ZTF Real-Time Pipeline

Frank Masci

ZTF ML working group Meeting, July 2017

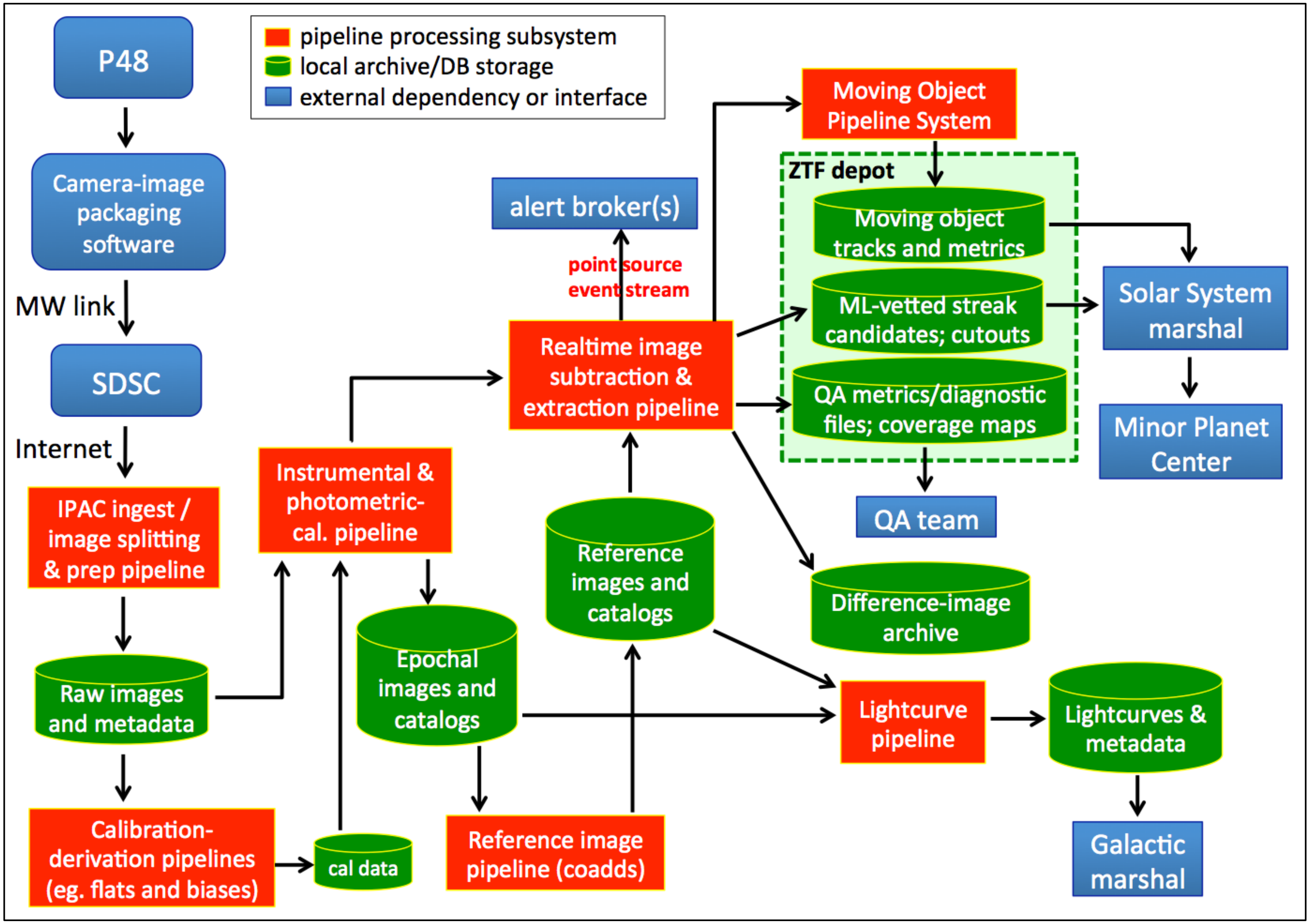
ZTF Real-time pipeline

- Does most of the heavy-lifting in real-time: to support near real-time (fast-response) science
- **Timing requirements:**
  - > 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)

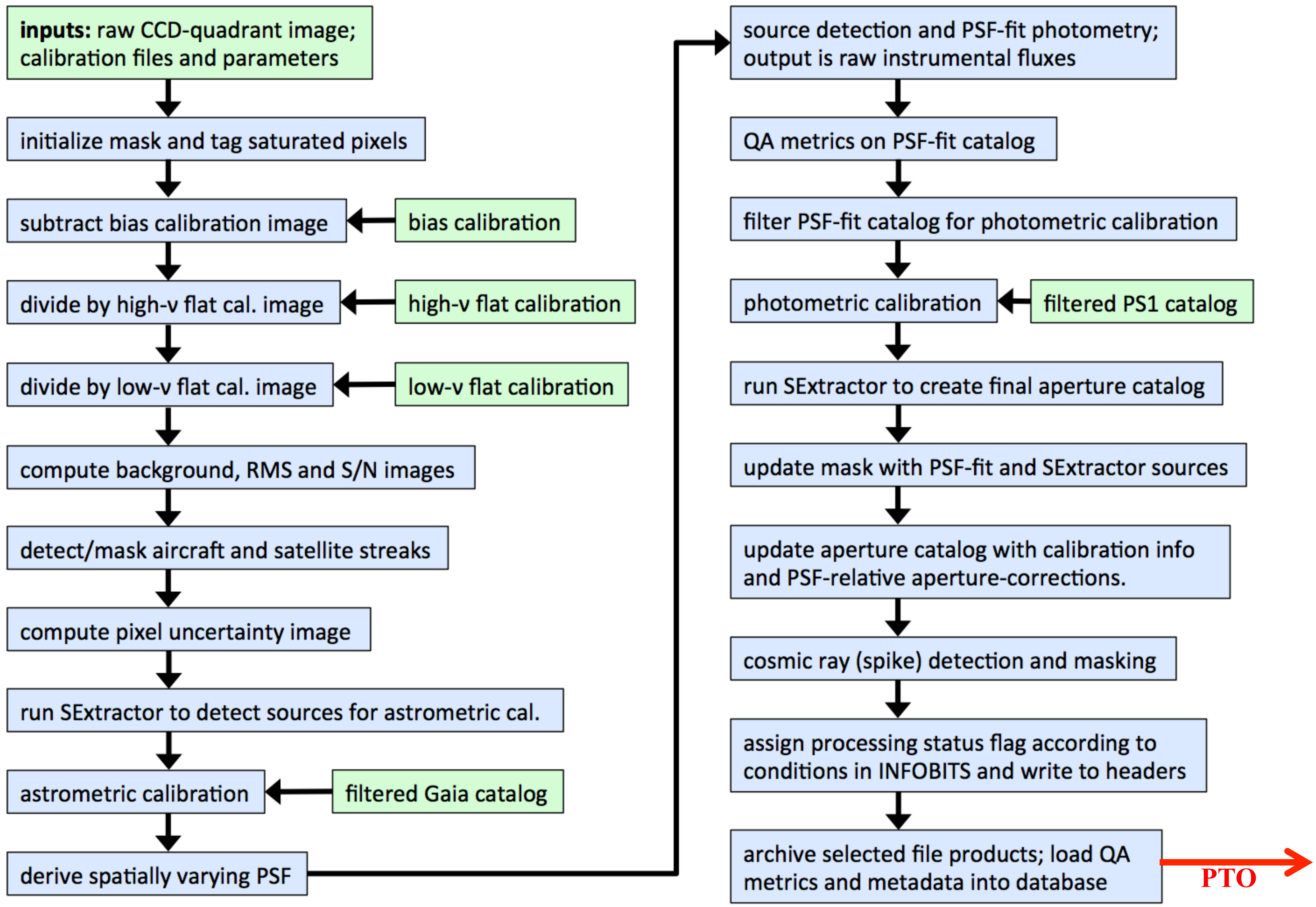
- Real-time pipeline consists of two phases:
  1. **Instrumental calibration** (bias-corrections, flat-fielding, astrometry, photometric calibration, masking of bad pixels, …): generates single-epoch image and catalog products for archive.
  2. **Image subtraction & extraction** of transient events (point-sources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

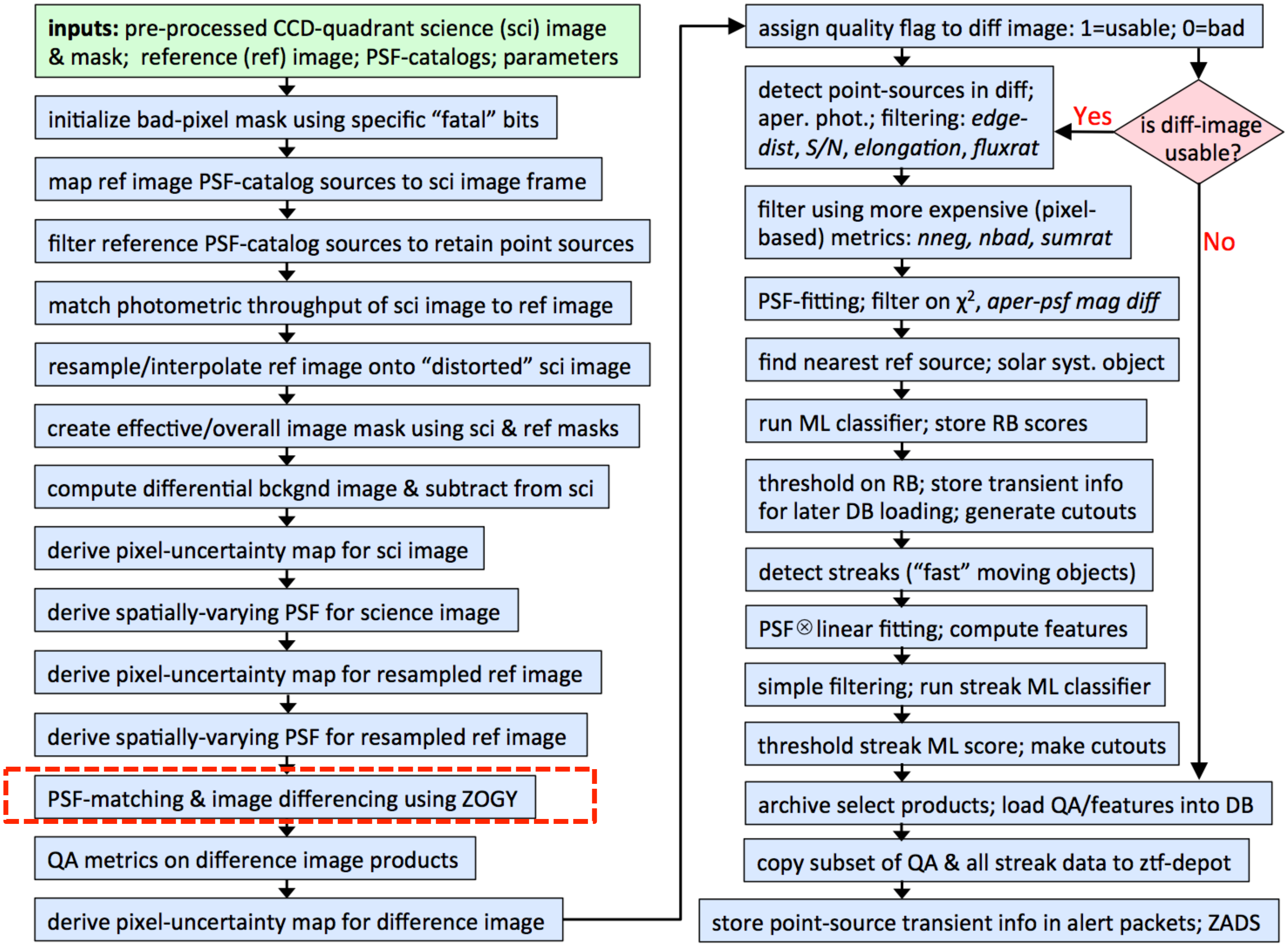
Overall processing & data flow



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



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



Initial goal for ZTF (automated vetting)

- Strive for reliability (purity).
  - For point-source transients, tune machine-learned vetting to a maximum tolerable false positive rate of 0.1% (?), for a given raw S/N extraction threshold (TBD)

- At expense of sacrificing completeness (inevitable)
  - Tuning: strive to minimize corresponding false negative rate (incompleteness) at above FPR

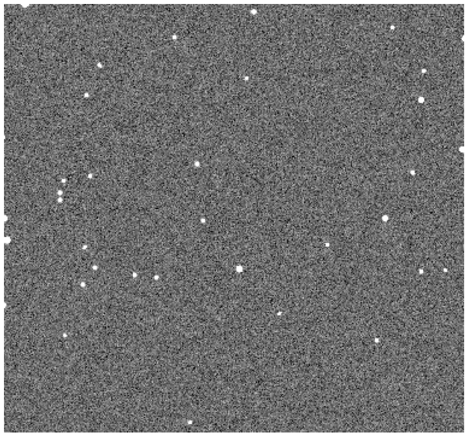
- Will depend on sky location, other unforeseen survey/stochastic variables not in training model

Implementation of ZOGY in image-subtraction pipeline

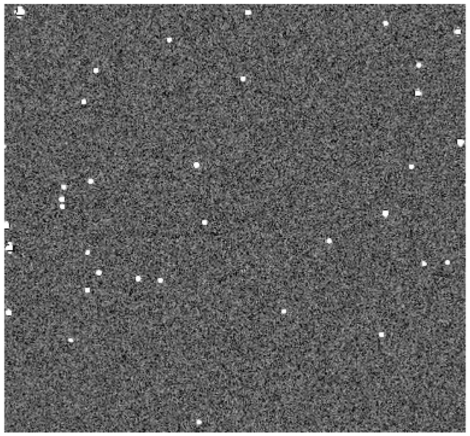
- ZOGY method: Zackay, Ofek, Gal-Yam (arXiv:1601.02655)
- First version implemented by Brad Cenko in Python. Uses pre-regularized image inputs.
- Parameter free! Optimality criterion: maximize S/N for point-source detection in sub-image.
  - Generates a “Scorr” (matched-filtered S/N) image for optimal 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:

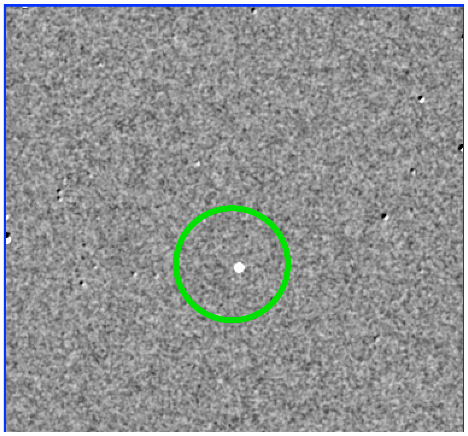
New image



Reference image

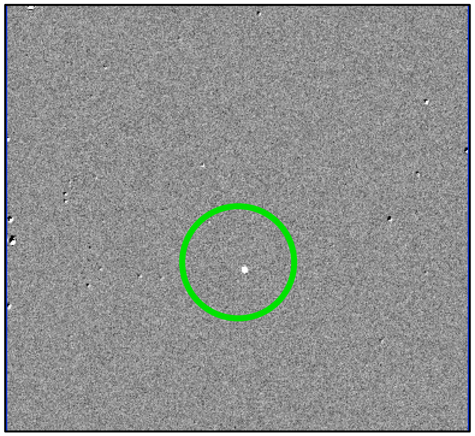


S/N (Scorr) image



used for detection

Sub-image



used for photometry

PTFIDE versus ZOGY

- **PTFIDE optimality criterion:** derive best PSF-convolution kernel κ by minimizing a weighted sum of the squared residuals between a model (reference) image I_2 and new image I_1.

Difference image in Fourier space:

$$D(k) = \left[I_1(k) - \kappa(k) I_2(k)\right]$$

Masci et al. 2016:

http://iopscience.iop.org/article/10.1088/1538-3873/129/971/014002/meta

- **ZOGY optimality criterion:** maximize the S/N (likelihood) for point-source detection in difference images, assuming images are dominated by uncorrelated Gaussian noise.

$$D(k) = \left[I_1(k) - \kappa(k) I_2(k)\right] \sqrt{\frac{\overline{\sigma}_1^2 + \overline{\sigma}_2^2}{\overline{\sigma}_1^2 + \kappa^2(k) \overline{\sigma}_2^2}}$$

Zackay et al. 2016:

http://iopscience.iop.org/article/10.3847/0004-637X/830/1/27/pdf

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

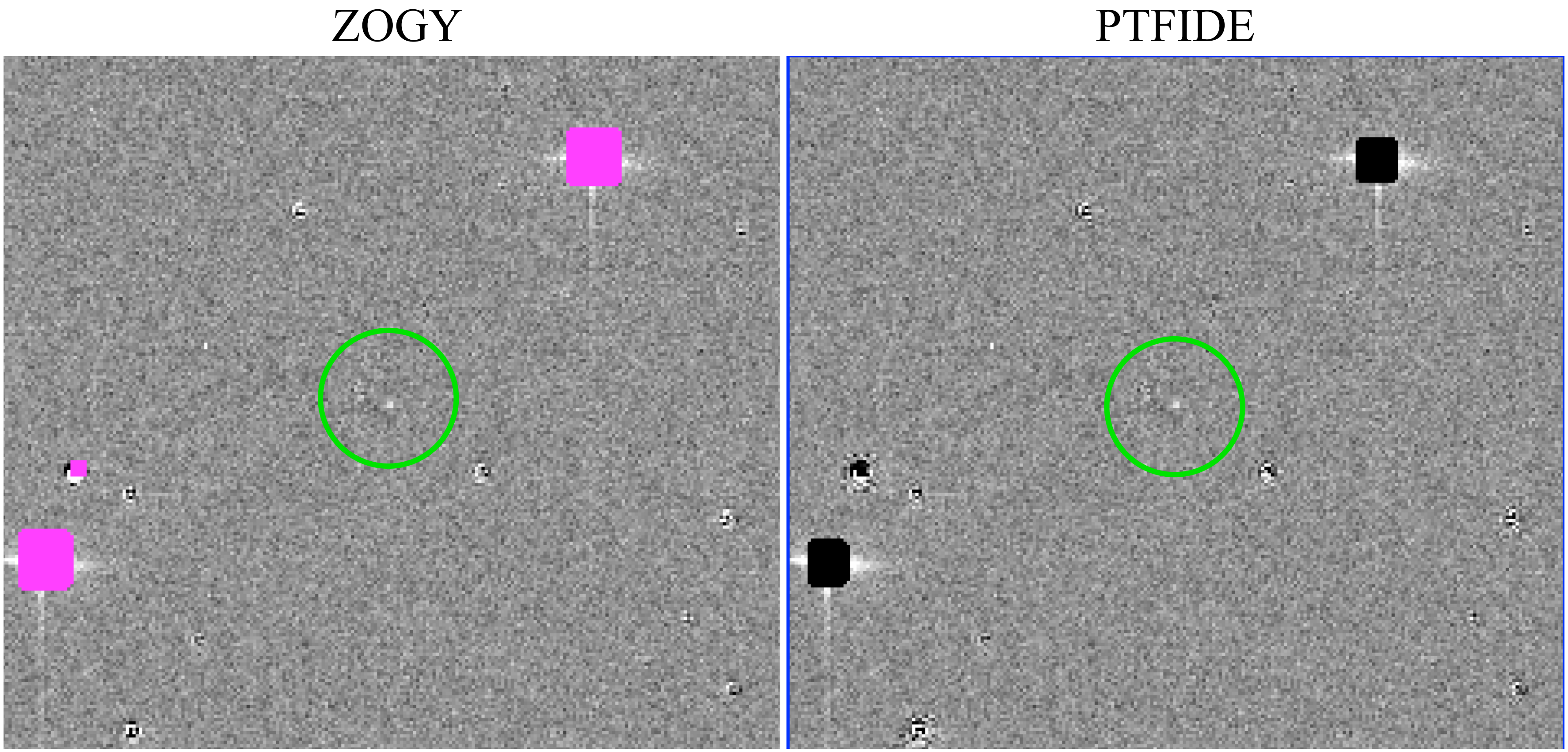
iPTF SEARCH FOR AN OPTICAL COUNTERPART TO GRAVITATIONAL-WAVE TRANSIENT GW150914 …

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

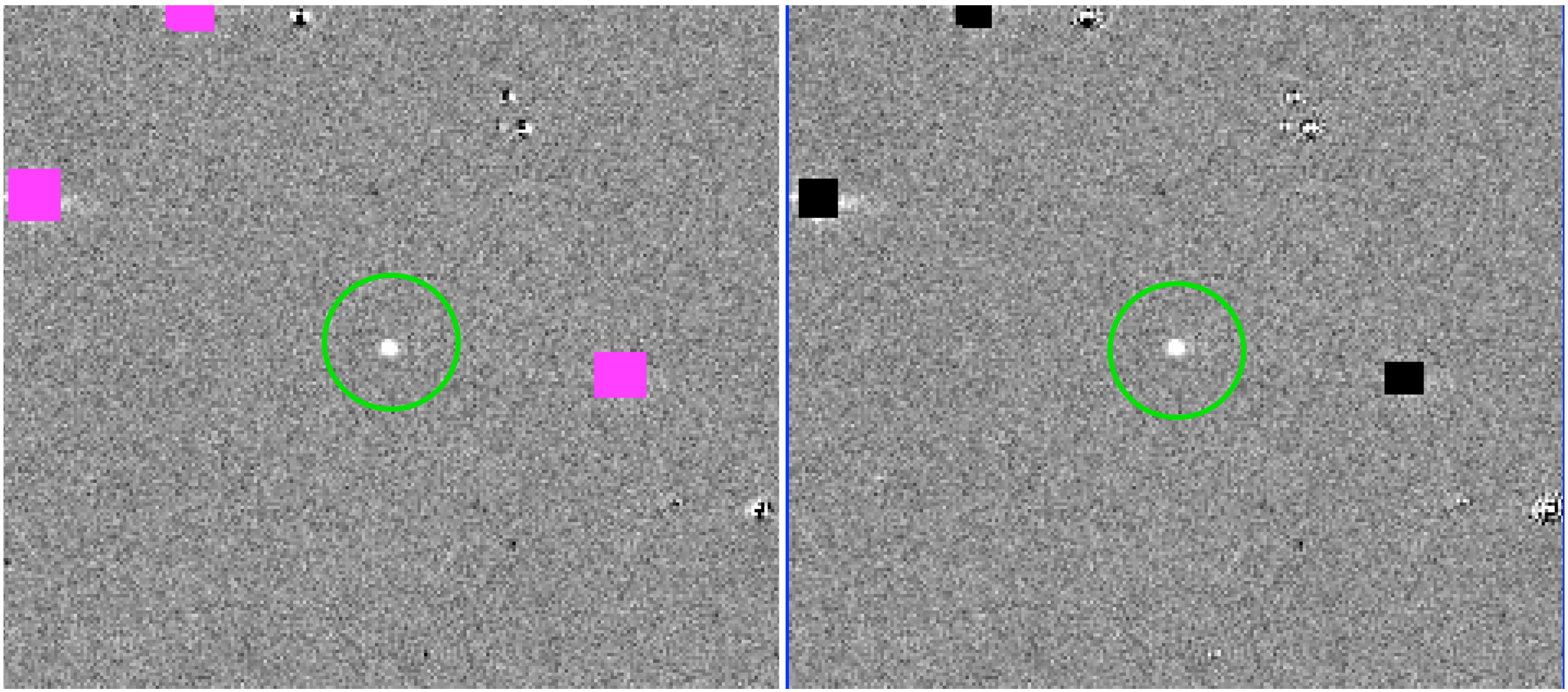
iPTF15cyk



PTFIDE versus ZOGY on iPTF data

iPTF15cyo

ZOGY PTFIDE



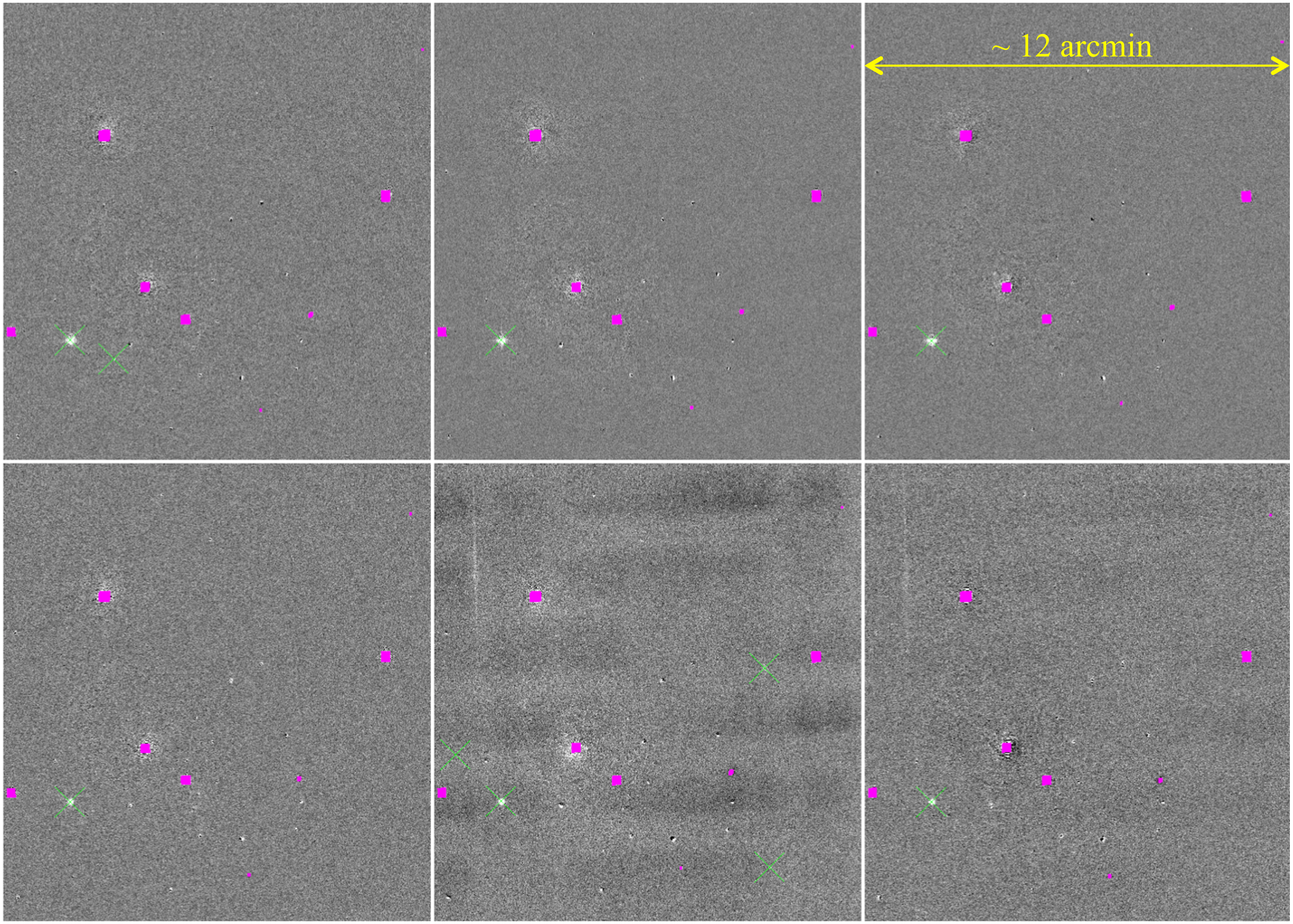
PTFIDE vs ZOGY: summary statistics

- Number of **<u>raw candidates</u>** 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.

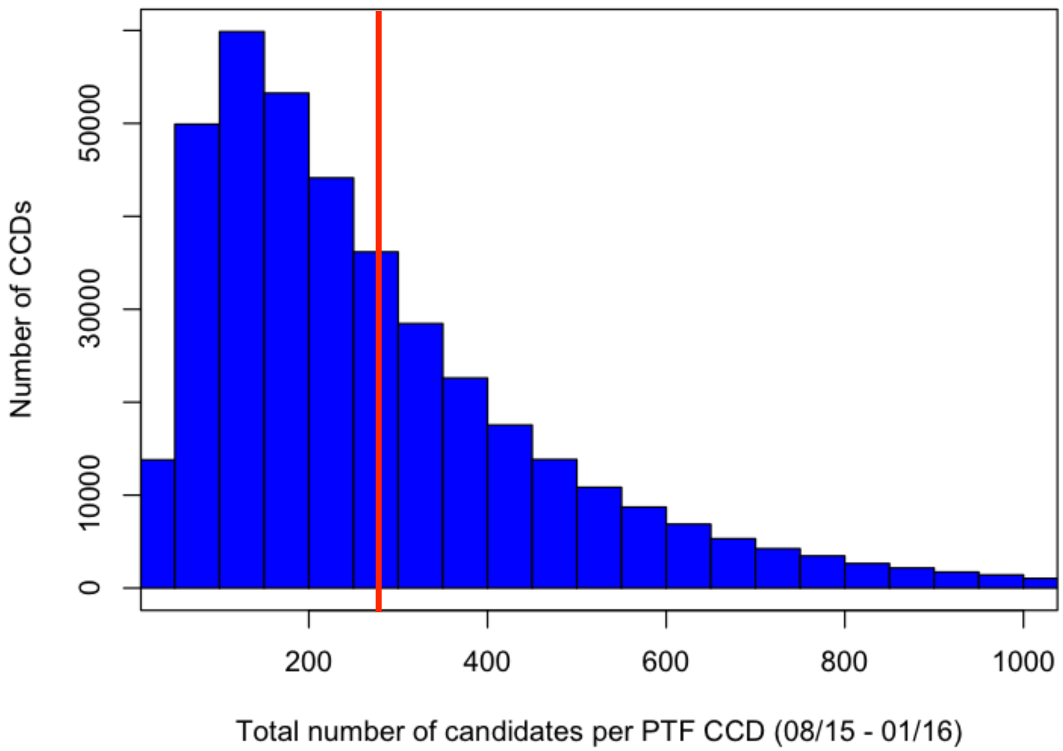
  ↓

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

Back up slides

Number of (raw) transient candidates

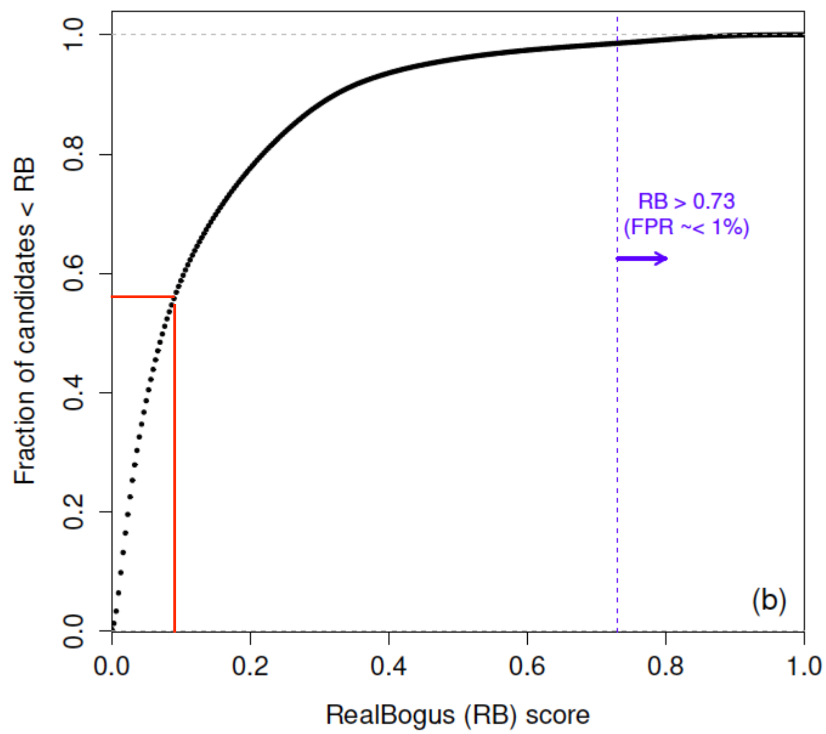
- From **PTF**, encounter ~ 260 raw, **non** machine-learned vetted candidates per CCD at > 4σ using PTFIDE.
- One ZTF CCD readout quadrant covers ~ one PTF CCD + ~ 10%. Hence we can extrapolate to ZTF.
- Have ~ 700 exposures * 64 readout quads: ~ 44,800 positive subtractions per night on average.
-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