

ZTF Development Completion Review Science Data System

Frank Masci, Ben Rusholme, Steve Groom

& the IPAC-Caltech ZTF Team

November 20, 2018



High level Data System Objectives & Requirements

From the MSIP (NSF) proposal & ZTF Management Plan (03 / 12 / 2014):

- Sustain processing & storage for three years of operations (initially: 2017 – 2019)
- Scale data processing and storage to sustain a 15 x PTF data-acquisition rate
- Generate data products similar to those as PTF and additionally:
 - Provide a lightcurve retrieval / search tool
 - Real-time transient alerts for public consumption, beginning in survey year three
- Support public release of archived products every six months, with the first occurring at the end of survey year one
- Process data and generate alerts at all galactic latitudes in the Northern visible sky

High level Data System Objectives & Requirements

More concrete specifications on products came together in a Science Requirements Doc. (E. Bellm; 12/16/2015):

- Data System development was guided primarily by this document.
- Two additional requirements presented a huge challenge:
 - Include a database of sources detected by PSF-fit photometry from each epochal image.
 - Maintain a database of lightcurves generated by positionally matching all epochal image extractions.
 - Not feasible following analyses on source statistics; instead, we devised a compromise for serving lightcurves.
- 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 published in < 10 min (goal: 5 min)

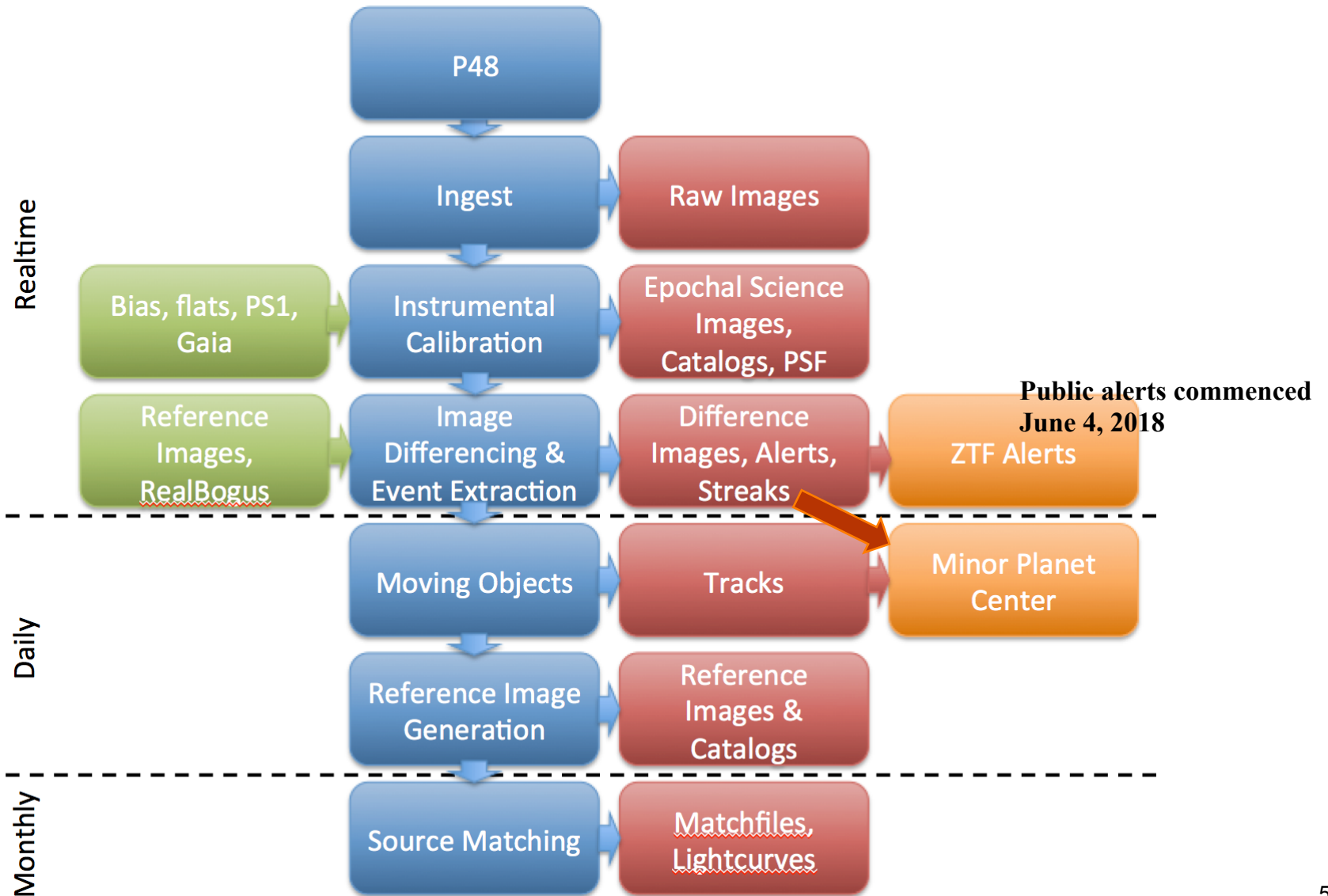
One requirement was updated in early 2017:

- Commence public alerts ~2.5 years earlier than planned: now in April 2018.
- DS was developed according to the above and best efforts to achieve best on-sky performance given the resources.

Data System Completion Checklist

- Production pipelines
- Baseline deliverables and products to enable the NSF-funded science (requirement)
- User interfaces and services for data access (requirement)
- Astrometric performance
- Photometric precision (repeatability)
- Photometric calibration accuracy
- Data processing latency (realtime alert production requirement)
- Public alert generation (requirement)
- Sky coverage for alert production (requirement)
- First Public Data Release (requirement)

All pipelines are in place & operational (timeline view graphic)



All Baseline Deliverables are complete; All Data Access portals are in place

1. **Instrumentally calibrated, epochal image products, bit-masks, source catalogs, PSFs, and difference images**
Archive (IRSA)
2. **Raw image data and image calibration products used in pipelines**
Archive (IRSA)
3. **Reference images (co-adds) from combining (1): coverage maps, uncertainty maps, and source catalogs**
Archive (IRSA)
4. **Alert (point-source event) stream from real-time image-differencing pipeline: packetized with metadata**
Kafka; Public Brokers; also archived in IRSA
5. **Products to support real-time Solar System / NEO discovery and characterization: both streaks and tracks**
ZTF-Depot (internal) and IAU-Minor Planet Center
6. **Lightcurves & metrics from matching sources across individual epochs using (1) to beginning of survey**
Archive (IRSA)
7. **Quality assurance metrics, summary statistics, and survey coverage maps: for performance monitoring**
ZTF-Depot (internal)
8. **Documentation: pipeline descriptions, recipes, and tutorials on data-retrieval and analysis**
Explanatory Supplement on ZTF Public Website; PASP paper in press

User interfaces to retrieve/analyze archive products are in place

- Search by position, timespan, image metadata, and object name; interactive manipulation; catalog overlays; visualization; analysis of lightcurves with periodogram service; cutouts on any image type.
- Accompanying APIs (command-line driven retrieval tools) are also available

Image viewer and file-product retrieval

IRSA DATA SETS SEARCH TOOLS HELP

Pixel Size: Eq-J2000: Image Pixel

PTF Image Lock By Click Background Monitor

Search by Position M83; Type=CENTER; Region=0.0083 deg; Image Size=0.2778 deg; Product Level=1

Level-1

Prepare Download 1 of 3 (1 - 50 of 117)

expid	obsdate	crval1	crval2	filter	ccdid	ptffield	seeing	a1
19123	2009-05-25 05:59:15.975000	204.5342815	-30.2159701	G	6	100006	3.36	2
19435	2009-05-27 04:31:09.425000	204.6302496	-29.9765794	R	6	100006	2.25	2
19499	2009-05-27 06:06:42.925000	204.6302877	-29.9764442	R	6	100006	2.44	2
51514	2009-12-30 13:29:56.358000	204.4431954	-30.6655180	R	6	100006	4.44	2
51651	2009-12-31 13:35:51.759000	204.4432088	-30.6654161	R	6	100006	3.90	2
51843	2010-01-01 13:20:49.858000	204.4434360	-30.6653985	R	6	100006	2.78	2
52350	2010-01-02 13:35:30.259000	204.4432805	-30.6654429	R	6	100006	3.11	2
52756	2010-01-03 13:14:24.108000	204.4432014	-30.6654362	R	6	100006	2.90	2
53520	2010-01-04 13:09:43.809000	204.3623019	-30.6769362	R	6	100006	null	2
53929	2010-01-05 13:05:51.809000	204.4432884	-30.6654451	R	6	100006	3.40	2
54294	2010-01-06 13:21:41.105000	204.4432339	-30.6654538	R	6	100006	3.78	2
55670	2010-01-11 12:43:17.405000	204.4433324	-30.6653873	R	6	100006	2.89	2
55748	2010-01-11 13:48:52.705000	204.4433366	-30.6653770	R	6	100006	3.42	2
56298	2010-01-15 13:55:58.154000	204.4432781	-30.6654010	R	6	100006	4.44	2
56723	2010-01-17 13:31:45.155000	204.4433075	-30.6656140	R	6	100006	2.90	2
57098	2010-01-25 12:03:55.055000	204.4433013	-30.6654146	R	6	100006	2.59	2

Image Preview Coverage Details

Lightcurve viewer/analyzer and retrieval

Return to Query Page Catalog List Process Monitor Program Interface Catalog Search Help

Catalog Search Result for PTF Lightcurve Table

Object/Coordinate	Source	Type	Glon	Glat	Equatorial J2000
312.333629 -1.078686	Coordinate		46.34431	-26.51339	20h 49m 20.07s -01d 04m 43.3s

Cone Search Constraints: No 33 sources found.

SDSS r 2x

17.5
17
16.5
16

55 400 55 420 55 440 55 460

MJD (days)

Result IPAC Table Column Key To Periodogram

obsid	mag	automagerr_at	oid	ra	dec	fid	transient_fla	astrometricnobs	ngood	inbest
55443.19982	17.172	0.028	26832000000	312.3336	-1.078652	0	5.484212e-07	59	58	46
55443.15561	17.236	0.028	26832000000	312.3336	-1.078652	0	5.484212e-07	59	58	46
56820.42195	16.498	0.028	26832000000	312.3336	-1.078652	0	5.484212e-07	59	58	46
56820.38975	16.823	0.030	26832000000	312.3336	-1.078652	0	5.484212e-07	59	58	46

Moving Object Search Tool (MOST)

Image Dataset

For PTF: Time Range = 2009-01-16 to 2017-03-02
For complete range, leave limits blank (but this may take a long time)

Observation Begin (UTC) Observation End (UTC)

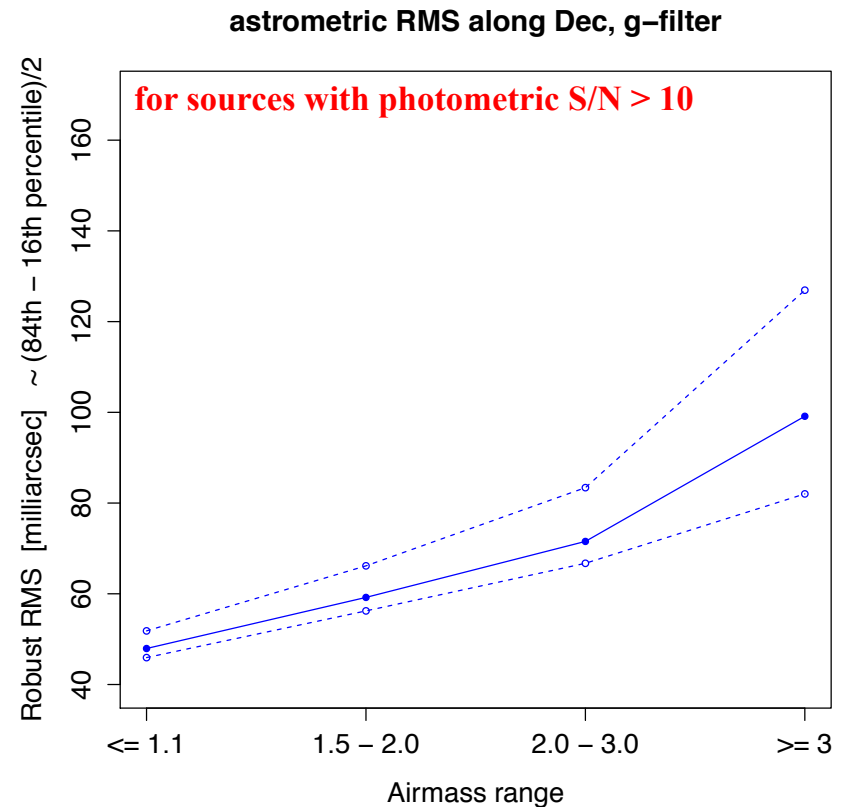
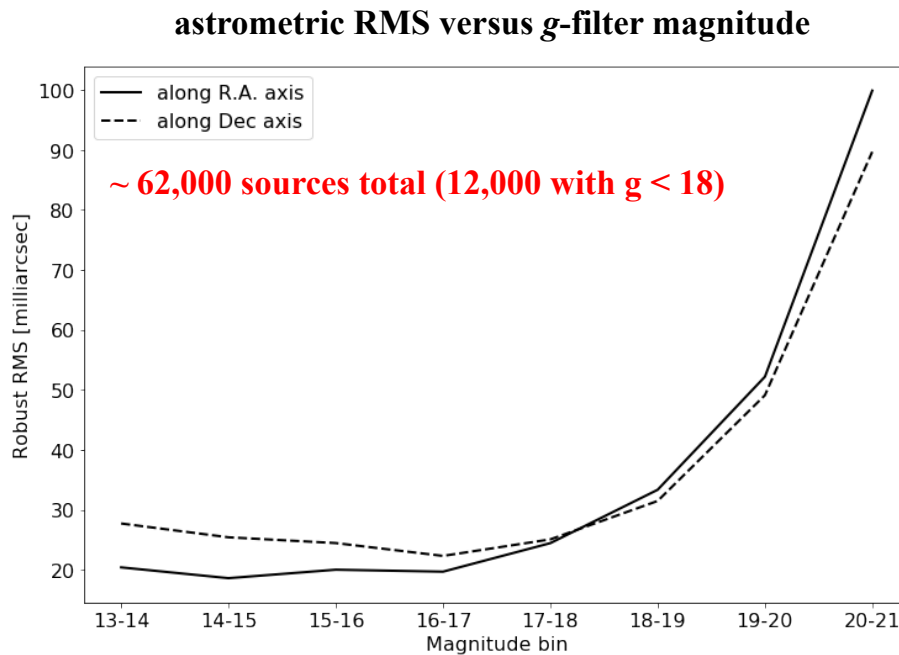
Ephemeris Step Size (day) Output Mode

Create Fits and DS9 Region Files Tarballs ☐ Create Cutout Images Page w/ Target ☐

Solar System Object Name Input:

Astrometric performance relative to Gaia

- Astrometric precision of bright stars with $r, g < 18$ mag at airmass < 1.2 is $< \sim 30$ milliarcsec (RMS per axis).
- Accuracy for sources with $S/N > 10$ ($g, r < 20$ mag) at airmass < 2 is $< \sim 65$ milliarcsec.
- No astrometric biases are present.
- Theoretical astrometric precision: for bright sources, limitation is atmospheric scintillation noise: ~ 22 milliarcsec.

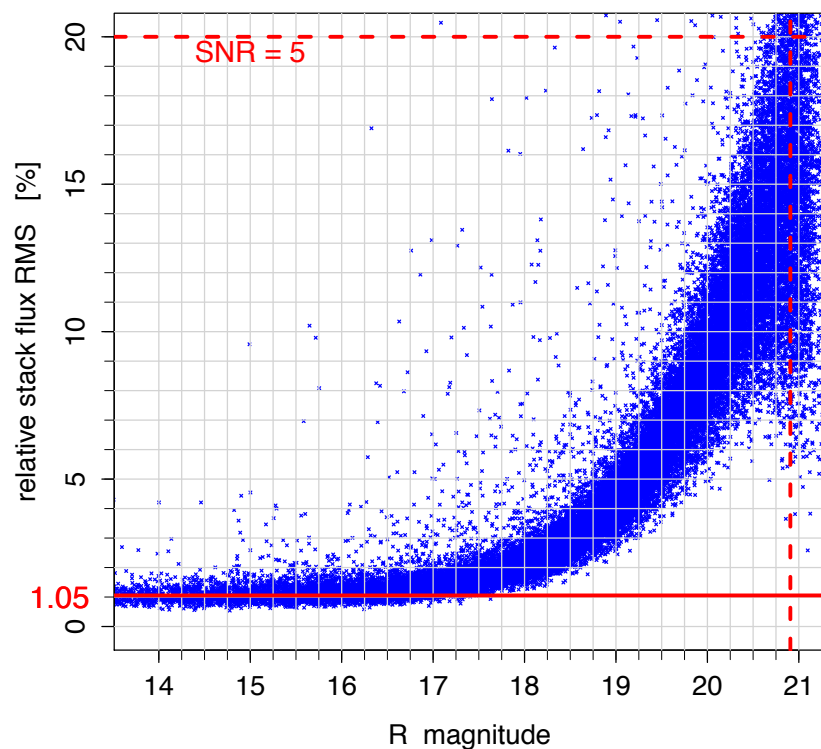


Photometric precision (repeatability)

- From matching epochal PSF-fit source catalogs: typical range is ~ 8 to 20 millimag; depends on airmass.
- 5- σ limiting depths are consistent with expectations and photometric uncertainties in PSF-fit catalogs.
- Plots represent relative flux-RMS from photometric repeatability: $\sigma(f) / \langle f \rangle$.

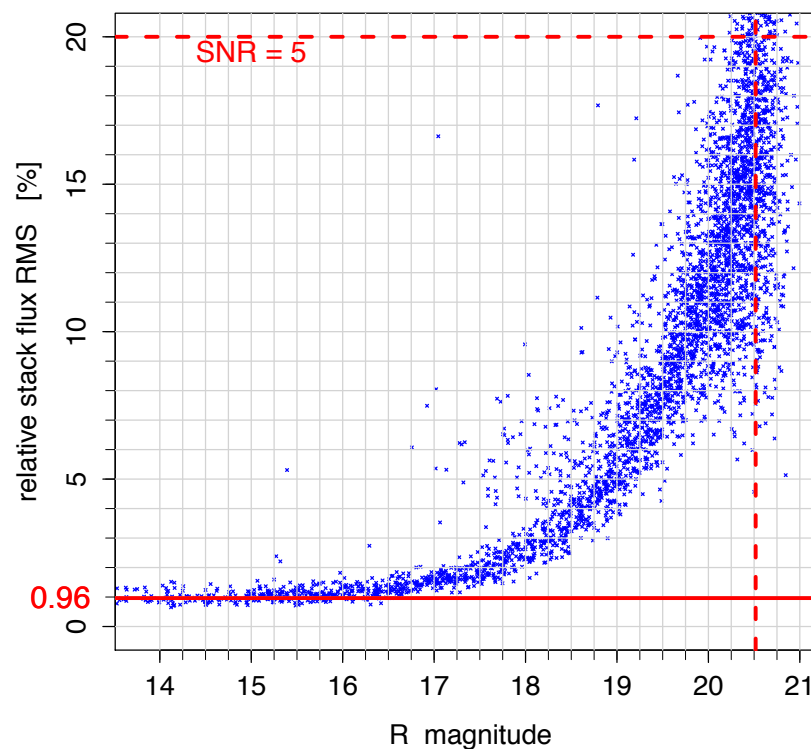
galactic plane field

ztf_000513_zr_c04_q1_mtchstack



high galactic latitude field

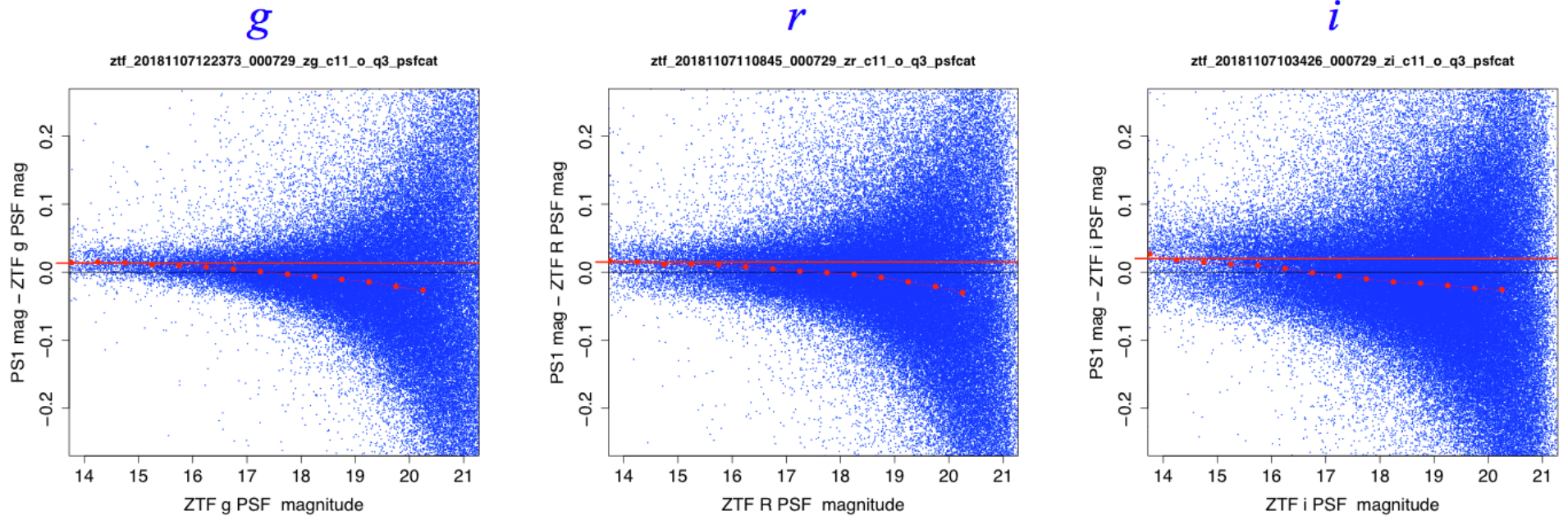
ztf_000520_zr_c12_q4_mtchstack



Photometric calibration check (data from Nov. 7, 2018)

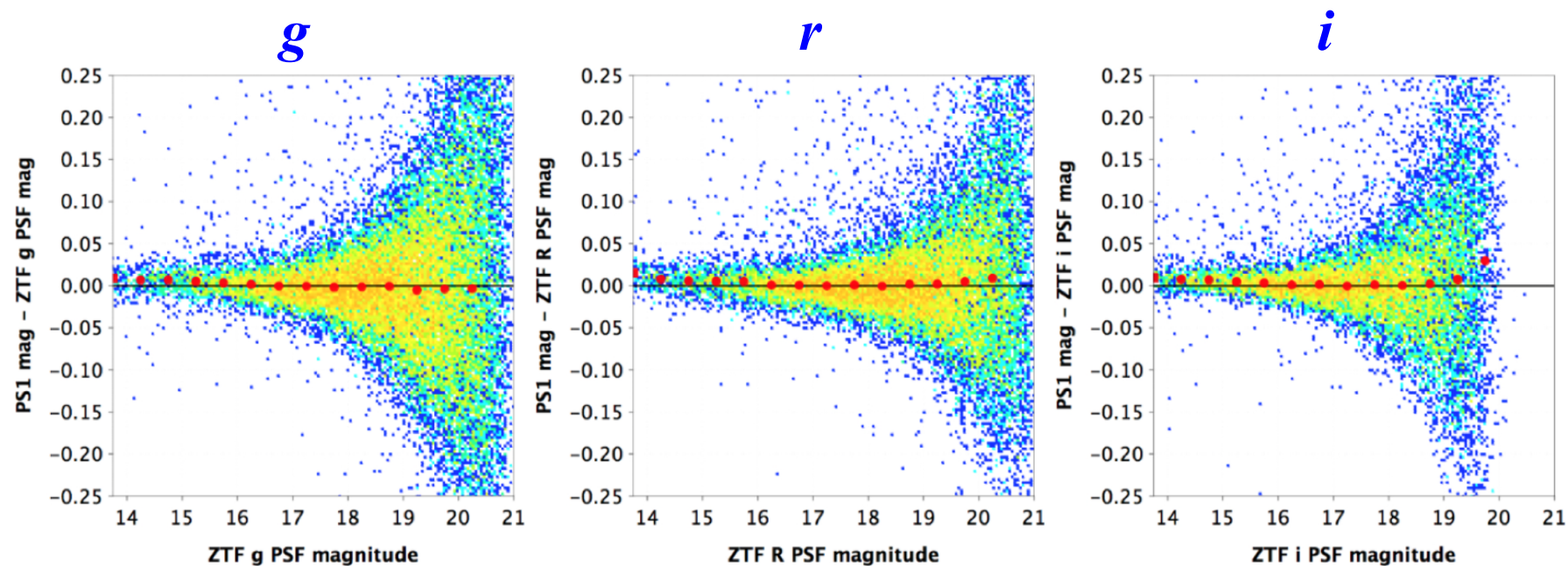
- Residuals are within ~ 0.025 mag in both PSF-fit and aperture-based catalogs *with respect to Pan STARRS1*.
- Following calibration, magnitude dependent biases are present. This is variable across fields.
- Analyses are required to track down the origin of these biases *with respect to Pan STARRS1*.

Below are from quadrant-based PSF-fit catalogs; all in the galactic plane ($|b| < \sim 8^\circ$)



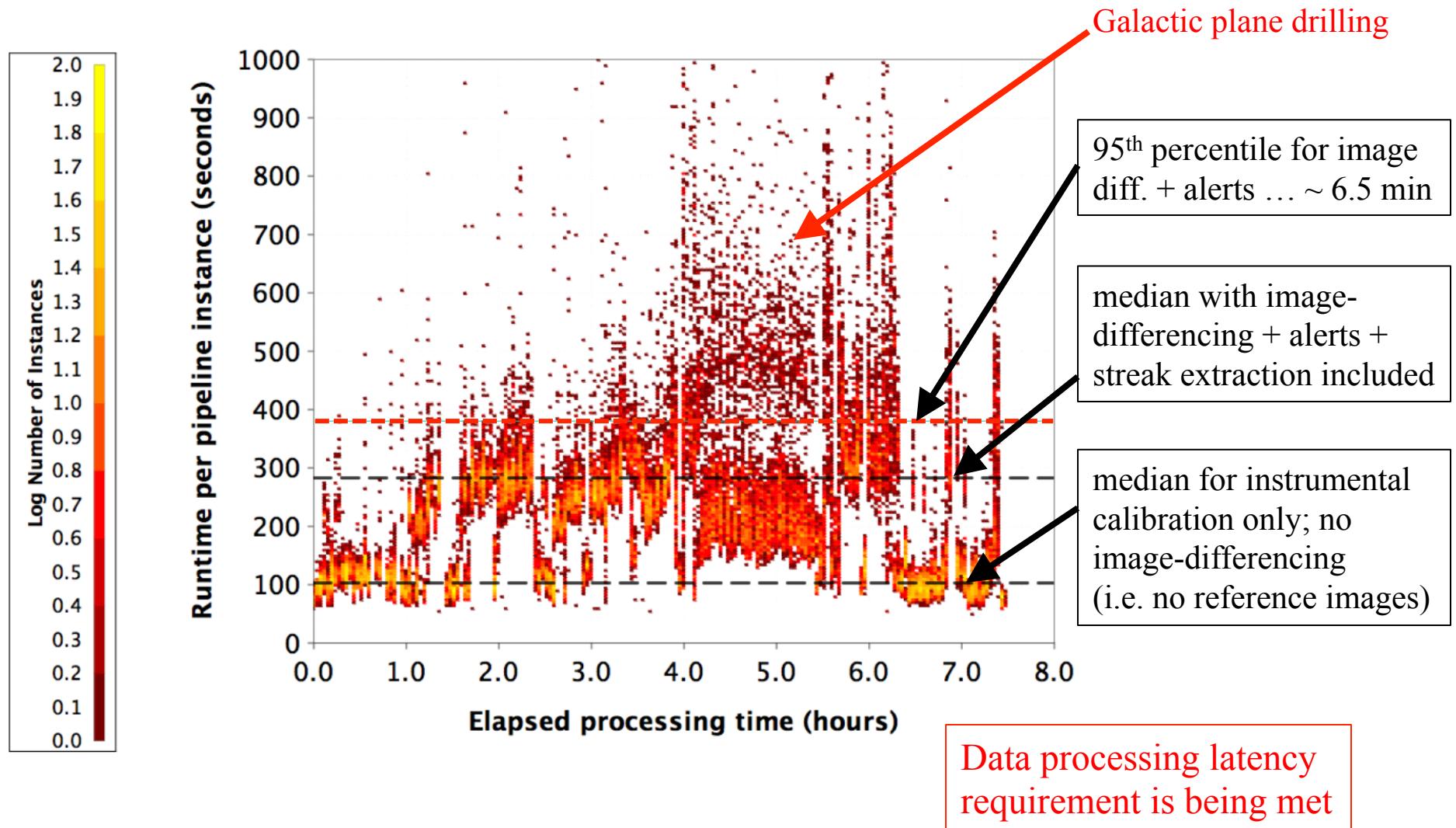
Photometric calibration check (data from June 10, 2018)

Residual biases were smaller for bright sources ($< \sim 0.015$ mag) for same fields *with respect to Pan STARRS1*.

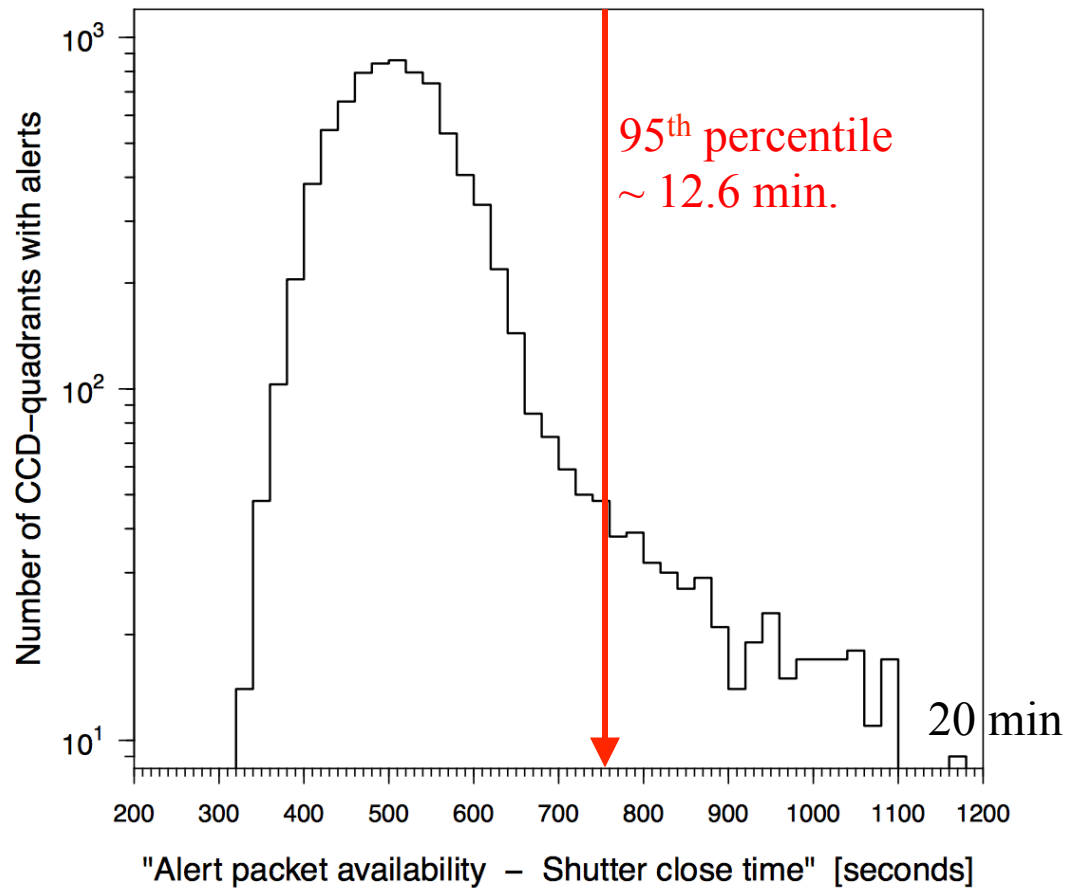


Real-time Pipeline Runtime

processing unit = one readout quadrant



Alert Packet Latency: Shutter Close → IPAC Kafka

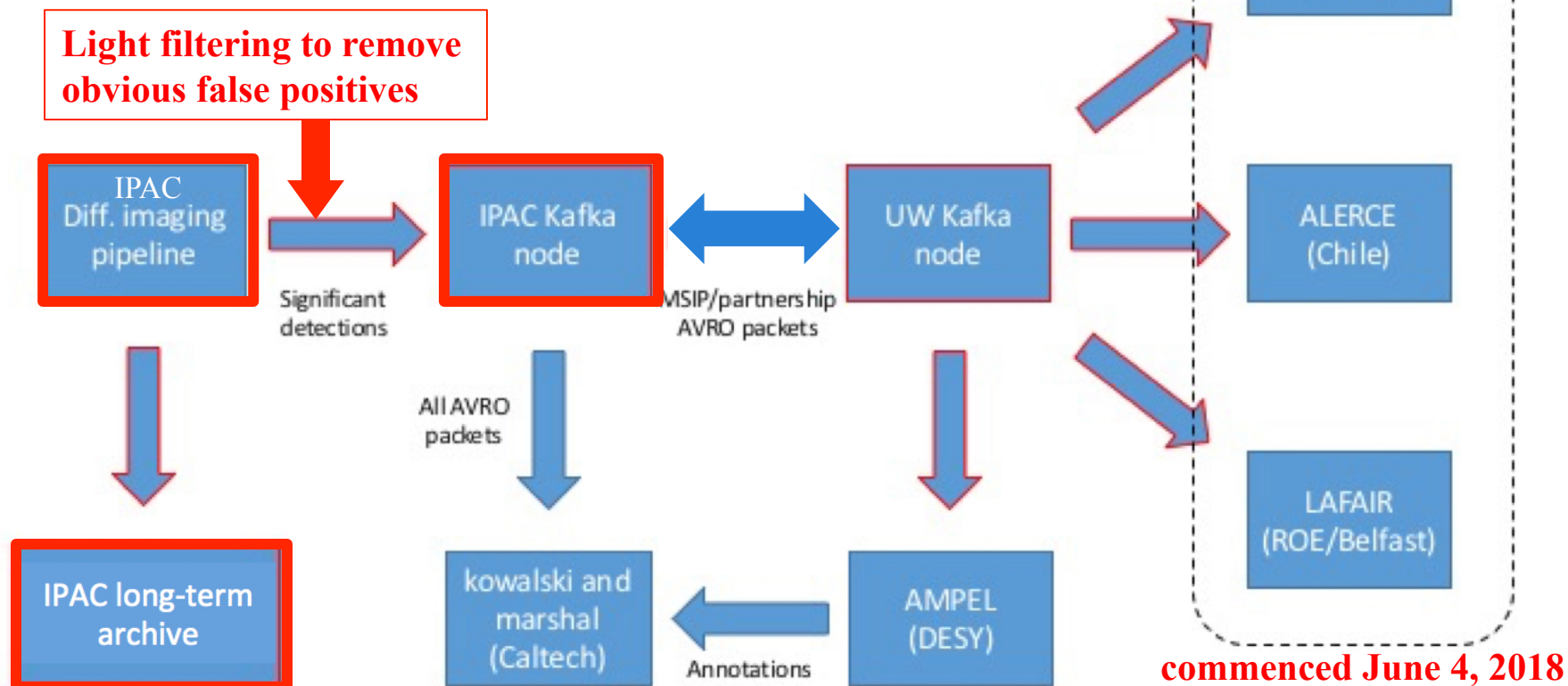


Data delivery timing
requirements are being met

Galactic plane processing
(25% of exposures in this night)

Alert Packet Distribution

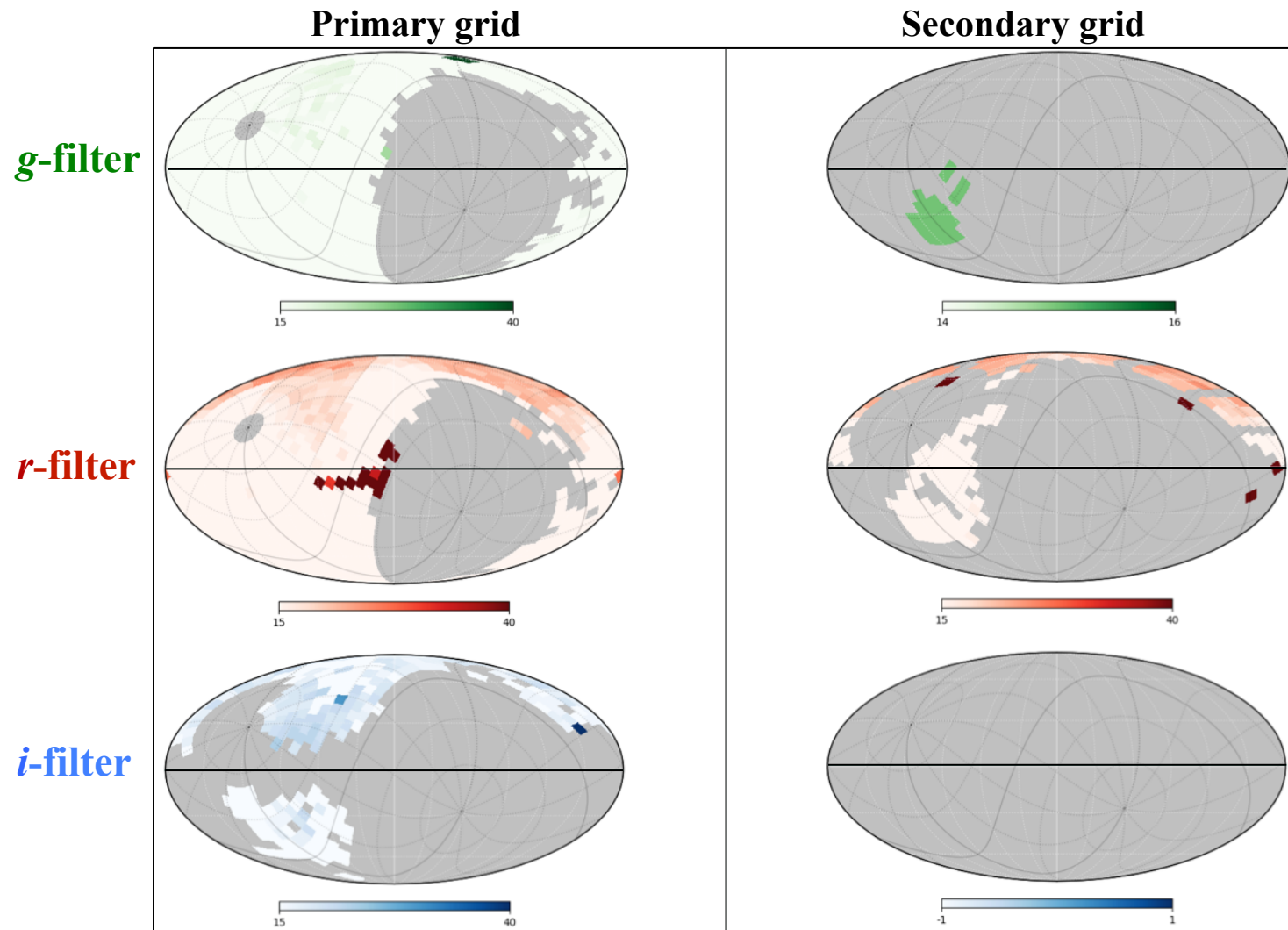
- **Alert packet:** a file containing metadata and contextual information on a single event extracted from a difference image.
- IPAC generates alerts, which are available via its Kafka node and also archived.
- The IPAC Kafka node is mirrored at UW, from where packets are distributed.



Reference Image Coverage (2018-11-14)

galactic projection

Reference image coverage drives image-differencing and hence where on-sky alerts are generated.



Reference Image Coverage Status

Filter	<i>g</i>	<i>r</i>	<i>i</i>
#quadrants with $N \geq 15$ archived science images	41,648	50,860	13,193
#current reference images (2018-11-14)	28,942	31,672	updated 2018-09-19 11,018
#reference images with new <i>provisional</i> cuts	38,277	46,862	11,018

- Initial criteria to select good quality science images for reference image generation were devised during commissioning/SV (January 2018).
- These criteria were found to be too tight.
- **Recommendation:** proceed with *automatic* reference image generation using looser criteria. The DS already has these tuned and tested.
- Pending approval by the ZTF Board.

First Public Data Release

- According to NSF Proposal: release data from public survey at end of survey year one: **March 17, 2019.**
- The Data System staff are on track to meet this date.
- Tools for data access and analysis are in place.
- Preparations:
 - Refine content/scope of release if needed.
 - Reprocessing subset of public data where necessary; requires analysis of holdings.
 - Updates by IRSA team; freezing of delivery contents.
 - Analysis, documentation, cautionary notes, and on-line tutorials.

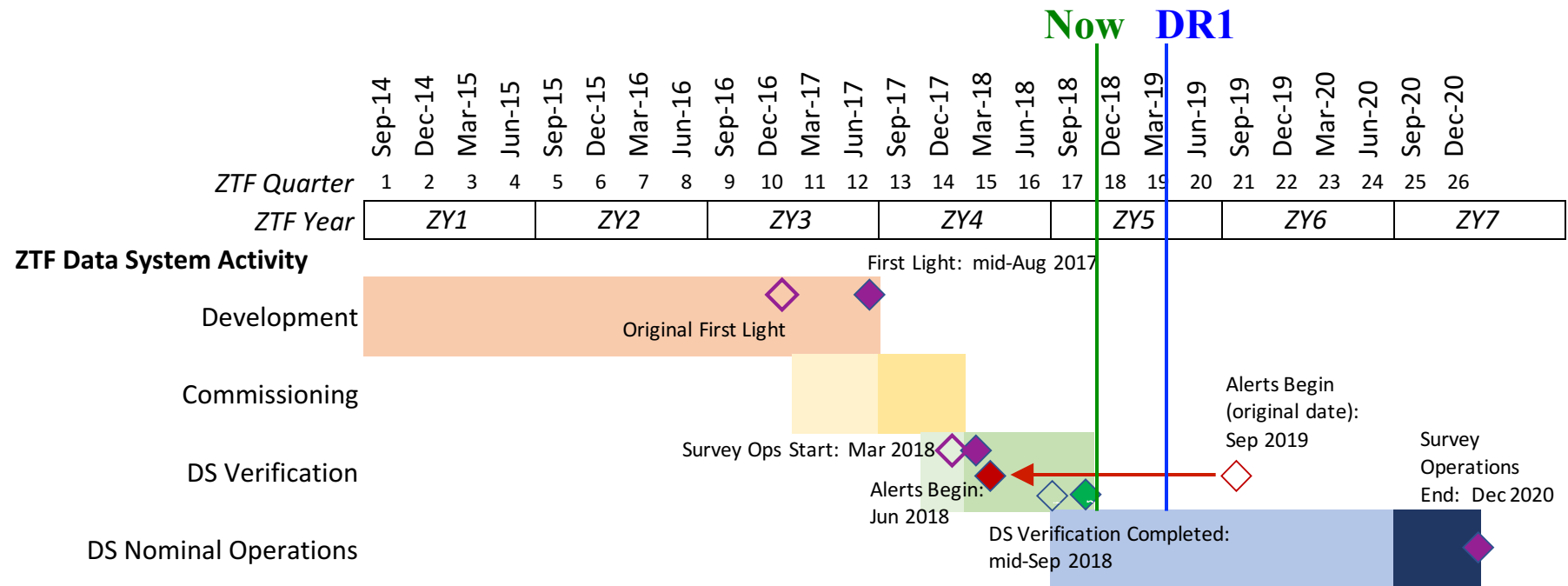
Summary (DS checklist items)

- Production pipelines: *complete and operational*.
- Baseline deliverables and products to enable the NSF-funded science (requirement): *complete*.
- User interfaces and services for data access (requirement): *complete*.
- Astrometric performance: $\text{RMS} < 65$ milliarcsec for $\text{S/N} > 10$ with respect to Gaia.
- Photometric precision (repeatability): varies from ~ 8 to 20 millimag.
- Photometric calibration accuracy: residuals of $< \sim 2.5\%$ with respect to Pan-STARRS1.
- Data processing latency (realtime alert production requirement): *satisfied*.
- Public alert generation (requirement): *in place; started June 4, 2018*.
- Sky coverage for alert production (requirement): *incomplete - pending ZTF board decision*.
- First Public Data Release (requirement):
 - on track for NSF-proposed schedule: to occur at end of survey year one (March 2019).
 - archive services for data retrieval are in place.

-
- Now settled into steady-state automated operations.
 - Refinements to calibrations and machine-learned classifiers continue.
 - Development of other services requested by project beyond baseline requirements are in progress.

Back up slides

ZTF Data System Timeline



- Science operations commenced on March 17, 2018
- Public Alerts officially commenced on June 4, 2018
- Data system verification period ended in mid September 2018
- We are currently in nominal operations; tuning/refinements have continued
- Public Data Release 1: March 19

Development Status of sub-systems

Component Group	DS Component	Status 2017-05	Status 2018-07	Notes
<i>Transfer</i>	Data Transfer Software Protocols from P48 to IPAC	100%	100%	
<i>Pipeline</i>	Ingest, CCD quadrant-splitting, floating bias correction	100%	100%	
	Calibration generation: biases, high & low-v flats	100%	100%	
	Instrumental calibration (astrometric & photometric)	100%	100%	
	Reference image generation (co-addition)	100%	100%	Pending approval to relax input criteria
	Source-matching & photometric corrections for lightcurves	100%	100%	
	Transient event discovery	100%	100%	
	Machine-learned vetting of transient events	0%	100%	ML Module integrated. Parameters being tuned.
	Pipeline executive: job scheduling/task orchestration	30%	100%	Completed start of commissioning
	Throughput testing: algorithm & cluster optimization	5%	100%	Completed end of commissioning
<i>Archive</i>	Image and catalog file product server	30%	100%	Completed start of survey operations
	Lightcurve retrieval service w/metadata	5%	100%	Completed July 2018
<i>Depot</i>	Event metadata	100%	100%	
	Stamp-image cutouts	100%	100%	
	Pipeline QA metrics	100%	100%	
	User access/server setup	0%	100%	Completed start of commissioning.
<i>Alerts</i>	Transient alert distribution infrastructure & interfaces	2%	100%	Alerts generated and available for brokers starting June 4, 2018

Development Tasks: in progress & near future

In progress:

- Forced photometry service; **delivery: end of February 2019.**
- Optimizing DB queries to better handle associations with historical events for alert packets.
 - Significant loads on Operations DB were recently detected.
- Lightcurve (*matchfile*) generation updates: filter “bogus” transients from image edges: **deliv: mid January 2019.**

Near future:

- Update from Gaia DR1 to DR2: for both astrometric calibration and alert association: **mid January 2019:**
 - Will require analysis, regression testing, regenerating static field-based catalogs, reformatting queries & code.
- Prepare for Public Data Release 1 (**March 2019; TBD**):
 - Refine content/scope, reprocessing subset of public data, analysis and documentation.
 - Finalize API tools to query lightcurves from IRSA DB (retrieval via GUI already exists).
- Update pointing / WCS offset file (*wcs.cfg*) to provide better CCD-quadrant priors: **end of February 2019.**
- **Recommendation:** start auto-generating reference images using relaxed input criteria (**see next slide**).
 - **Pending approval by ZTF Board.**
 - Analysis needed: assess quality of subtractions & alerts; work with ML team to retrain classifiers.
 - Reassess database loads and possibly reconfigure alert-generation infrastructure to handle increase in #alerts.

Reference Image Coverage Status

Filter	<i>g</i>	<i>r</i>	<i>i</i>
#quadrants with $N \geq 15$ archived science images	41,648	50,860	13,193
#current reference images (2018-11-14)	28,942	31,672	updated 2018-09-19 11,018
#reference images with new <i>provisional</i> cuts	38,277	46,862	11,018

- **Diagnosis:** initial criteria used to select good quality science images for reference image generation were too tight.
- **Recommendation:** proceed with *automatic* reference image generation using looser criteria (already tuned: see box on right).

```

obsdate ≥ 2018-02-05
1.8" ≤ FWHM ≤ 4.5"
25.3 ≤ ZP/mag ≤ 26.5
-0.05 ≤ color_coeff ≤ 0.22
maglim ≥ 19 mag
global_median ≤ 1200 DN
robust_spatial_rms ≤ 80 DN

```

Example of new provisional cuts (r filter)

Ongoing Development/maintenance Tasks

- Continued refinements to automated classifiers for point-source alerts and streaks.
- Importing and testing of improved S/G classification scores for PS1 to associate with alert streams.
- Check / re-derive bad-pixel masks used in production.
- Updates to reflect findings and unforeseen “issues” to Camera / OS / TCS configurations, scheduling.
- Updates to Explanatory Supplement, in particular Cautionary Notes sections.
- There are other operations / maintenance / monitoring related tasks – **not the subject of this review.**

Refinements contingent on support from partnership

Optimizations to primarily support partner science programs:

- Correct dome flats for edge / scattering / CCD-etching effects prior to stacking.
 - Includes optimal (re)weighting when combining LED-sets of exposures per filter.
- Star-flat assessment and application (DESY group input).
- Exposure-time correction map (flat augmentation, $\sim 0.2\%$ max, at focal-plane edges).
- *i*-filter fringe correction (DESY group input).
- We expect the partnership to provide standalone drop-in software modules if approved by the project.
 - However, DS resources are still needed to integrate and test new software modules; also reassess runtime.
 - If resources available, these will be scheduled according to priorities.
 - Maintaining the baseline operational system and responding to “hiccups” has priority.

Future ideas (contingent on resources)

- Currently, alerts are distributed as *avro* packets; consumers ingest these into their databases to enable positional association and retrieval of metadata.
- We are currently storing all alert packets in tar-files *per CCD-quadrant* in the archive.
 - The files are only searchable on an image-basis using the standard API and GUI, but no search capabilities exist at the alert (source) level.
 - It would be extremely versatile to search *for individual* alerts, their photometric histories, metadata & cutouts, all of which are archived; this information already resides in a DB at IPAC.
- Sandbox (work space) environment for users to perform analyses close to where the data resides.

Accomplishments over last nine months (1)

- Infrastructure and software for distributing Alert Packets: includes Kafka; hardware; UW interfacing
- Alert packet schema and contents stabilized following feedback from science working groups
- R&D on filtering of difference-image events for alert packets to mitigate obvious false-positives
- Improved quality of differential photometry in alert packets with meaningful uncertainties
- Solved depth-issue for alerts generated from deeper (300 sec) exposures to support ToOs
- Refinements to PS1 Star/Galaxy scores for associating with alerts; combined with Gaia parallax/PM
- Long-term archiving of alert packets at IRSA – now subject to same user/*programID* access policies
- Tuned/optimized Moving Object detection pipeline that links events to create tracklets (*ZMODE*)
- Tuned/optimized Fast-Moving Object (streak) detection pipeline (*ZSTREAK*)
- More accurate reporting of known asteroids to associate with alert streams and *ZMODE* output
- Accurate reporting of known comets to associate with alert stream and *ZMODE* output
- Ghost prediction and masking (both co-moving and counter-moving types)
- All ancillary file products are now downloadable through the archive GUI

Accomplishments over last nine months (2)

- Lightcurve (*matchfile*) products from linking epochal image extractions now routinely made
- Lightcurve (*matchfile*) products containing only partnership data now also made
- Lightcurve query GUI and Time Series / visualization tool now ready for partnership
- Automated reference-image generation (more later)
- Better real-time reporting of QA metrics, pipeline status and failures for Observing System team
- Enabling of image-cutouts on archived (compressed) difference images – IRSA service
- Significantly improved subtractions in “challenging” galactic plane fields
- Synopsis of reference image holdings and diagnosed incompleteness: retuned input criteria/filters
 - Relaxed input criteria for *i*-filter reference images on 2018-09-19
- Automated daily generation of all-sky coverage maps for reference images
- Data System paper accepted by PASP, in press.

Real-time pipeline monitoring dashboard (ready to go)

