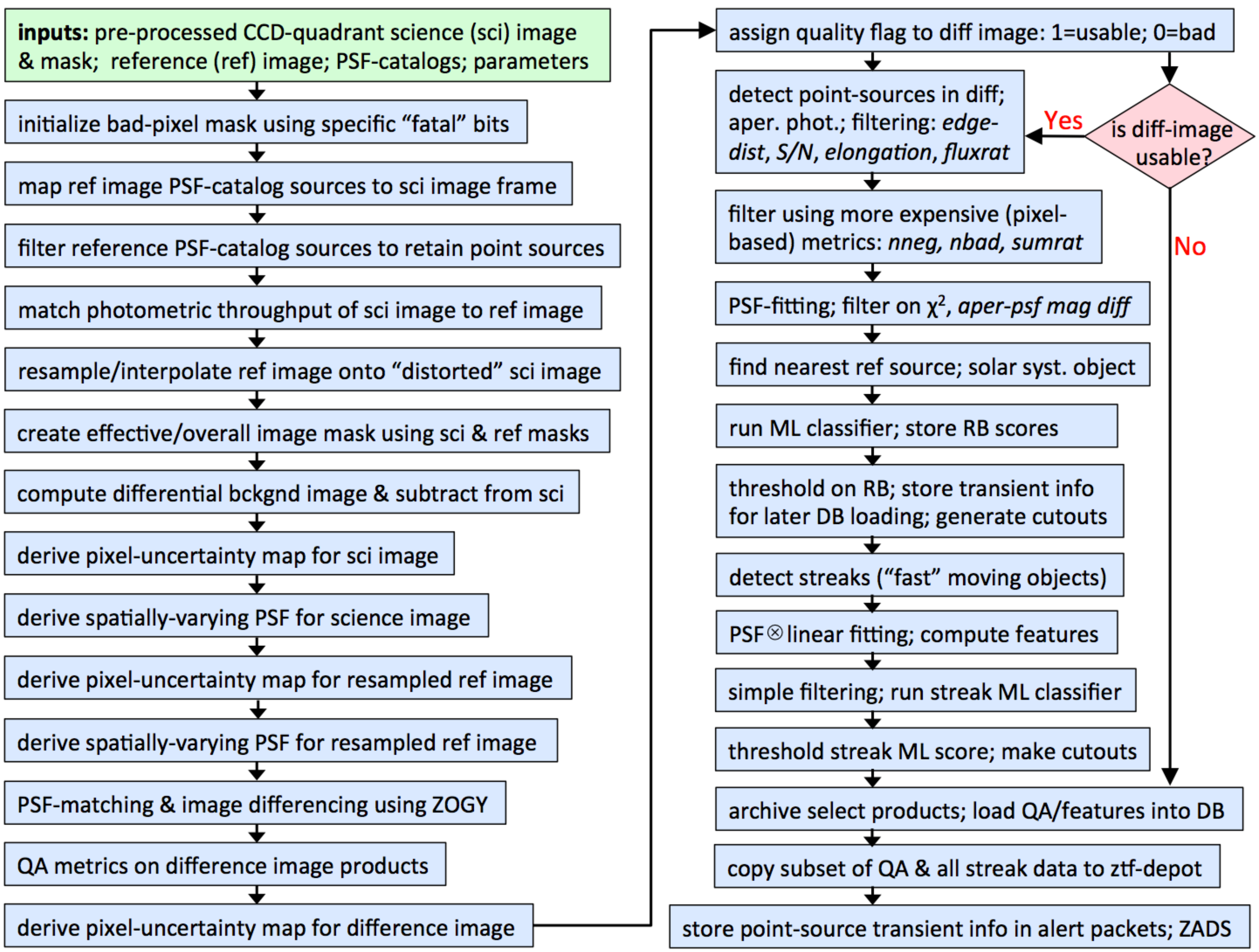
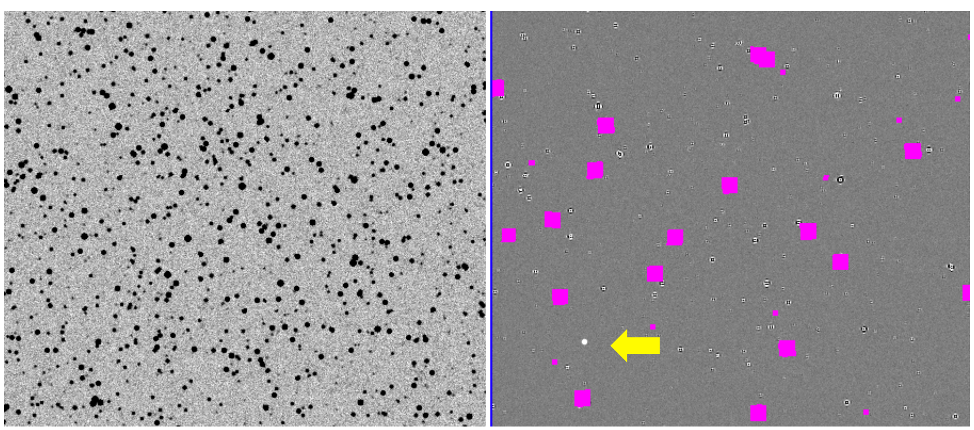


Image Differencing & Event Extraction

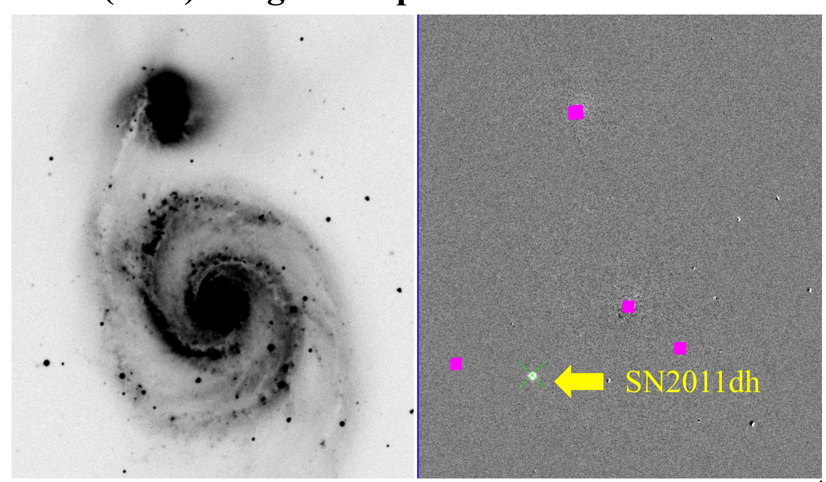
- PSF-matching, differencing, & matched-filtering for event detection is done using ZOGY (Brad)
  - small algorithmic updates are in progress
- Benefits of ZOGY (*statistical optimality*) crucially depends on accuracy of upstream calibrations & prep. steps: flat-fielding, gain-matching, astrometry (+ distortion), PSF-estimation
- Over last several months, much R&D has gone into designing metrics to filter obvious false-positives from the raw candidate detection stream (no formal ML-classification, yet)
- From testing on PTF data, a few to tens of likely real events remain per 0.7deg^2 readout image down to S/N ~ 5 following simple filtering: this all events (variables, [super]novae, asteroids, …)
- Assuming ~ 10 likely real events per readout image, expect ~ 0.5 million events/night to S/N = 5

simulated ZTF-like image zoom:

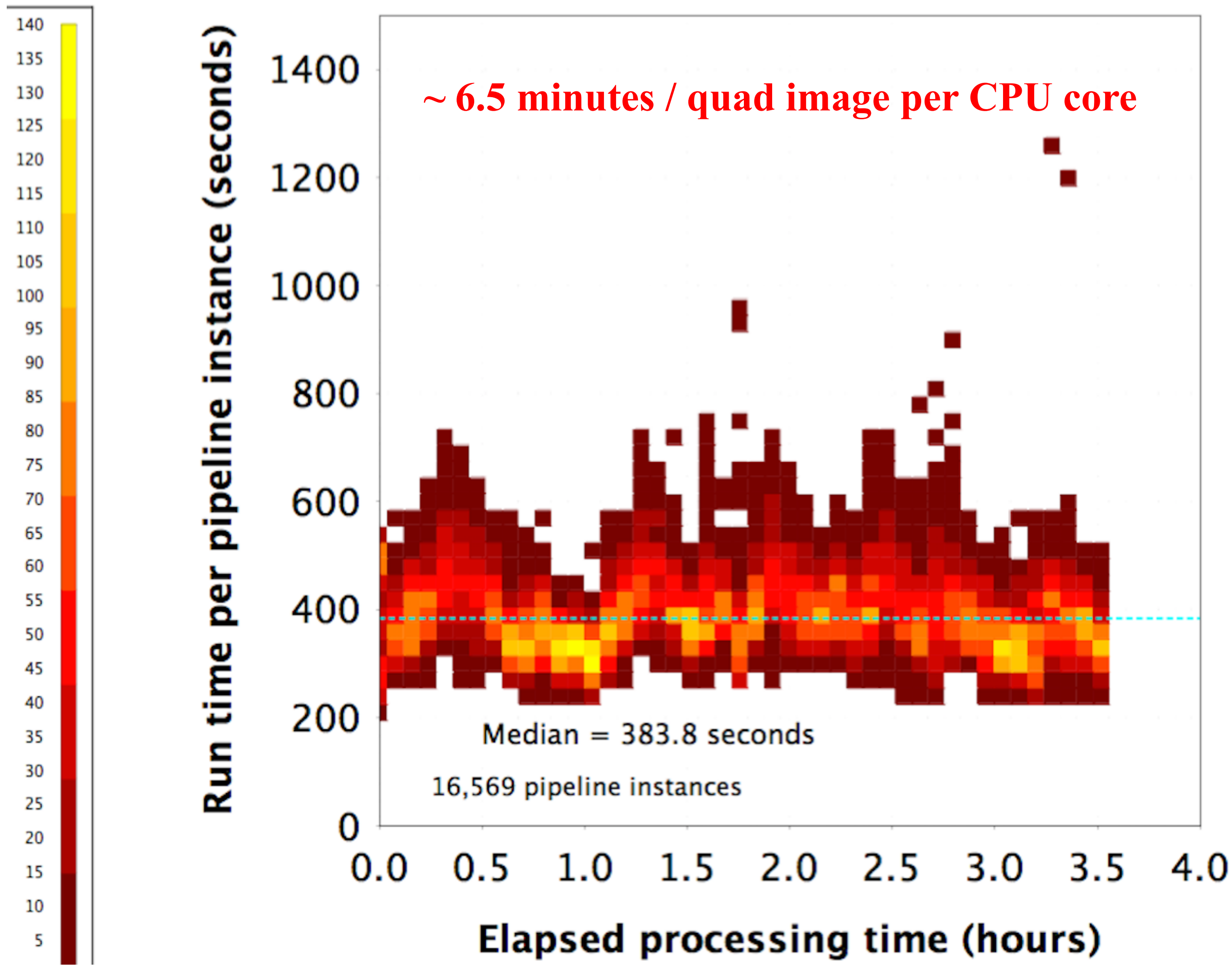
science image **subtraction image**



real (PTF) image example:



ZTF real-time pipeline runtime (processing unit = one readout-quad image)



- 32 machines × 16 cores each
- 45 sec between exposures
- 265 input exposures: ~ 3.3 hrs
- approx. half a summer night that crosses galactic plane

- full realtime pipeline with alert packets containing 30 day histories per event, streaks, ML-vetting, ztf-depot..

Hardware procured to commence science operations

- Until now, development, testing, and benchmarking of realtime pipeline has utilized a compute cluster of 32 nodes (16 physical cores each)
- Looking ahead, other processing needs and demands need to be accommodated. Hence, we recently purchased another 34 compute nodes.
- Also purchased network switches; disks for archive (1PB), pipeline ops & DB; file and DB servers

Racks containing 66 compute nodes

Archive fileservers/disk arrays (holds up to 4 PB)

Baseline deliverables & data access portals

1. **Instrumentally calibrated, <u>readout-quadrant based</u> epochal image products:**
   - images with photometric zero-points from PSF-fit photometry; bit-mask images
   - two source catalogs per image: PSF-fitting and aperture photometry
   - difference images, accompanying matching PSFs and QA metadata
   - ➢ archive via GUI or API at IPAC; can interface with Moving-Object Search Tool (MOST)

2. **Reference images (co-adds), coverage, unc maps, and two source catalogs per image**: PSF-fitting and aperture
   - ➢ archive via GUI or API at IPAC

3. **Lightcurves and collapsed-metrics using "object-based" searches by position and/or metrics**
   - ➢ archive via GUI or API at IPAC; can interface with lightcurve viewer/analyzer

4. **Match-files: all lightcurves per readout-quadrant:** from source-matching of epochal PSF-fit extractions
   - ➢ restricted (galactic marshal)

5. **Raw image data (CCD-based files with metadata) and image calibration products used in pipelines**
   - ➢ archive via GUI or API at IPAC

Baseline deliverables & data access portals

6. **Alert (point-source event) stream** from real-time image-differencing pipeline: packetized with metadata, 30 day photometric histories, upper limits, ML-scores, cutouts on new, reference and difference images, …
   - ➢ transmitted to UW; access via specific science/filtering channel using "Kafka consumers"

7. **Products to support realtime Solar System/NEO discovery and characterization:**
   - streaks (fast moving objects) from difference images: metrics, ML-scores, and cutouts
   - moving object tracks from linking point-source events; known objects are tagged
   - ➢ ZTF-depot (restricted audience)

8. **Quality assurance metrics, summary statistics, and coverage maps for performance monitoring**
   - ➢ ZTF-depot (restricted audience)

9. **Documentation:** pipeline descriptions, recipes and tutorials on data-retrieval

---

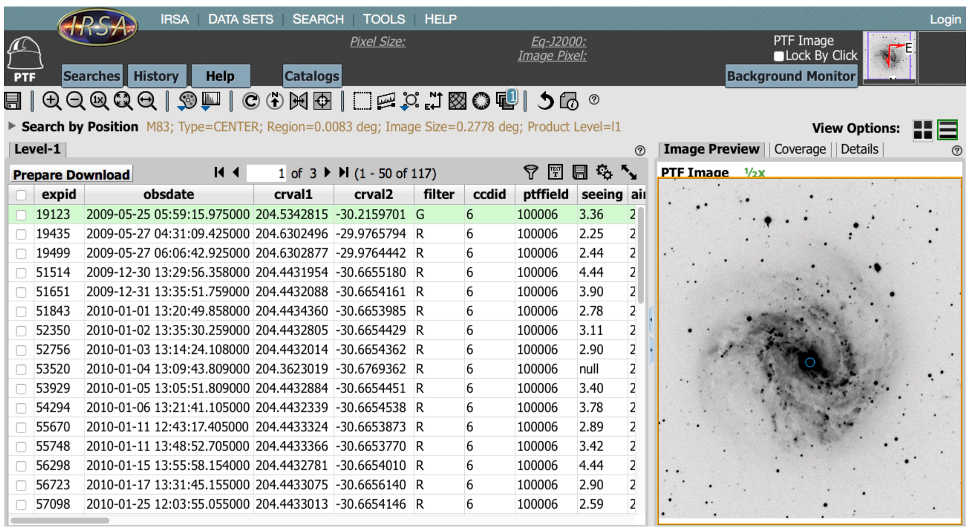
Not a deliverable:

- Archival difference image photometry; not funded; only incorporated into alert products.
- All events, difference image photometry and source-features are being stored in a database.

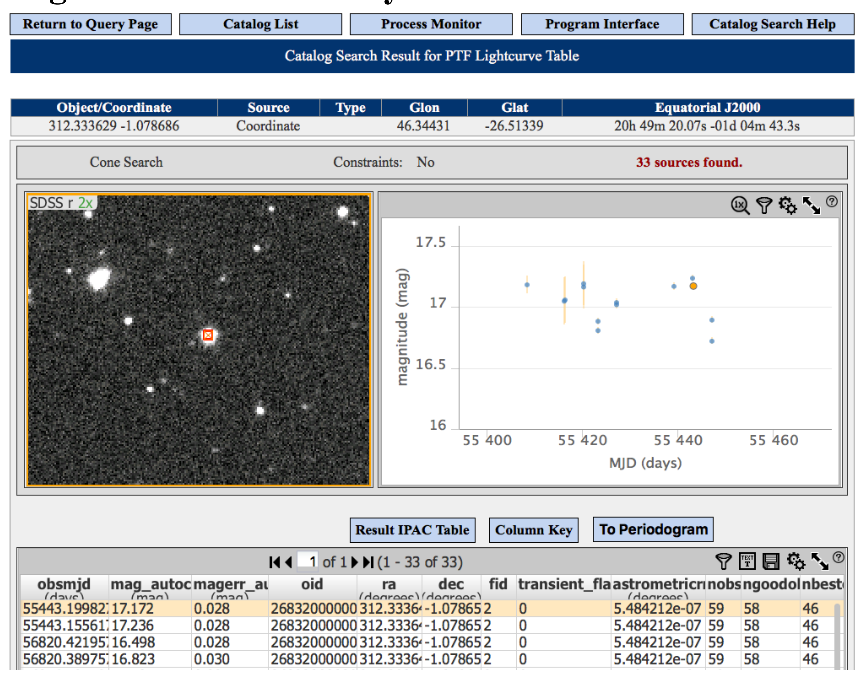
User Interfaces to archive products

- Will allow search by position, time-windows, filtering on metadata, object name, interactive manipulation, catalog overlays, visualization, basic analysis of lightcurves with periodogram service.
- Accompanying APIs (command-line driven retrieval) also available

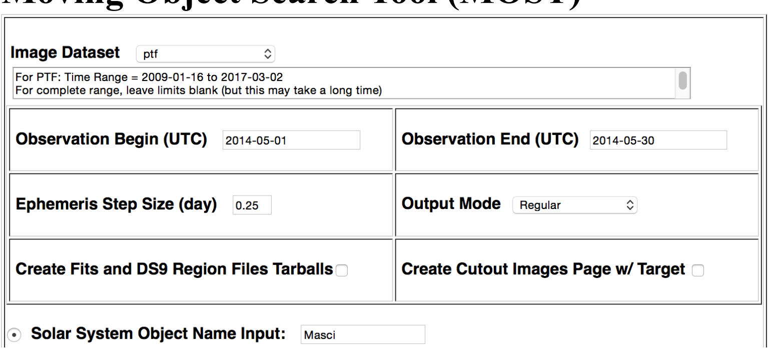
Image viewer and file-product retrieval



Lightcurve viewer/analyzer and retrieval



Moving Object Search Tool (MOST)



Data access for collaboration during commissioning / science-validation

- Archive user-interfaces and tools will not be in place until October 2017.
- To facilitate access to pipeline products during the commissioning and science-validation by the collaboration, a file-based webserver “ztfweb” will be set up.
- Will enable access to non-primary (intermediate) pipeline products to support analyses.
- May also provide early access to alert packets for testing: in *avro* or *json* or other format.
- No GUI or search functionality connecting to a database; purely file based, following archive layout.
- Users will be authenticated through individual accounts, with prior consent by Project Scientist / P.I.
- **This access method is temporary and will disappear** when archive GUI becomes available.

Data System status

green => in progress
red => to be developed, characterized or need input
black => done

Data transfer software/protocols from P48 to IPAC

Pipelines
- Ingest, CCD quadrant-splitting, floating bias correction, QA
- Calibration generation: biases, high & low-ν flats
- Instrumental calibration (astrometric & photometric & signatures)
  > new instrumental signatures? bright source artifact/ghost masking
- Reference image generation (co-addition)
- Source-matching & photometric corrections for lightcurves
- Image subtraction, event extraction, assoc. & alert packet generation
  > small algorithmic updates: registration-error term in ZOGY
  > train machine-learned classifiers for point sources & streaks
  > updates to alert-packet schema; alert-IDs & naming
  > update S/G classification flags for PS1 in DB with ML probabilities
- Moving-object track-generation from event linking
- Pipeline executive; job scheduling/task orchestration: ingest to archive

Filesystem view-only web server

Ztf-depot for realtime products
- streaks, stamp-image cutouts
- moving-object tracks & metadata
- pipeline QA metrics / aggregate stats
- nightly survey / progress coverage maps
- server setup / user authentication

Temporary ztfweb for ad-hoc products
- view to pipeline/archive filesystem only
- select products to support analyses
- file listings only, no search GUI.
- user authentication

Long-term archive with search GUIs, APIs & user tools
- image and catalog **file** product server (single epoch & co-adds)
- moving-object search tool (MOST)
- lightcurve retrieval service, time-series viewer/analyzer
- on-demand finder-chart service
- future: forced photometry service operating on archive diff images

Alert Distribution System (ZADS)
- Kafka cluster (producer) to mirror w/ UW
- transmission/scheduling of alert packets

Data System schedule and activities

- **Black => Data System activities / milestones**
- ★ **Green => Key project dates**

- **Aug 17, 2017:**
  - raw camera image data / pipeline interface finalized
  - data-transfer to ingest and pipeline triggering and rules finalized
- **Aug 25, 2017:**
  - Data System needs (experiments) from commissioning finalized
  - Observing System needs from Data System finalized
  - user-access to *ztf-depot* and *ztf-web* file-based webservers
  - ad-hoc/algorithmic updates in pipelines implemented
- ★ **Sep 6, 2017:** first light (start of engineering commissioning)
  - *some* calibrations / detector parameters for pipeline (gains, distortion, astrometry parameters, bad pixel masks)
  - characterize detectors; identify new artifacts and devise correction/masking methods for pipeline **(need help)**
- ★ **Sep 18, 2017:** flat-field illuminator ship (latest)
  - high spatial-frequency responsivity maps (flats) and analysis using flat-screen images
- **Sep 20, 2017:**
  - moving object pipeline (point source event linking) in place and operational in test
  - update PS1 table in operations DB with ML-based S/G classification probabilities
- ★ **Oct 2, 2017:** start of science validation **(…see next slide)**

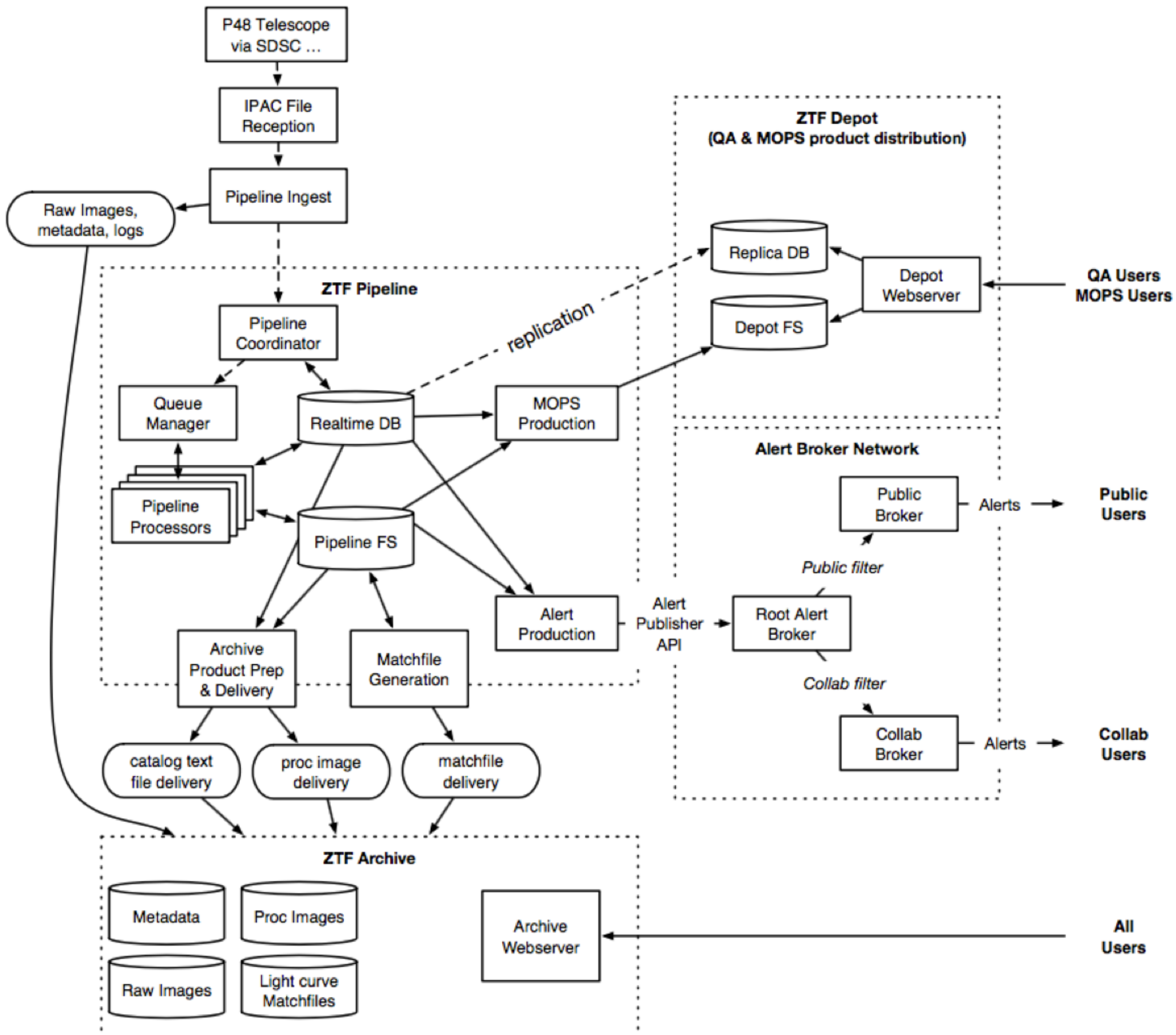
Data System schedule and activities

- **Black => Data System activities / milestones**
- ★ **Green => Key project dates**

- ★ **Oct 2, 2017:** start of science validation
  - user-interfaces to retrieve image/catalog file products from archive in place
  - pipeline tuning & optimization
  - optimal depth for reference images per filter
  - full *routine* calibration generation and analysis for pipeline input (lo/hi-ν flats, biases, final masks …)
  - exercise/refine daily operational routines (P48 – IPAC – archive – customers – points of contact)
  - performance analyses: astrometric/photometric precision, depth, FWHM variation **(need help)**
  - train ML classifiers (point sources & streaks)
- **Nov 30, 2017:**
  - user-interface to search and retrieve lightcurves from match-files
  - integration of Moving Object Search Tool
- **Dec 10, 2017:**
  - alert-packet schema finalized with "naming mechanism" in place
  - Kafka "producer" interfacing / mirroring with UW
- ★ **Jan 2018:** start of science survey
  - finder-chart service interfacing with archive in place and operational
- ★ **Apr 1, 2018:** public alerts commence
- **May 2018:**
  - forced photometry service operational on archived difference images
- ★ **Jan 2019:** first public data release of MSIP products (TBD)

Back up slides

More detailed data flow

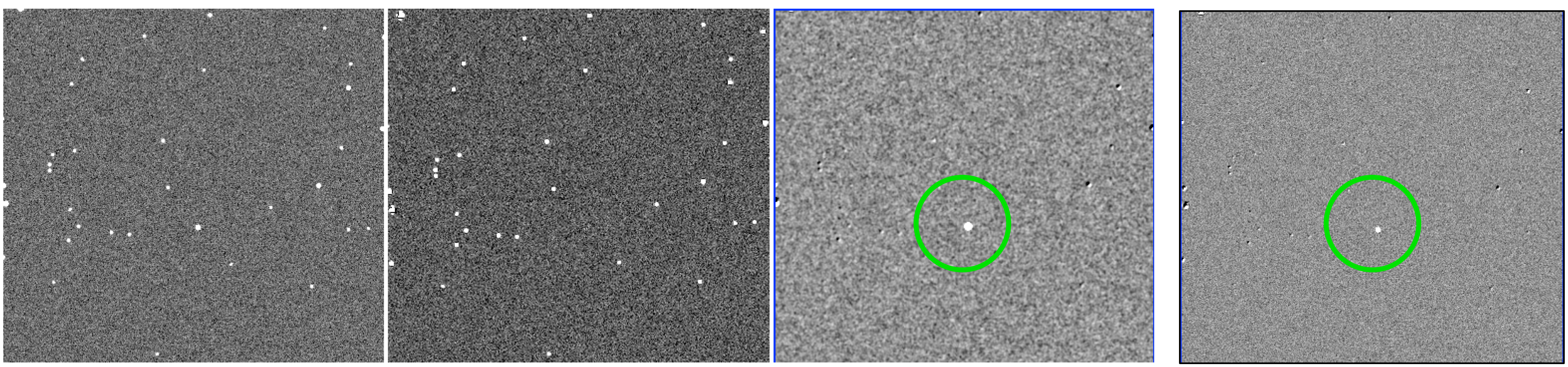


Implementation of ZOGY in image-subtraction pipeline

- ZOGY method: Zackay, Ofek, Gal-Yam (arXiv:1601.02655)
- First version implemented by Brad Cenko in Python. Uses pre-regularized image inputs.
- Parameter free! Optimality criterion: maximize S/N for point-source detection in sub-image.
  - Generates a “Scorr” (matched-filtered S/N) image for optimal point-source detection
  - de-correlates the pixel noise in subtraction image used for photometry
  - also generates an estimate of the effective PSF for the sub-image.

Products from simulated images:

New image | **Reference image** | **S/N (Scorr) image** | **Sub-image**



used for detection | *used for photometry*

PTFIDE versus ZOGY on iPTF data

- Adapted ZTF image-subtraction pipeline (that executes Brad Cenko's Python implementation of ZOGY) to process PTF image data
- ZTF pipeline then applies filters to raw ZOGY detections (research by Frank).
- Experimented on 6 iPTF fields containing transients discovered from ToO on event GW150914

The Astrophysical Journal Letters, 824:L24 (9pp), 2016 June 20

CrossMark

iPTF SEARCH FOR AN OPTICAL COUNTERPART TO GRAVITATIONAL-WAVE TRANSIENT GW150914 …

M. M. Kasliwal[1], S. B. Cenko[2,3], L. P. Singer[2,20], A. Corsi[4], Y. Cao[1], T. Barlow[1], V. Bhalerao[5], E. Bellm[1], D. Cook[1],

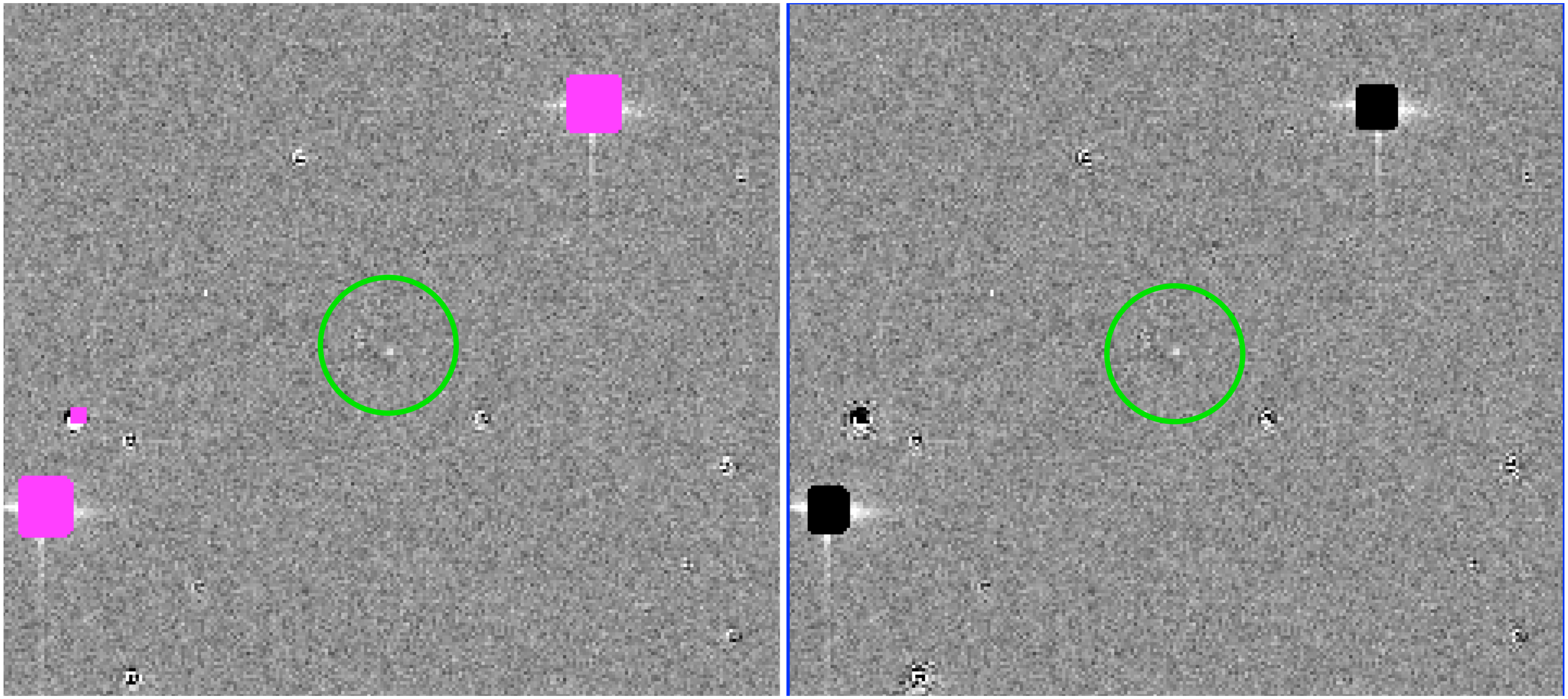
	Name	RA (J2000)	DEC (J000)	Discovery Time
→	iPTF15cyo	$8^{h}\,19^{m}\,56^{s}.18$	$+13$ d $52'\,42''.0$	2015 Sep 17 05:54:55.6
	iPTF15cyp	$8^{h}\,21^{m}\,43^{s}.68$	$+16$ d $12'\,42''.0$	2015 Sep 17 05:56:31.6
→	iPTF15cys	$8^{h}\,11^{m}\,55^{s}.59$	$+16$ d $43'\,10''.1$	2015 Sep 17 06:05:16.6
→	iPTF15cym	$7^{h}\,52^{m}\,35^{s}.67$	$+16$ d $45'\,59''.6$	2015 Sep 17 05:46:17.1
→	iPTF15cyq	$8^{h}\,10^{m}\,00^{s}.86$	$+18$ d $42'\,18''.1$	2015 Sep 17 05:57:16.3
→	iPTF15cyn	$7^{h}\,59^{m}\,14^{s}.93$	$+18$ d $12'\,54''.9$	2015 Sep 17 05:47:20.5
	iPTF15cyt	$7^{h}\,38^{m}\,59^{s}.35$	$+21$ d $45'\,43''.2$	2015 Sep 17 06:08:09.3
→	iPTF15cyk	$7^{h}\,42^{m}\,14^{s}.87$	$+20$ d $36'\,43''.4$	2015 Sep 17 05:38:38.3

PTFIDE versus ZOGY on iPTF data

iPTF15cyk

ZOGY

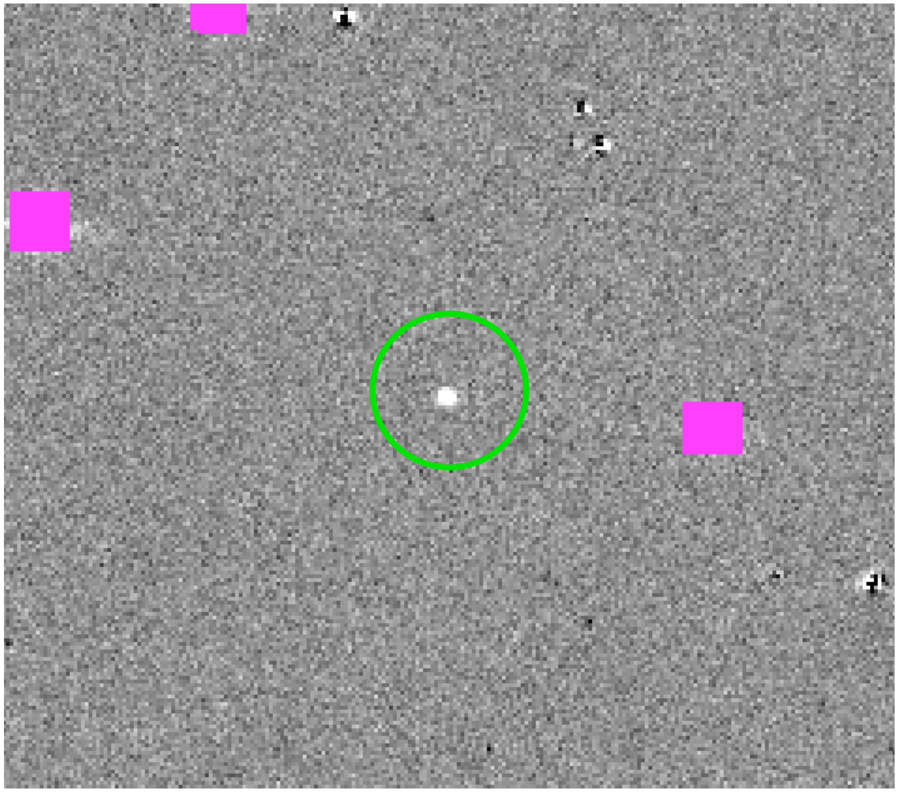
PTFIDE



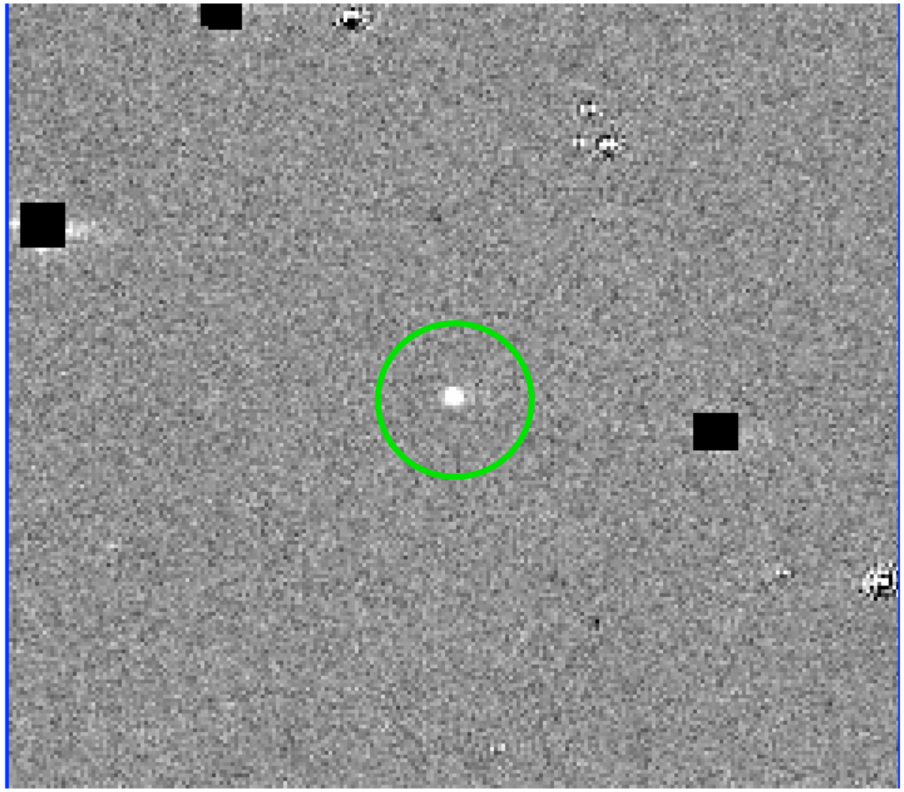
PTFIDE versus ZOGY on iPTF data

iPTF15cyo

ZOGY



PTFIDE



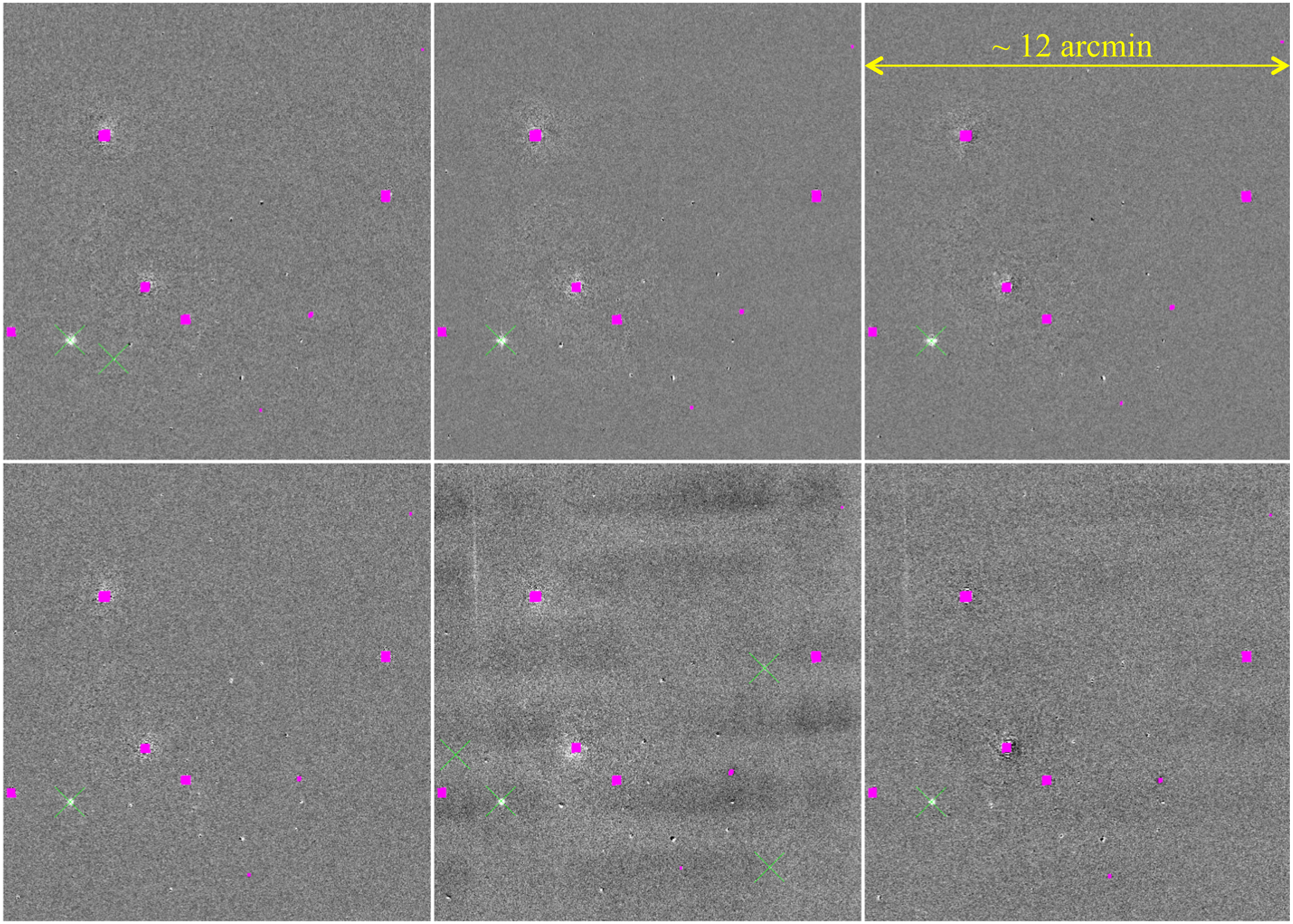
PTFIDE vs ZOGY: summary statistics

- Number of **raw candidates** extracted to S/N = 5.
- Following ZOGY, we use simple PSF-shape/morphology & local pixel filters to remove obvious false-positives; **no machine-learned (RB) vetting here.**

real transient	Field/CCD	#candidates (PTFIDE)	#candidates (ZOGY + *filt*)	#asteroids
iPTFcyk	3658 / 8	181	5	2
iPTFcym	3459 / 6	472	6	1
iPTFcyn	3560 / 7	343	10	7
iPTFcyo	3359 / 8	268	4	3
iPTFcyq	3561 / 6	210	6	2
iPTFcys	3460 / 9	350	11	4

- **NOTES:**
  - same archival PTF reference image co-adds were used in PTFIDE and ZOGY subtractions, created using an old/non-optimal method --- will be different for ZTF
  - PTF epochal images used old astrometric calibration method --- will also be different for ZTF

M51 with SN 2011dh (from PTF): ZTF pipeline with ZOGY + *filtering*



PTFIDE versus ZOGY

- **Conclusion:** PTFIDE and ZOGY appear to show similar performance on PTF data, **at the raw level (with no filtering)**, noting the non-optimal calibrations upstream.

- ZOGY with *simple* filtering of raw candidates is better!

- This exercise shows that **raw** difference-image quality is primarily driven by quality of upstream calibrations (systematics): astrometry, flat-fielding, gain-matching, PSF-estimation.

- Upstream calibrations must be accurate before one starts to benefit from the *statistical*-optimality property underlying ZOGY, i.e., maximum point-source S/N in limit of background dominated noise

ZOGY caveats and limitations

- From discussions with the LSST DIA working group (David Reiss & Robert Lupton).
- Crucial inputs to ZOGY are prior estimates of the PSF for the new and reference images.
- These must be as accurate as possible to avoid systematics in the difference-image products.
- Currently, ZTF pipeline automatically derives PSFs on a readout-channel basis (~ 0.65 deg^2).

- Two limiting cases **will** present a challenge:
  - fields containing very few stars, or a sufficient number of bright enough stars.
  - very dense fields, approaching galactic-plane densities with high source confusion.

- PSF estimation in ZTF pipeline uses an updated version of *DAOPhot*, with iterative de-blending. Very robust process with quality metrics generated.
- Limitations need to be explored and quantified.
- Current ZTF simulations seeded by PS1 do indeed show problems at the above extremes.

Number of transient candidates (PTF vs. ZTF)

- **PTF experience:**

  Raw transient stream: ~ 200 – 300 candidates per image (chip).

  ↓

  Machine-learned RB vetting, ~ five to ten(s) *likely real* candidates per image; all transient flavors;

  with ~ 250 PTF exposures/night × 11 chips × 20 candidates/chip, ~ 55,000 candidates/night.

  ↓

  Marshal automated-vetting for specific science cases, e.g., ≥ 2 detections in night, etc.

- **Expectation for ZTF:**

  Very raw transient stream (no filtering): <~ 150/image ?

  ↓

  Simple filtering on candidate metrics, ~ ten(s) *likely real* candidates per image;

  with ~ 700 PTF exposures/night × 64 images × 20? candidates/image, <~ 890,000 candidates/night.

  ↓

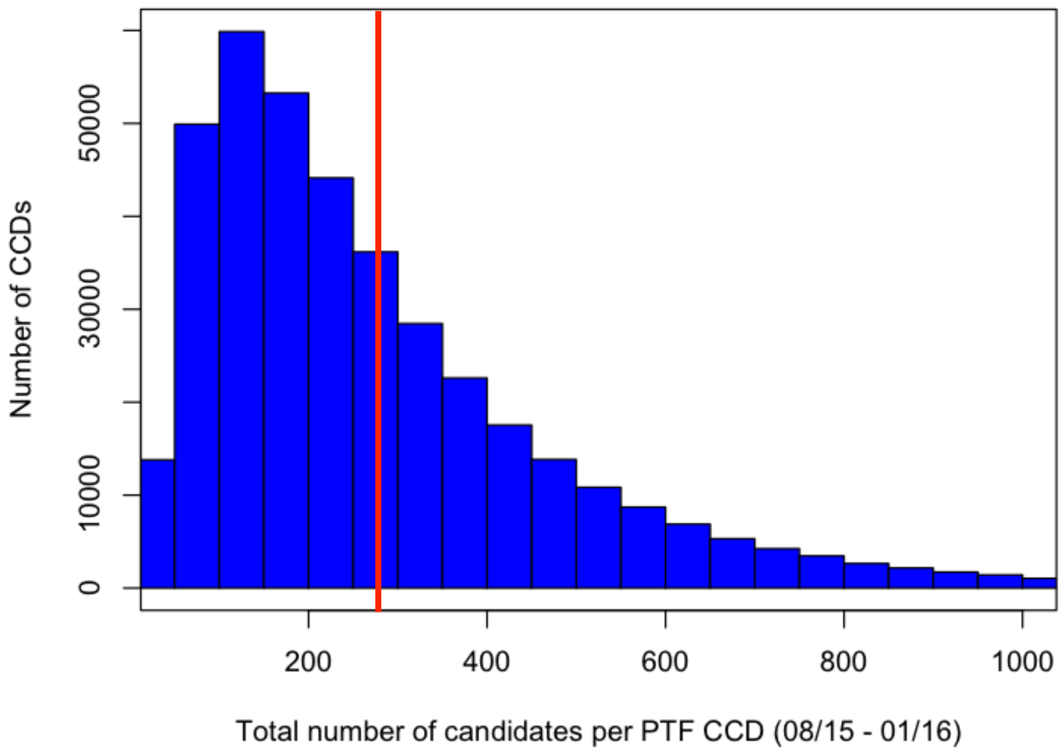
  Automated (machine-learned) vetting in pipeline is likely to reduce the above nightly count.

  ↓

  Alert packets sent to broker for further filtering based on specific science-use cases.

Number of (raw) transient candidates

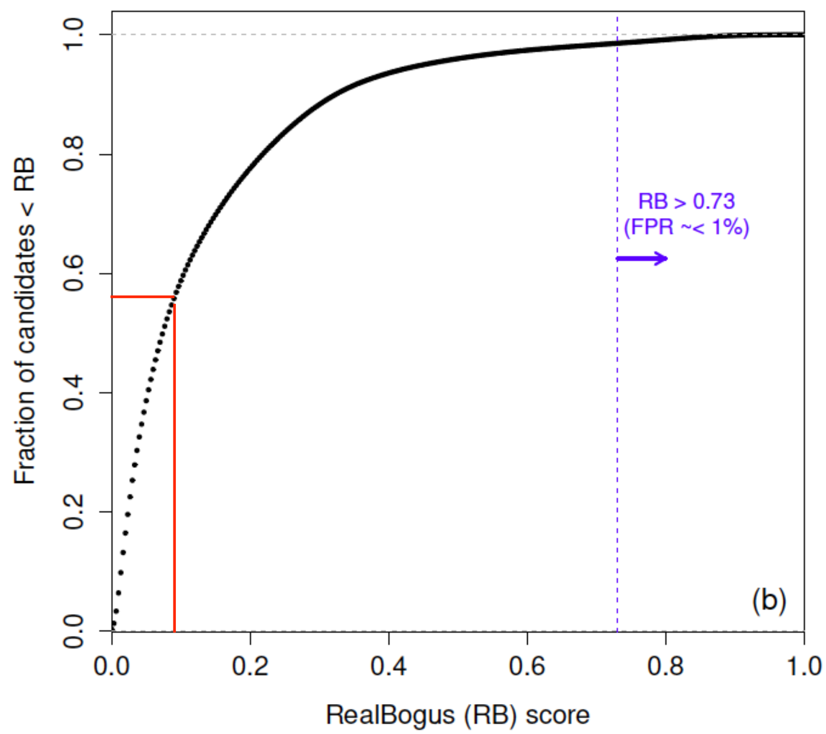
- From **PTF**, encounter ~ 260 raw, **non** machine-learned vetted candidates per CCD at > 4σ using PTFIDE.
- One ZTF CCD readout quadrant covers ~ one PTF CCD + ~ 10%. Hence we can extrapolate to ZTF.
- Have ~ 700 exposures * 64 readout quads: ~ 44,800 positive subtractions per night on average.
- Implies **~ 13 million transient raw candidates** per night for ZTF. Includes all transients (+ variables + asteroids)



or ~ per ZTF readout quadrant

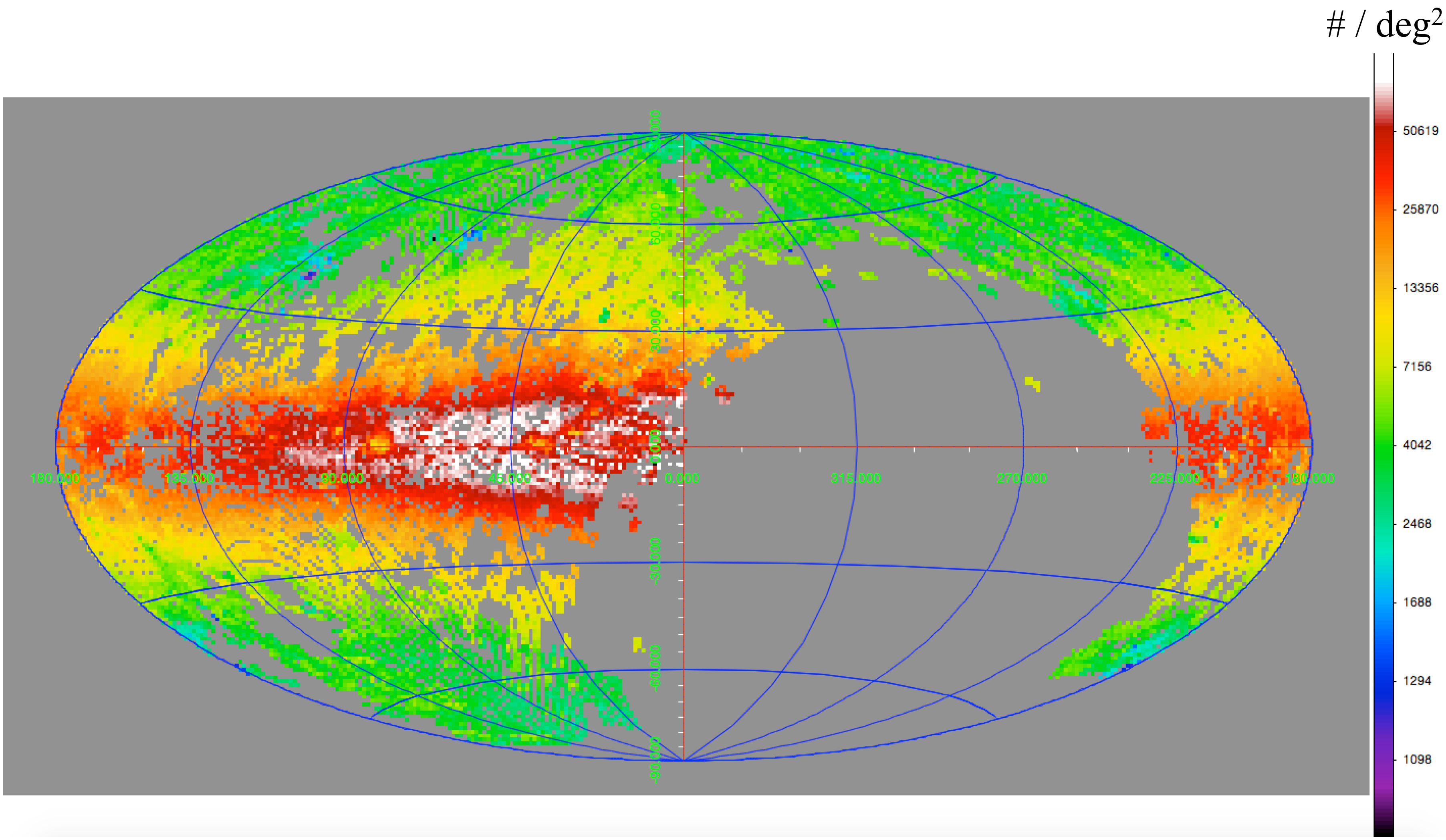
Benefit of Machine Learning

- Use the *RealBogus* (RB) quality score from a machine-learned classifier: crucial for PTF (down to 4σ).
- If avoid everything with a RB score < 0.1, only need to store ~ 6 million candidates per night in DB for ZTF.
- If use RB > 0.73 (< 1% false-positive rate) found for PTFIDE subtractions, need to scan <~ 400,000 cands/night.
- Translates to <~ 10 candidates per ZTF quadrant image or <~ 14 candidates/deg^2 on average (<u>all transients</u>).

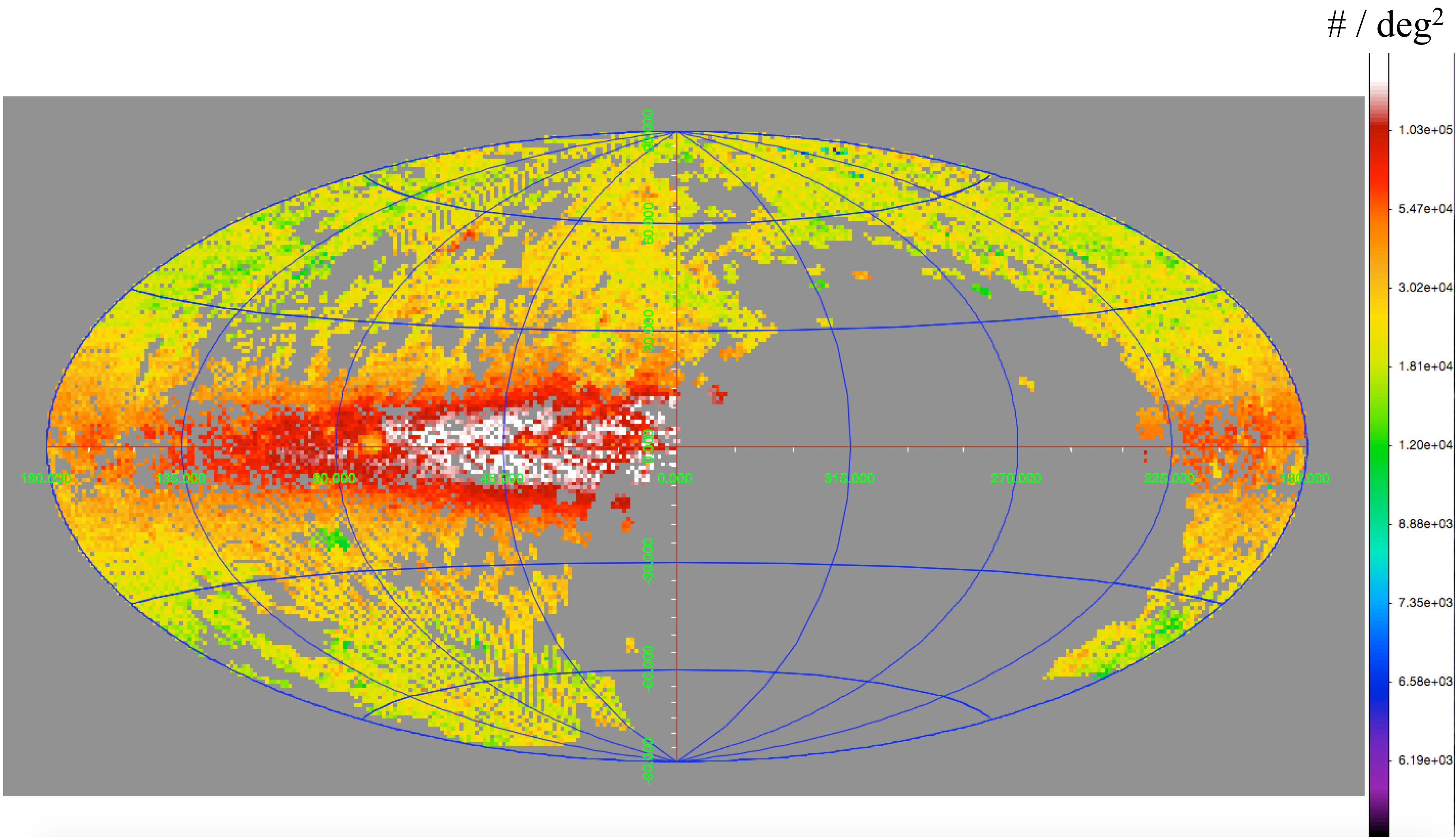


Cumulative fraction of transient candidates versus RB score from ~22,000 PTFIDE subtractions (Masci et al. 2016).

Density of aperture (SExtractor) extractions from PTF CCDs

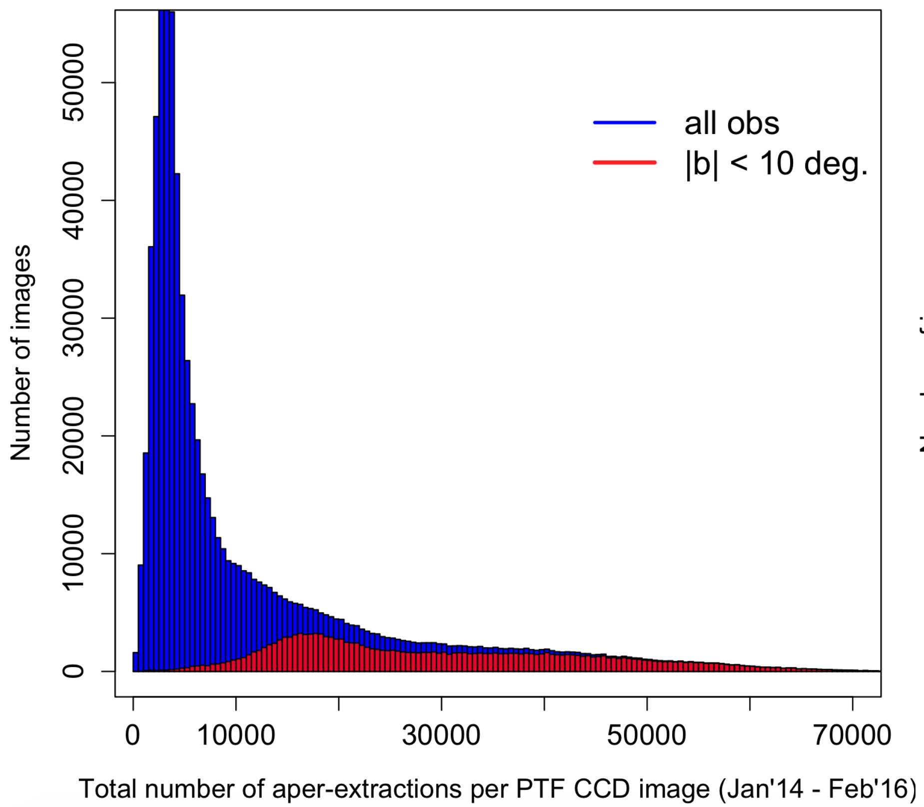


Density of PSF-fit extractions from PTF CCDs

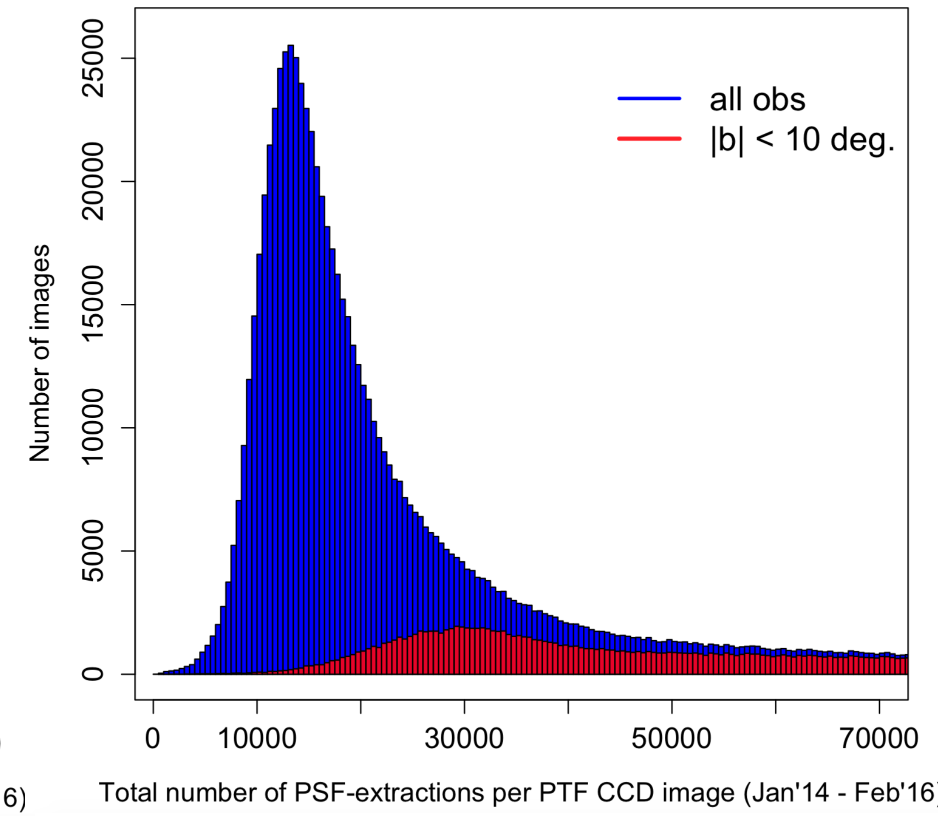


Number of sources extracted from PTF CCDs

Aperture (SExtractor)



PSF-fitting (DAOPhot)

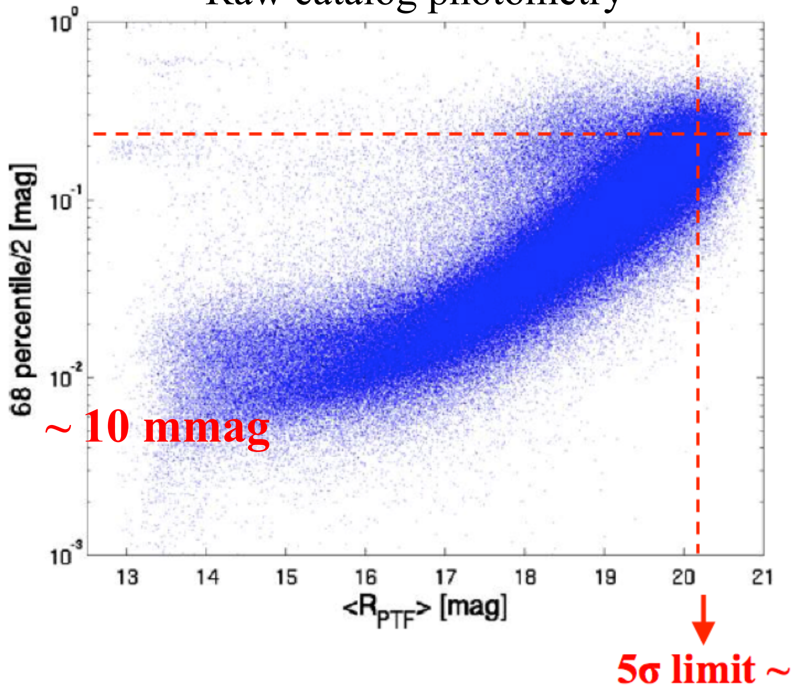


ZTF Lightcurve Pipeline

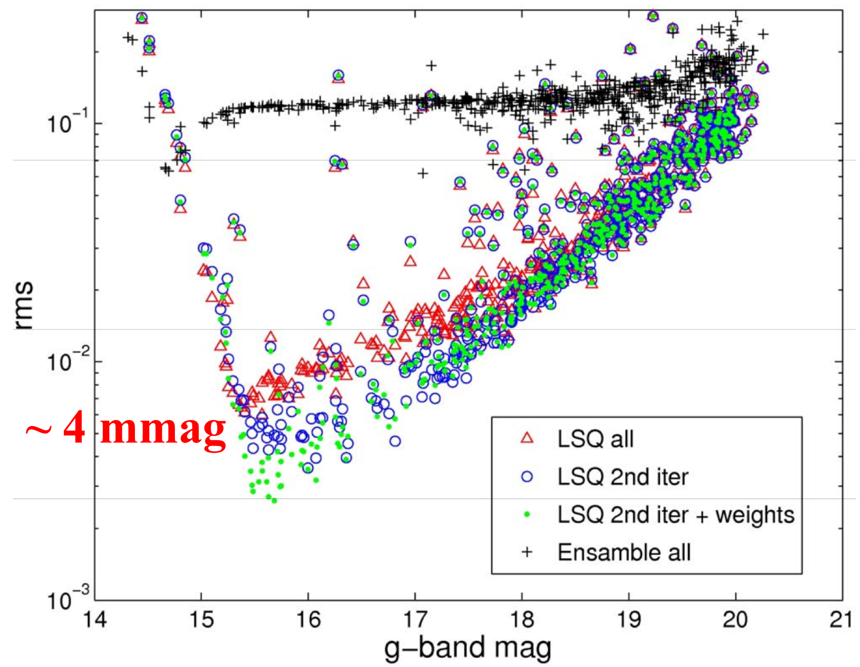
- All sources detected in epochal images are matched against the reference-image source catalog for a given field, CCD image quadrant, and filter
- The “cleanest” least variable sources are used as anchors for the relative photometric calibration
- Individual image gain-correction factors are computed using a least-squares fitting method
- These gain-correction factors are applied the image photometric zero-points
- The refined zero-points improve relative photometry to a few millimag for bright sources
- This pipeline will be triggered on timescales of typically 2 to 3 weeks (TBD)
- All lightcurves for a single CCD image quadrant and filter are stored in a “matchfile” (hdf5)
- Accompanying each lightcurve is a set of >100 metrics: RMSs, Skews, Stetson indices …
- All lightcurves and metrics are seeded by an object ID; these IDs with positions are loaded into a database to support spatial searches; associated lightcurve is retrieved from the “matchfile”
- Expect of order 1.3 billion objects (individual lightcurves) for ZTF

PTF Photometric Performance (internal)

Raw catalog photometry



Refined (relative) photometry for lightcurve DB

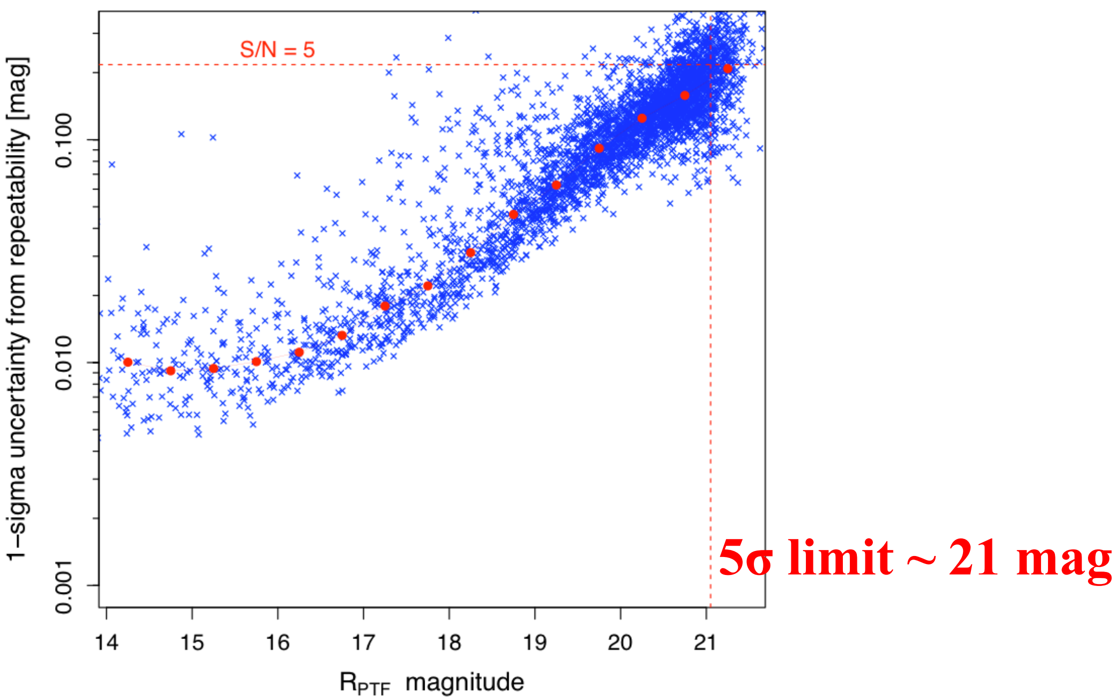


Basic Photometric Calibration

- Photometric calibration will be performed with respect to an external catalog (e.g., Pan-STARRS1) using PSF-fit extractions on a readout-quadrant image basis:

$$m_i^{PS} - m_i^{ZTF} = ZP + b\left(g_i^{PS} - R_i^{PS}\right) + \varepsilon_i \quad \Rightarrow \quad \text{solve for } ZP,\, b \text{ per image}$$

- Expect an *absolute* precision of ~ 2 - 3%.
- *Relative* photometric precision using PSF-fitting on PTF images ~ 1% (no refinement of ZPs across epochs)
  - ➢ Biggest limitation is flat-fielding!



ZTF real-time processing throughput (naïve estimate)

- Incoming data rate (set by cadence):
  - one exposure or 64 quadrant images / 45 sec.
  - *inprate* ~ 85 quad images / minute, on average

- Processing rate (median as of today):
  - *outrate* ~ 1 quad image / 6.5 minutes / CPU core

- If processing was purely CPU-limited, no or negligible I/O latency, *minimum* number of CPU cores needed to keep up with input data rate is:

  $N_{cores} = inprate \, / \, outrate = 553$ cores

- This estimate is naïve since it ignores I/O, network speed, other interleaved processing tasks. Goal is to process faster than incoming data rate.

- Our currently "active" ZTF compute cluster has 16 physical cores × 32 nodes = 512 cores (or 16 × 2 × 32 = 1024 admissible simultaneous threads, contingent on shared resources)

Data System staff functions

Task Management

Coordinate data system activities; report to ZTF project manager; coordinate with project scientist.

Pipeline Development

Design, prototype, and implement pipeline software modules; integrate into production system, interfacing with database and archive

Archive Development

Scripts to transfer ZTF data to formats consistent with permanent storage at IRSA; adaptation of IRSA user interfaces for ZTF-special capabilities.

Simulation, Analysis, and Performance Monitoring

Provide synthetic data sets for pipeline testing; analysis of instrument performance; evaluate instrument calibration and stability with recommendations for pipeline tuning/ optimization during survey (works with Pipeline Developer)

Database Administration

Design ZTF schema; optimize database tables for quick access (both on archive and pipeline operations side); manage daily large database merges for new data.

PTF / iPTF Operations

Complete ingest and processing of PTF data; requested bulk reprocessing to support future public release

Pipeline Operations and Archive Ingest:

Execute and monitor pipelines; periodic transfers of raw and processed data to archive; daily cleanup/set-up for forthcoming night; schedule and manage reprocessing requests as needed.

Datacenter Operations

Procure and install all ZTF-related hardware: compute drones, storage arrays and disks, network gear; offsite backups; environment monitoring.