ZTF Science Data System Status & Plans

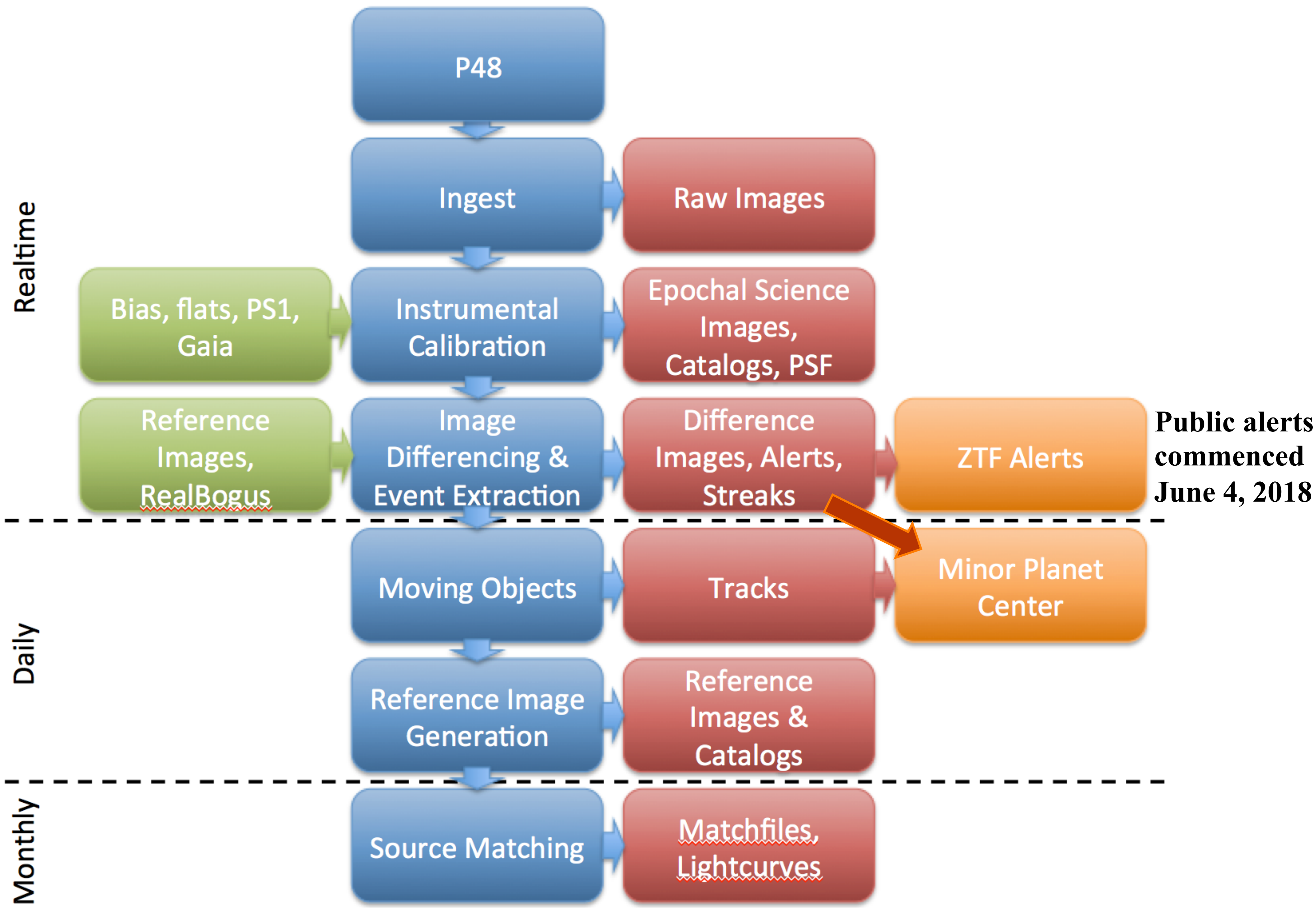
Frank Masci & the IPAC-Caltech ZTF Team

March 12, 2019

Outline

- Refresher: overview of processing and products
- Current data holdings and statistics
- Recap of some on-sky performance
- Recent significant updates
- Data quality & advisories
- First Public Data Release
- Validating & correcting photometric uncertainties
- Ongoing tasks

Pipeline summary: timeline view



Baseline deliverables / data access portals

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
   Marshal(s); Public Brokers; Archived in IRSA

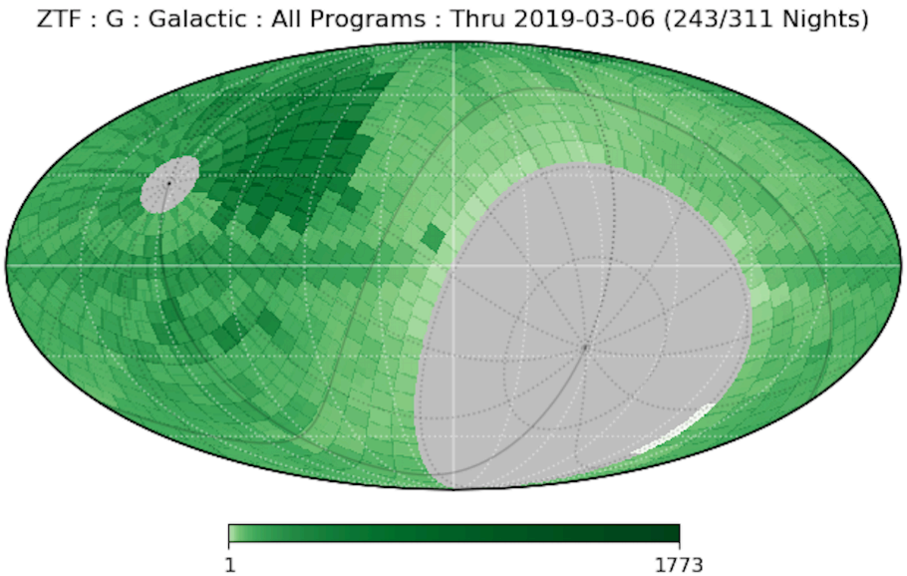
5. **Products to support real-time Solar System / NEO discovery and characterization:** both streaks and tracks
   ZTF-Depot (internal) and IAU-Minor Planet Center

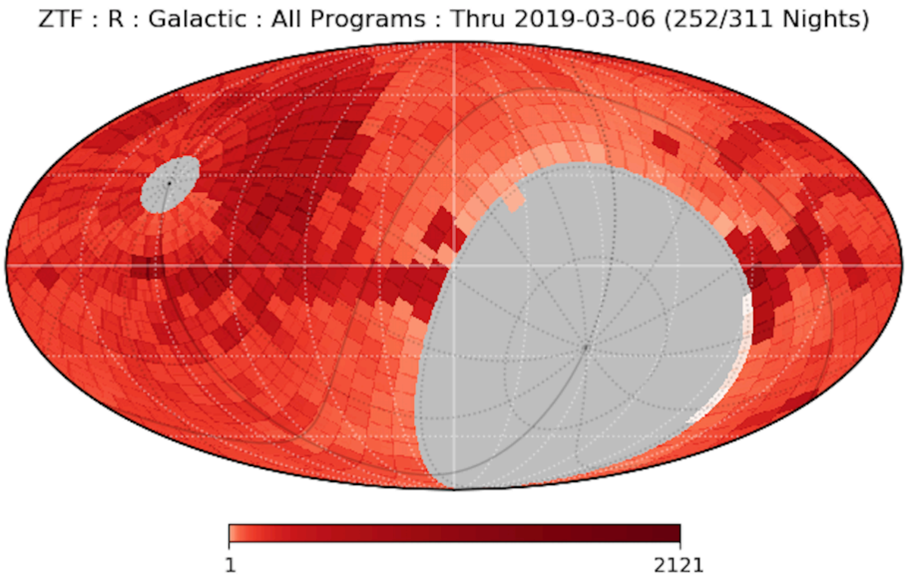
6. **Lightcurves & metrics from matching sources across individual epochs using (1) to beginning of survey**
   Archive (IRSA); ZTF-Depot (raw matchfiles)

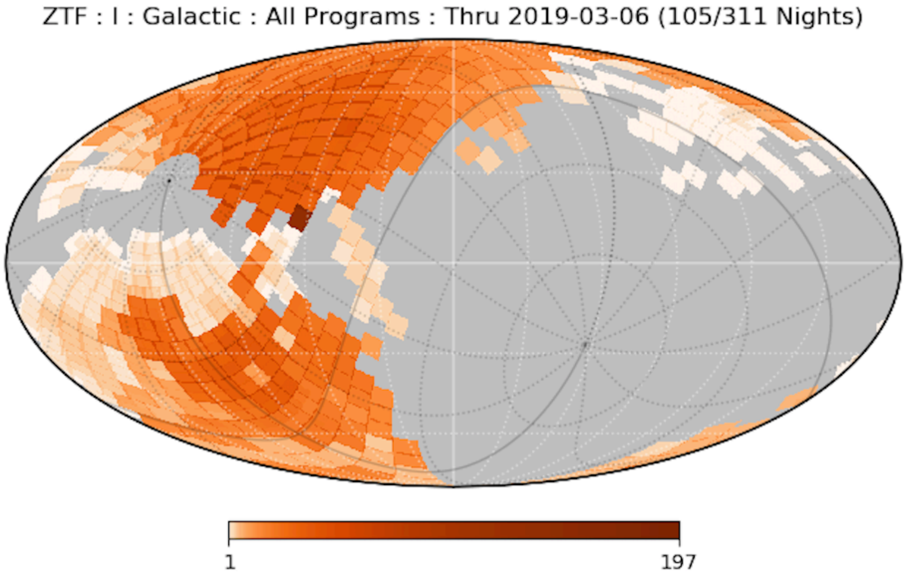
7. **Quality assurance metrics, summary statistics, and survey coverage maps:** for performance monitoring
   ZTF-Depot (internal)

8. **Documentation:** cautionary notes, recipes, and tutorials on data-retrieval and analysis
   Explanatory Supplement on ZTF Public Website; PASP paper published in Dec 2018

Sky coverage: all programs
Mar 17, 2018 (survey start) – Mar 6, 2019







Data volumes & Statistics
Mar 17, 2018 (survey start) – Mar 6, 2019

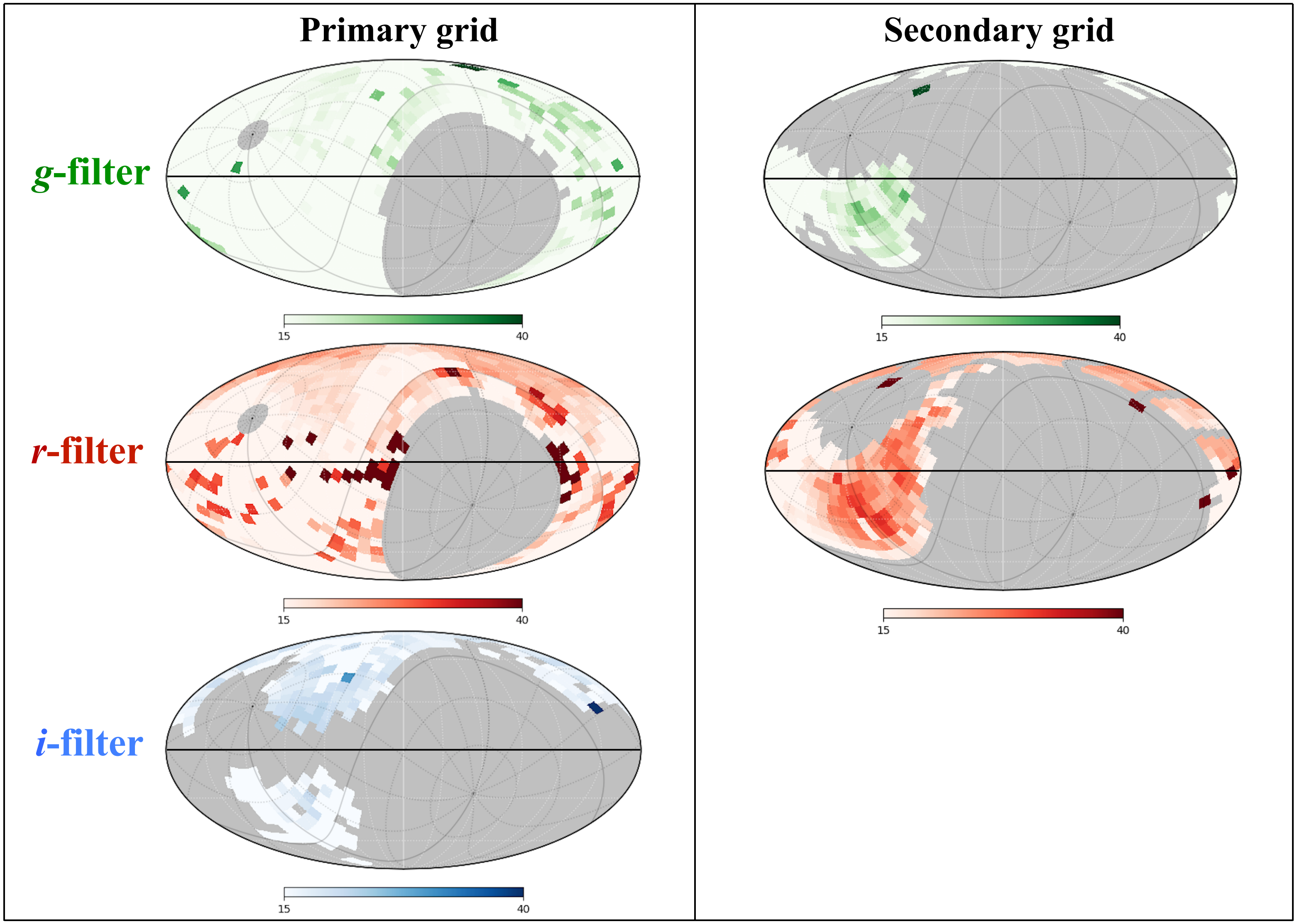
Exposure/Image Metric	*g*	*r*	*i*
Raw on-sky exposures	67,781	103,366	5,510
Survey-ready quadrant-based reference images (#quadrants N≥15 visits)	45,087 (47,934)	53,392 (56,259)	11,127 (13,193)
Lightcurve matchfiles (last made Dec. 15, 2018)	40,822	51,076	11,018
Epochal science image products archived (all CCD quadrants)	~ 10.6 Million (788.6 TB)		

Source Extraction Metric	Number
Epochal science image PSF-fit extractions	183 B
Epochal science image aperture-based extractions	113 B
Reference image PSF-fit extractions ("seeds" for lightcurves)	4.5 B
Reference image aperture-based extractions	1.5 B

Event Extraction Metric	Number
Raw candidate events from all difference images (+ and – diffs)	+ 274 M – 136 M
Alert packets generated from all difference images (+ and – diffs)	+ 58 M – 29 M
Alert packets associated with known solar system objects (≤ 3 arcsec)	2.6 M
Streaked detections from new SSOs	30
Streaked detections from known SSOs	> 12 K
Moving object tracklets not associated with known SSOs & delivered to MPC	~ 5 K
Moving object tracklets associated with known SSOs & delivered to the MPC	> 850 K

Reference Image Coverage (Mar 6, 2019)
galactic projection

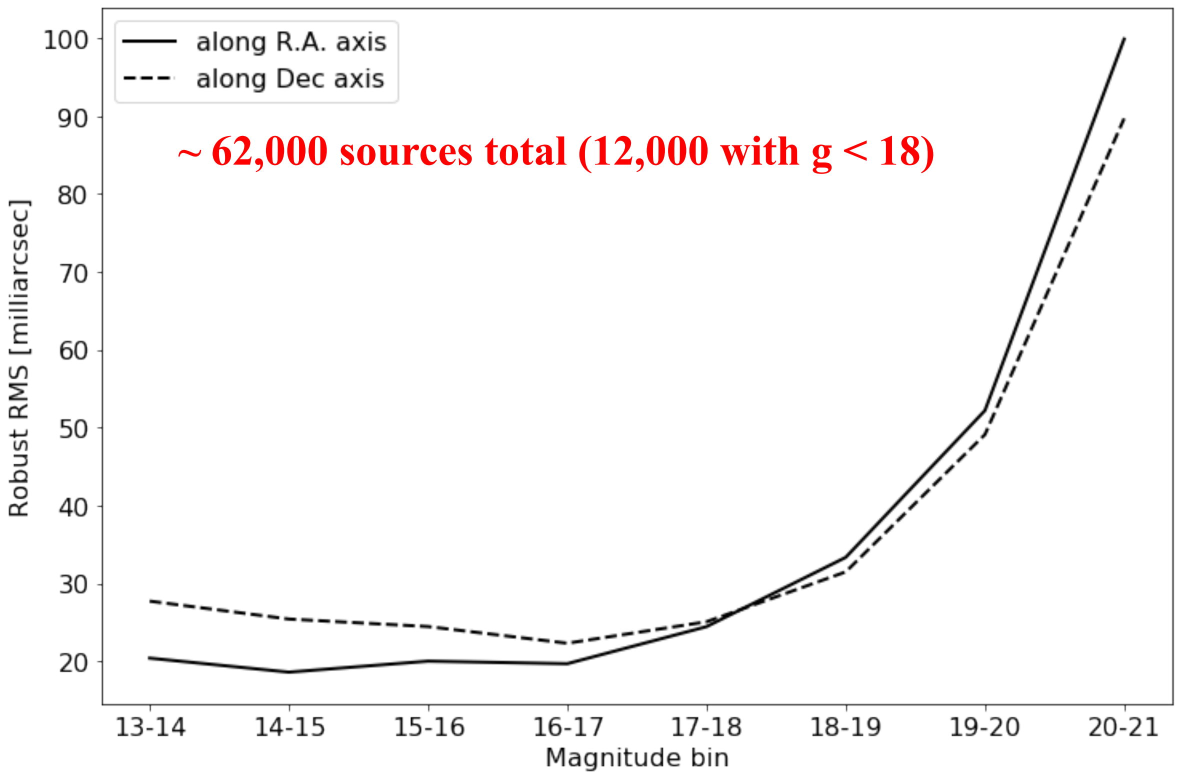
Determines where on the sky alerts are generated.



Astrometric performance relative to Gaia

- Astrometric precision of bright stars with *r, g* < 18 mag at airmass < 1.2 is <~ 30 milliarcsec (RMS per axis).
- Accuracy for sources with S/N > 10 (*g, r* < 20 mag) at airmass < 2 is <~ 65 millarcsec.

Astrometric RMS versus *g*-filter magnitude

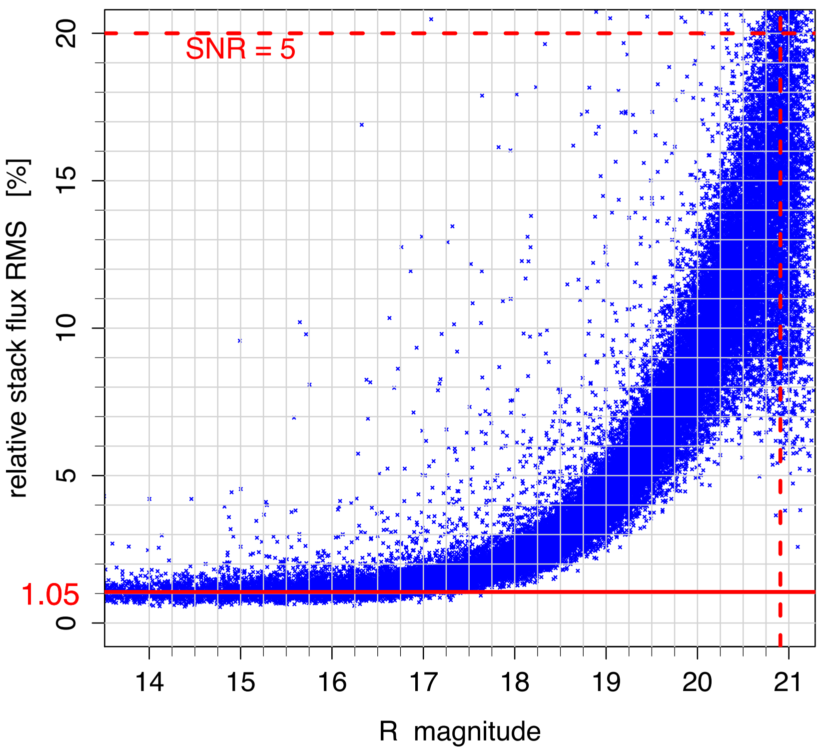


Photometric precision (repeatability)

- 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.
- Plots represent relative flux-RMS from photometric repeatability: σ(*f*) / <*f*>.

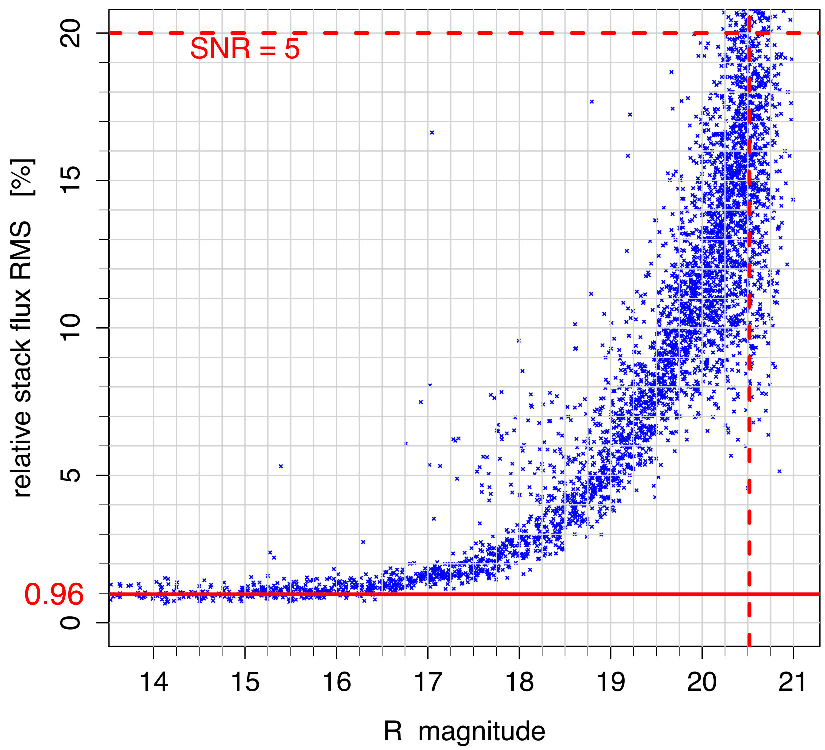
galactic plane field

ztf_000513_zr_c04_q1_mtchstack



high galactic latitude field

ztf_000520_zr_c12_q4_mtchstack

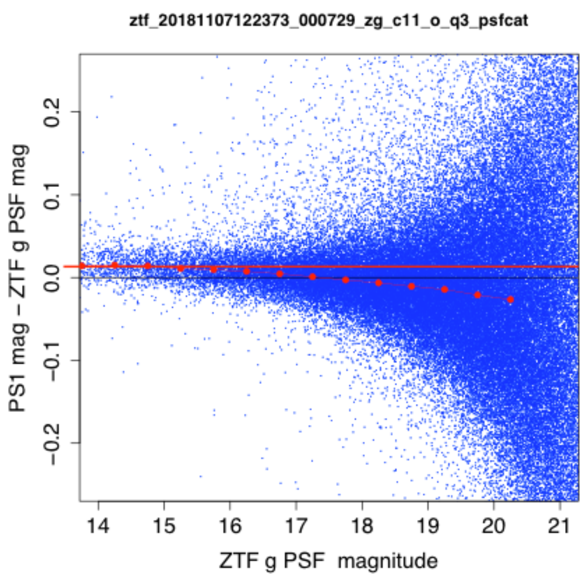


Photometric calibration check
(data from Nov. 7, 2018)

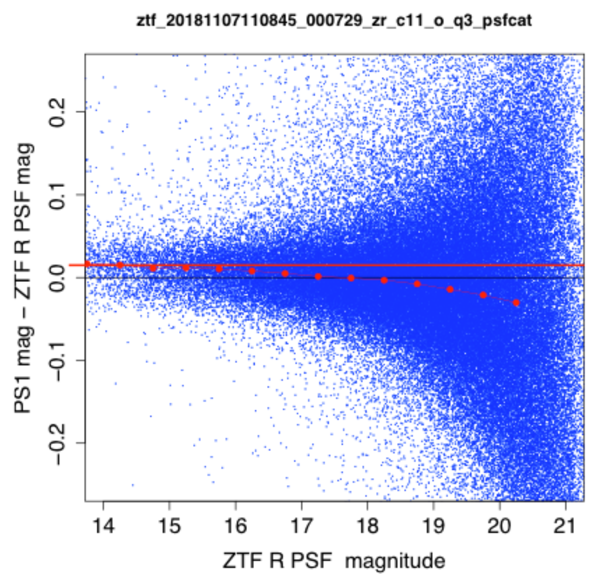
- Residuals are within ~ 0.025 mag in both PSF-fit and aperture-based catalogs *with respect to Pan STARRS1*.
- Following calibration, magnitude dependent biases are present. This is variable across fields.
- Analyses are required to track down the origin of these biases *with respect to Pan STARRS1*.

Below are from quadrant-based PSF-fit catalogs; all in the galactic plane (|b| <~ 8°)

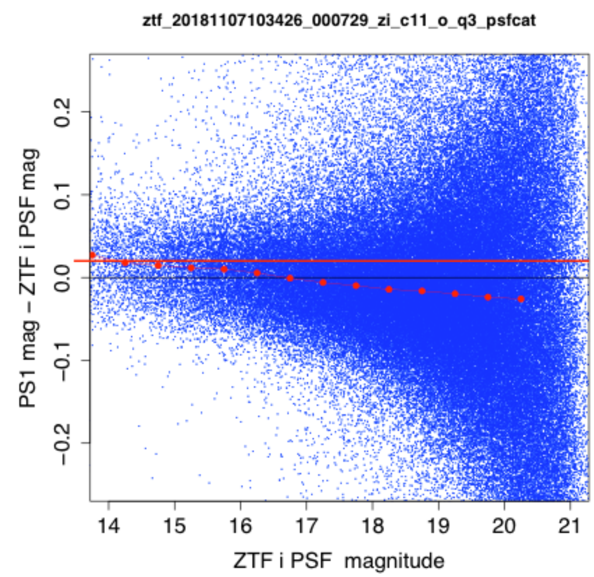
g



r



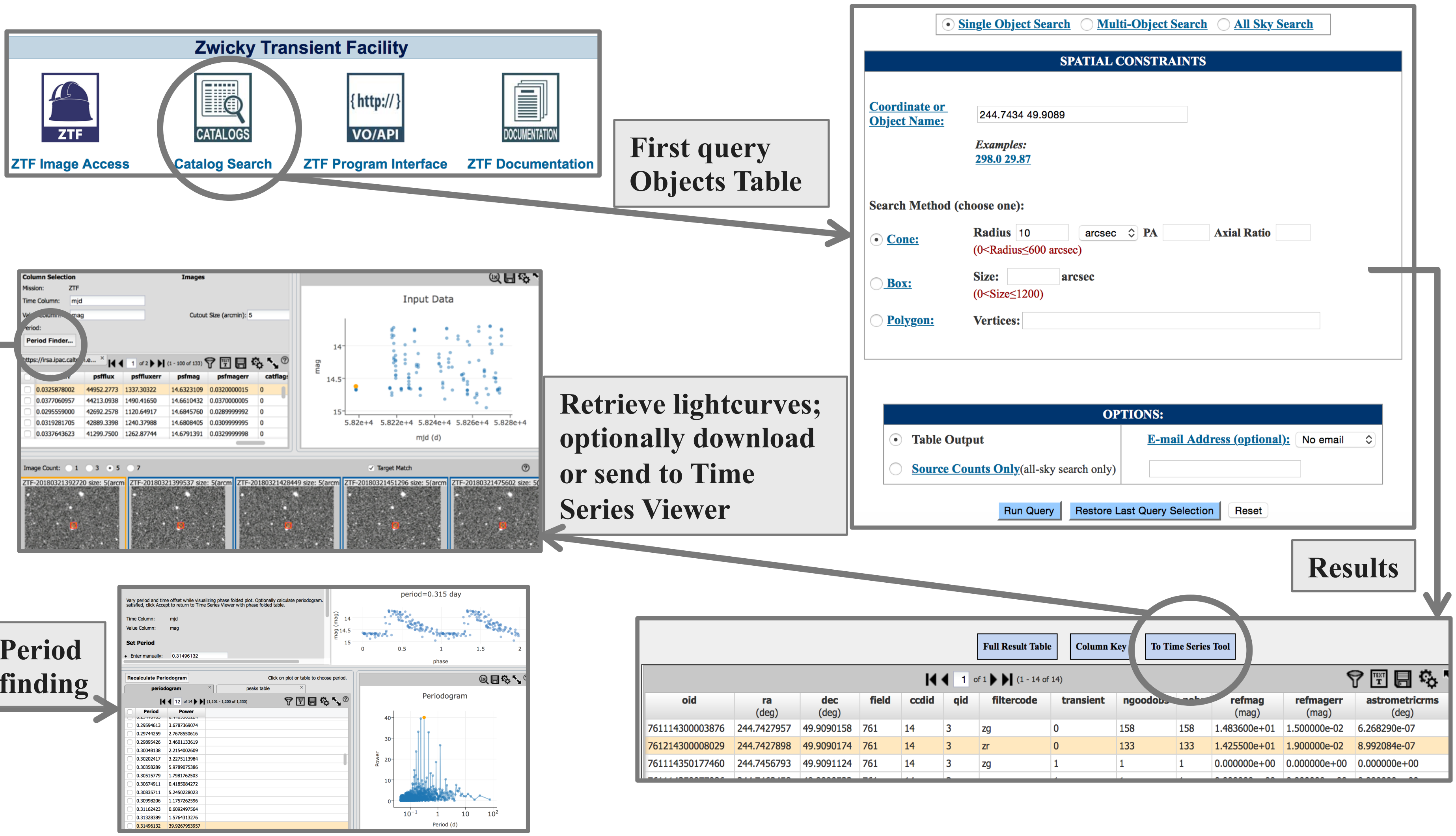
i



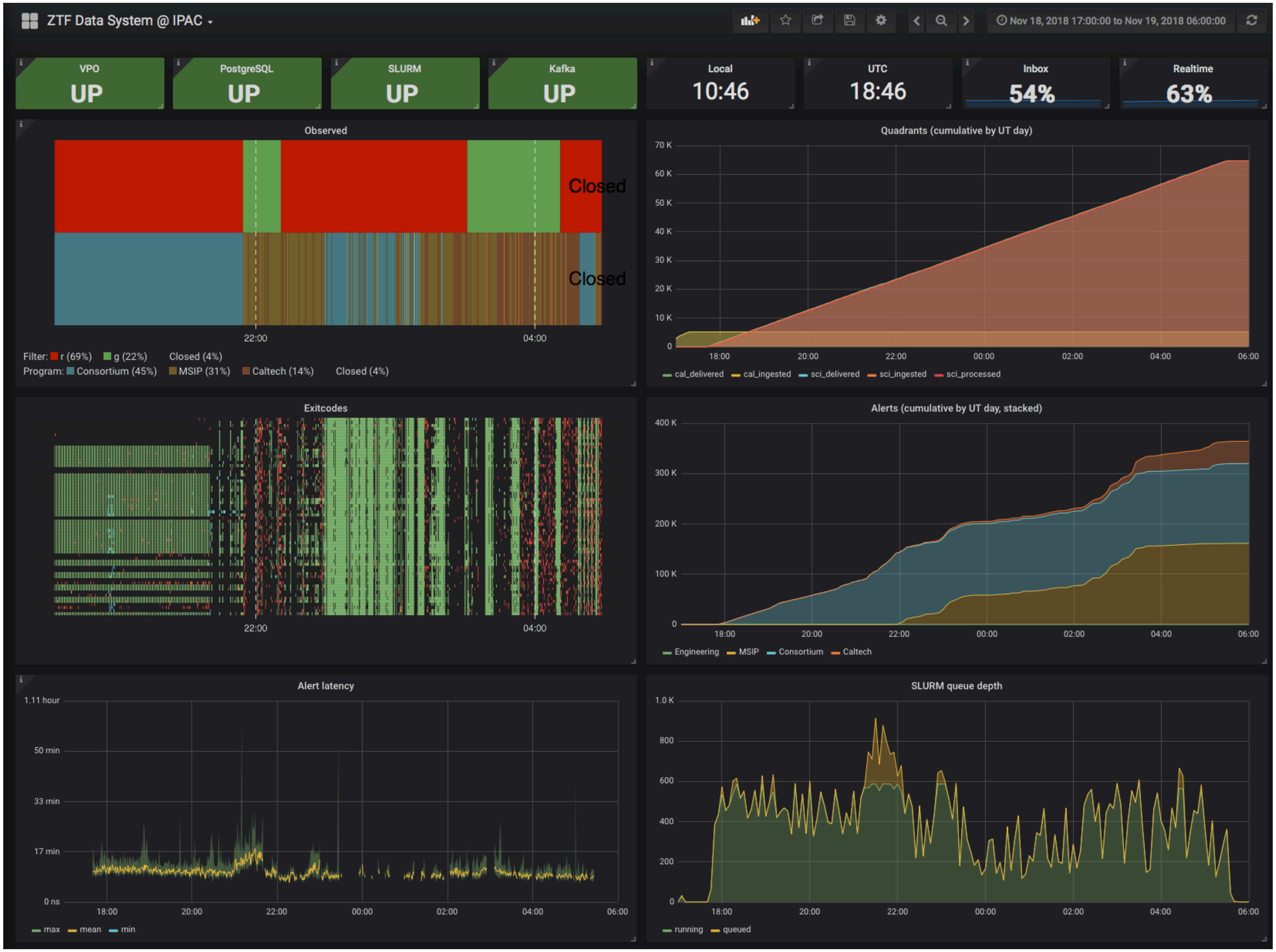
Some significant updates
(~ last six months)

- Improved galactic plane image-differencing and photometry; some challenging fields encountered!
- Improved quality of differential photometry in alert packets with more plausible uncertainties.
  - Lots of feedback received from partnership. Thank you!
- Added more metrics to alert packets:
  - Nearest Gaia source metrics; reference image info to enable DC photometry; color terms and calibration info.
- Refined criteria for selecting science images for reference image generation: significant increase in sky-coverage.
- Cutouts on archived (compressed) difference images now possible: both through GUI and API.
- End-to-end optimization of processing throughput (primarily database tuning).
- Tuning of moving-object pipeline (tracklet-generation via ZMODE): increase in detection efficiency.
- Moving Object Search Tool (MOST) improvements: faster integration of orbital-elements with improved accuracy.
- Data quality analysis:
  - Updates to documentation, cautionary notes, and advisories on data usage.
- Dashboard for realtime visualization of pipeline processing, in particular alert packet production.
- Lightcurve retrieval and time series analysis tools (from multi-epoch source-matching).

Lightcurve retrieval GUI / Time Series Tool

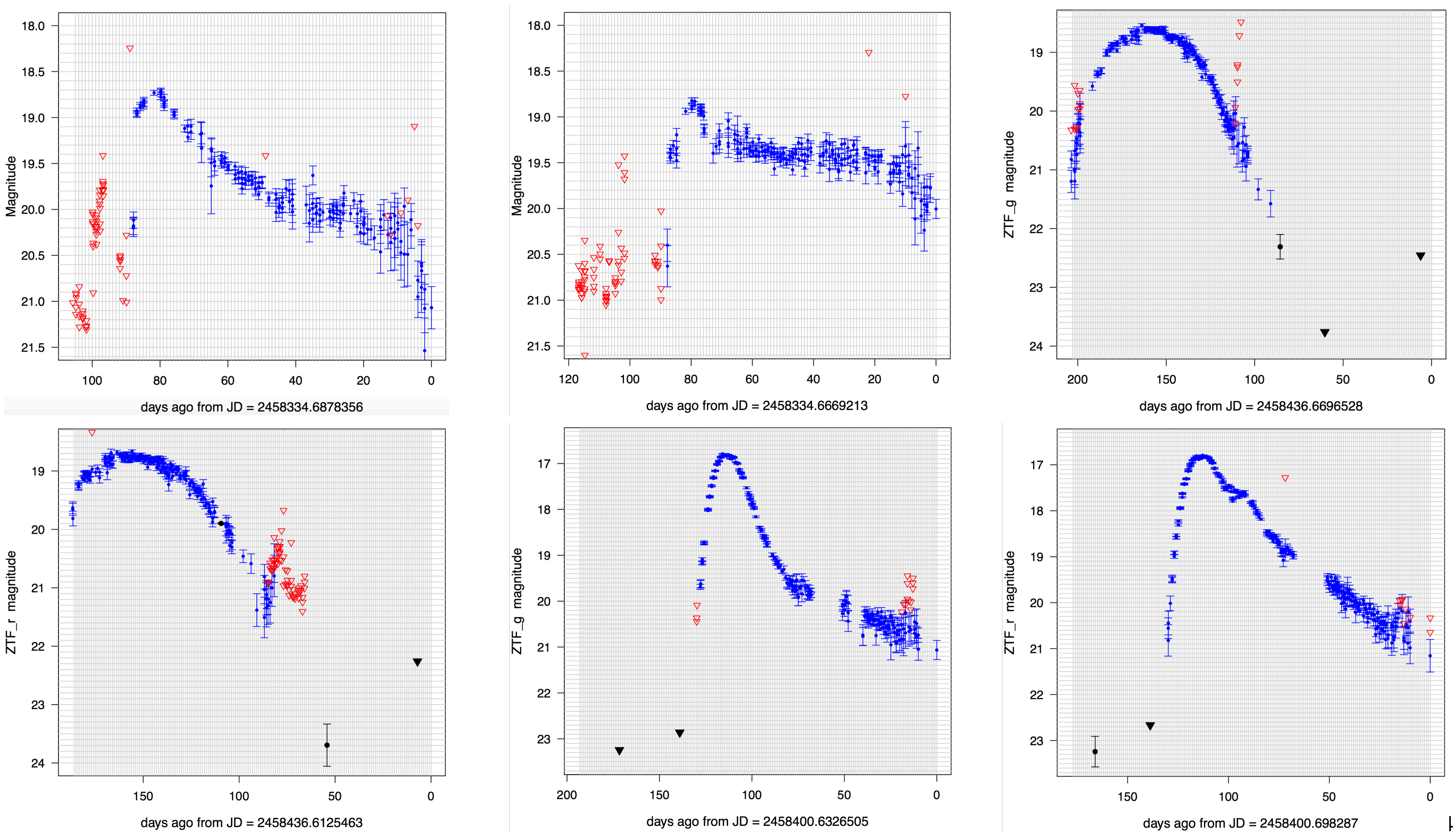


Real-time pipeline monitoring dashboard



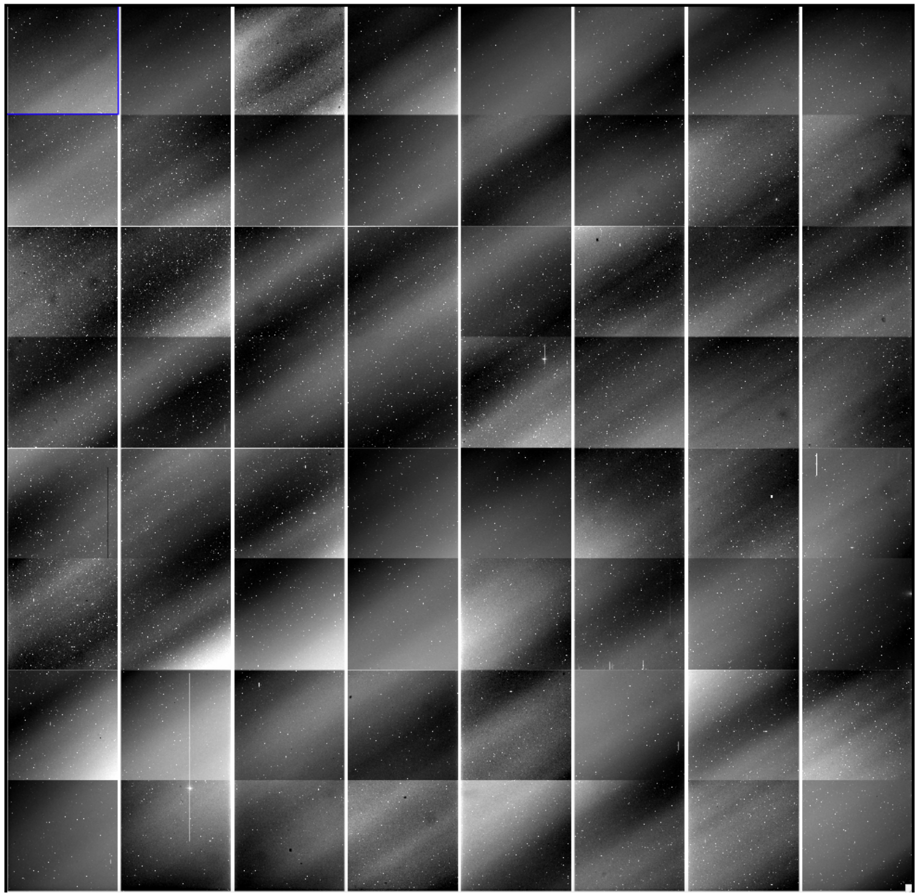
Forced Photometry Service

Currently being tested by Caltech members: refinements in progress, primarily crowded fields.



Data Quality Awareness

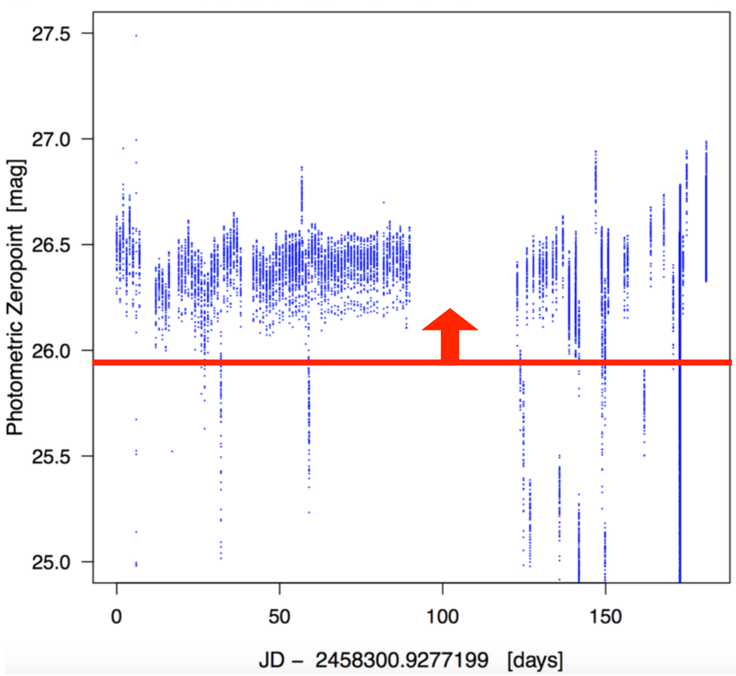
- Recipes and metrics for identifying bad quality data, as well as advisories are being documented in the Cautionary Notes section of the Explanatory Supplement.
- An example is clouds and scattered moonlight!
- Hard to predict in advance; raw data are still calibrated, but solutions are nonsense and products not usable!
  - ➢ This includes difference images and alerts generated therefrom.

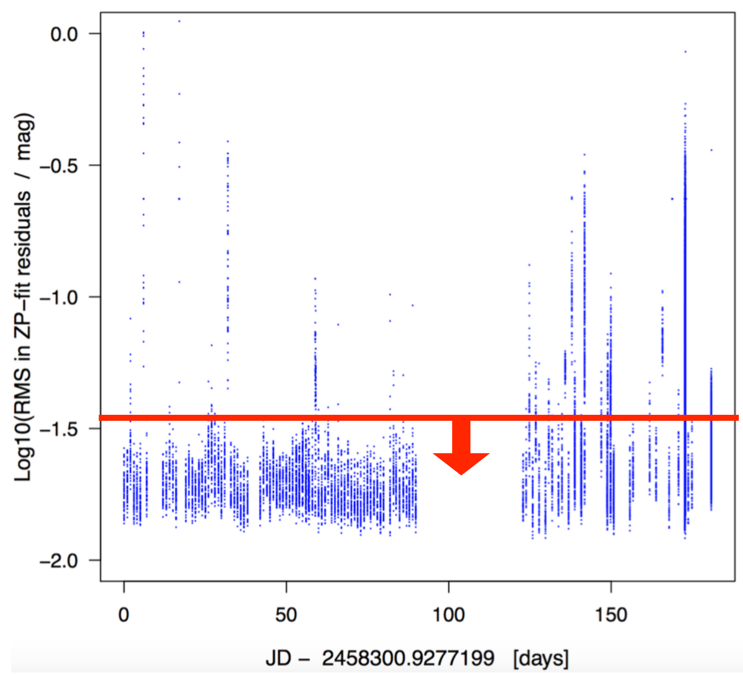


Strategies for flagging bad data

- In progress: identification of metrics for trending and thresholding so data can be flagged in archive.
- Want to catch the worst possible data. Entire nights can be bad.
- Automatic flagging to retain good quality data is a delicate (and subjective) process.

Example of trending **MAGZP** and **MAGZPRMS** metrics for all quadrants in Field 806:





- Please let us know of other useful DQA metrics because we have a Public Data Release coming up ...

First Public Data Release

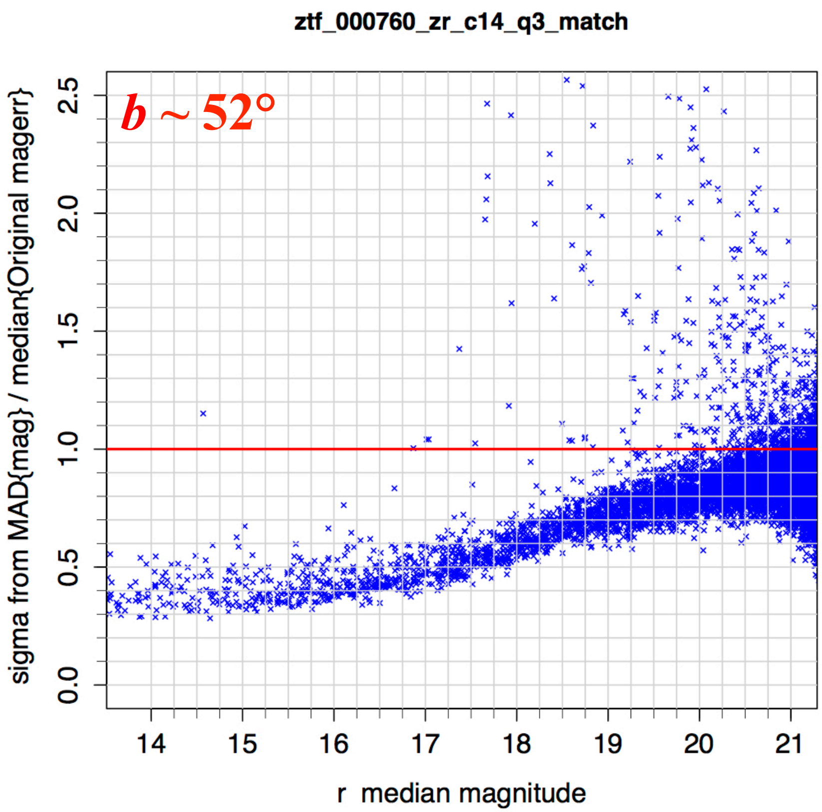
- Anticipated release date: **May 1, 2019.**
- Data span: **March 17, 2018 – December 31, 2018**; only epochs tagged with *programID* = 1 (MSIP).

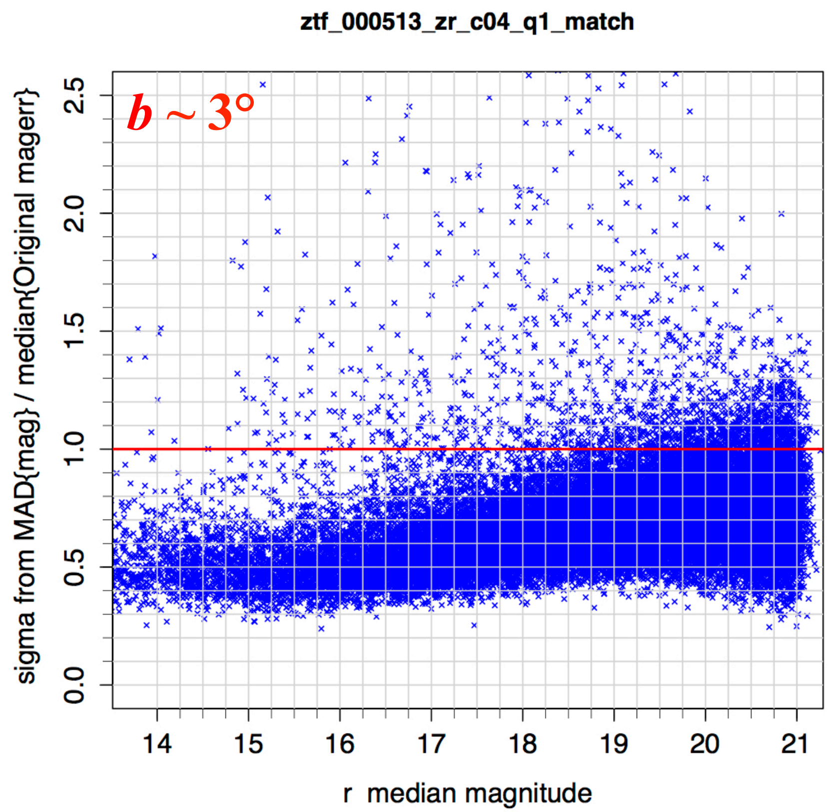
- **Release products:**
  - raw CCD image files
  - epochal instrumentally calibrated science images and all associated ancillary products
  - epochal source catalog table files
  - reference images and associated ancillary products
  - reference image catalog table files
  - calibration image files
  - object source database with collapsed-lightcurve metrics to facilitate lightcurve retrieval
  - lightcurves from matched epochal PSF-fit photometry

- **Tasks in progress:**
  - Lightcurve API to accompany GUI that interfaces with matchfiles: cone search with user-specifiable timespan, magnitude range, #observations, and other constraints.
  - Data quality analysis & filtering / tagging of bad data.
  - Photometric uncertainty validation and correction [next slide].
  - Documentation, cautionary notes, recipes, and on-line tutorials.

Photometric uncertainties in *matchfiles (lightcurve files)*

- Currently, photometric uncertainties reported in *matchfiles* are overestimated by factors of ~ 2 at bright magnitudes relative to photometric scatter across epochs.
- More an issue with relative photometric refinement. Unrefined measurements are better behaved.



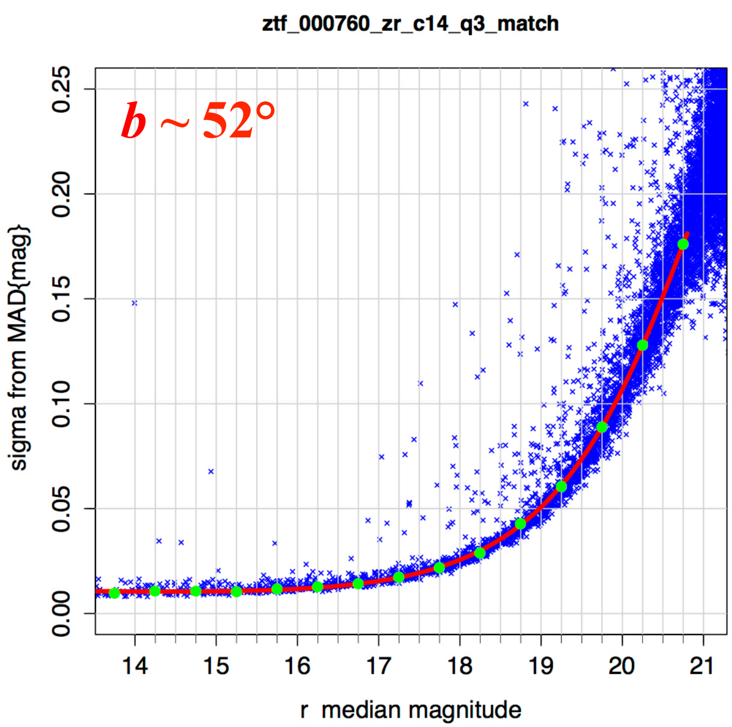


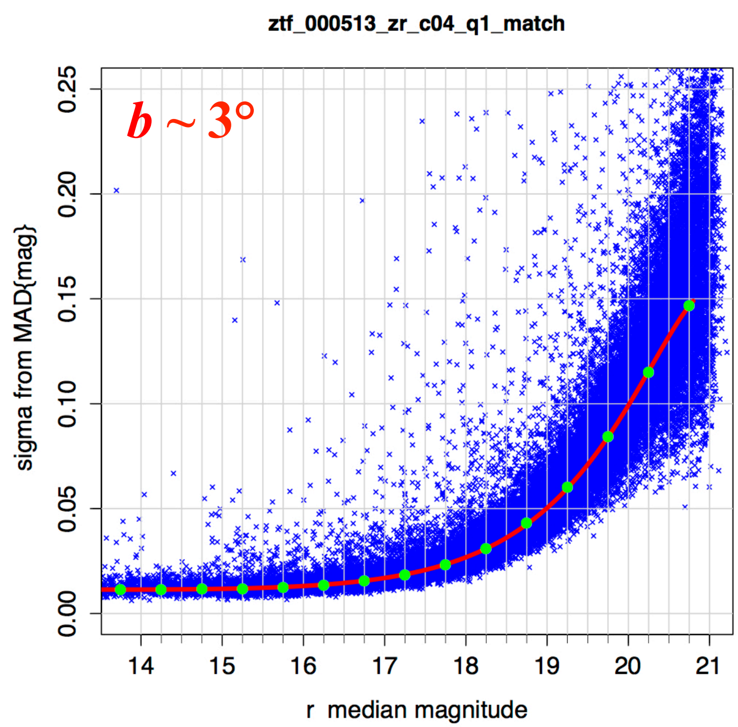
Photometric uncertainties: correction plan in progress

- For each CCD-quadrant-based matchfile, median-bin the robust sigma-MAD lightcurve metric vs. magnitude.
- Fit a function to these binned medians of the form:

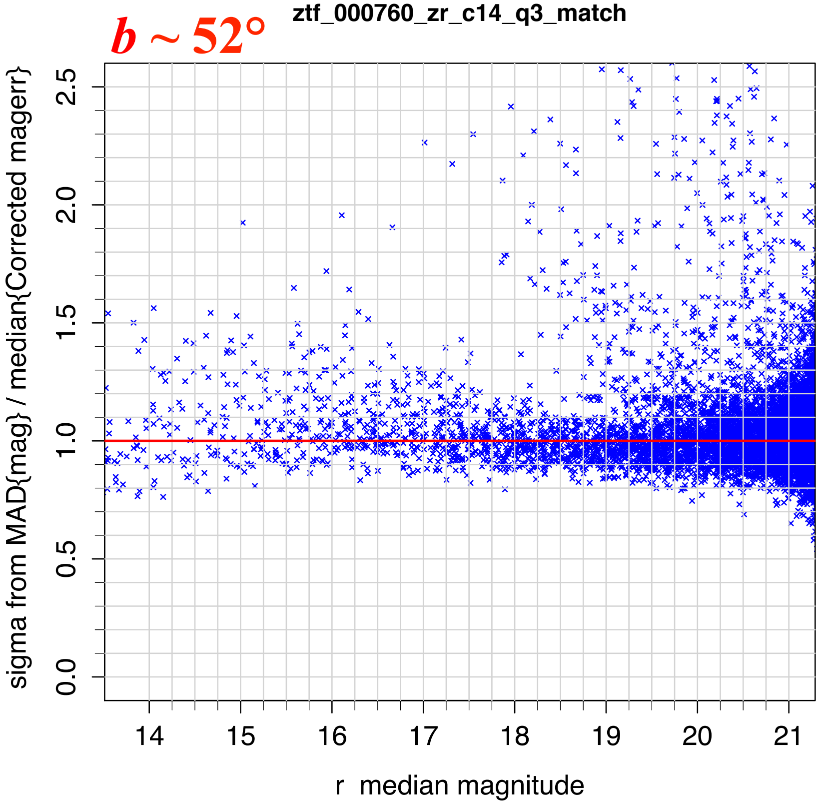
$$\sigma_{mag} = A + B * mag + C * 10^{0.4 mag} + D * 10^{0.8 mag}$$

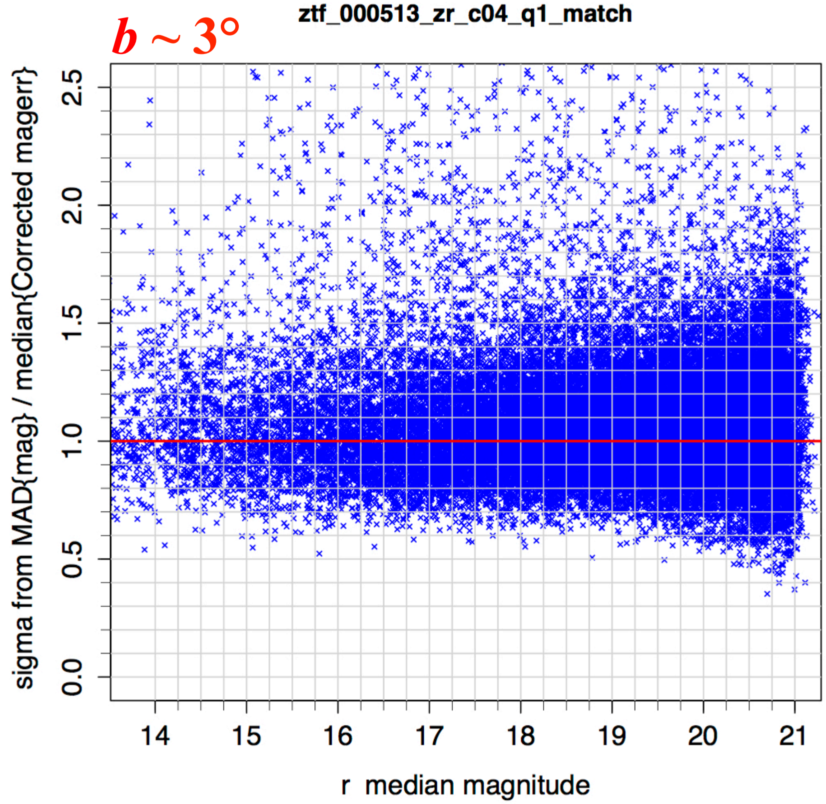
- Use this function to predict σ_{mag} for any epochal magnitude measured from that CCD-quadrant given A,B,C,D.





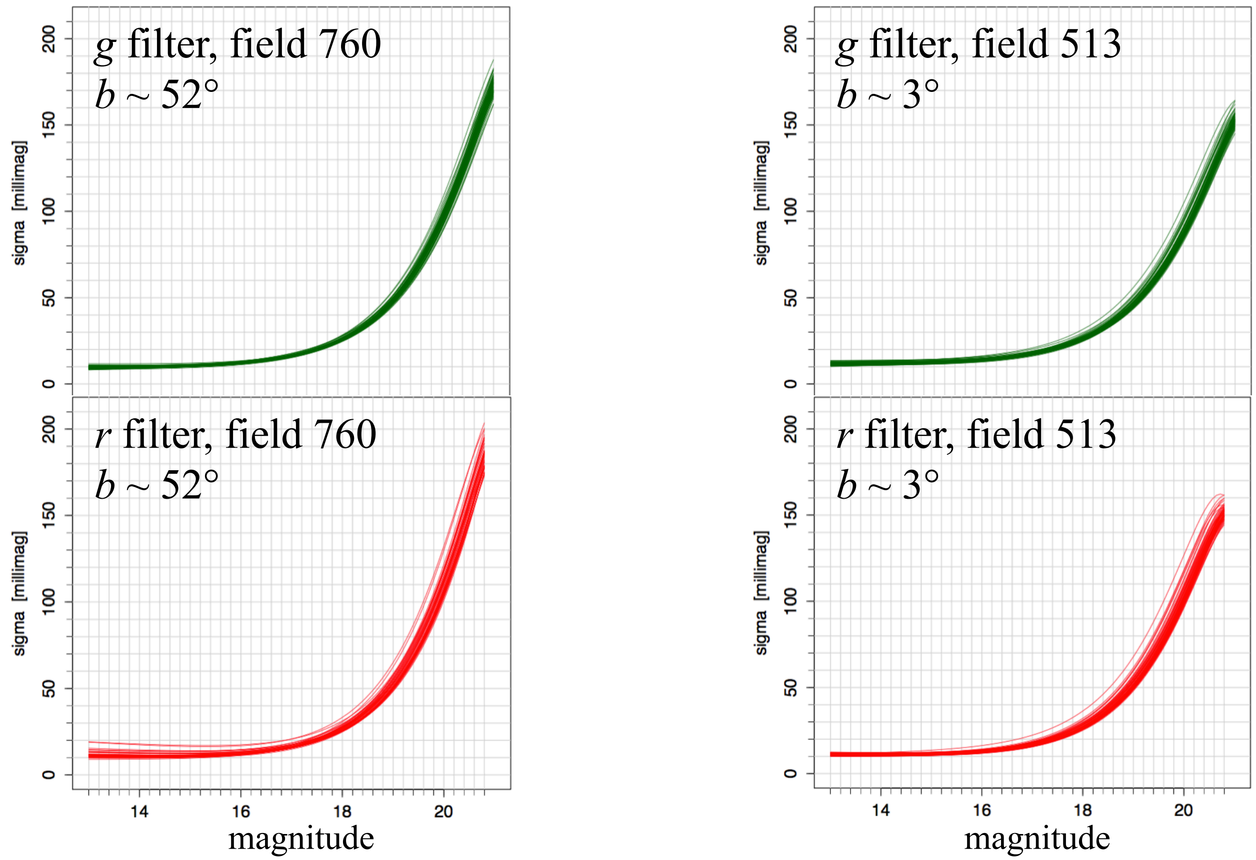
Photometric uncertainties: examples following correction





Photometric uncertainties from repeatability: global analysis

- Fit function across multiple fields and all CCD quadrants to explore variation versus environment and other parameters (airmass, seeing,…)
- Goal: find orthogonal/principle components of variation and have one formula for all of ZTF!



Data System Tasks: in progress & planned

- Preparations for Public Data Release 1 (previous slide).

- Update to Gaia DR2: for both astrometric calibration and alert association.
- Consider updating to Pan STARRS DR2: for both photometric calibration and alert association.
  - ➢ both require analysis, regression testing, regenerating static field-based catalogs, reformatting queries...

- Continue monitoring reference image quality:
  - ➢ recreate on a per-case basis if enough inputs satisfy older (tighter) criteria.

- Update pointing / WCS offset file to provide better CCD-quadrant priors.

- TBD: Setting up "Globus" to enable more efficient transfer of data across institutations.

Refinements contingent on support from partnership

- 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: <~ 2 millimag across FOV or <~ 0.3 millimag across CCD quad)

- *i*-filter fringe correction (DESY group input).

In Closing …

- A number of limitations and deficiencies have been identified and most are understood.
- Please continue reporting possible issues, however small you think they are.
- It is in our best interest to document everything for the community.

Back up slides

Reminder on documentation

- **ZSDS Explanatory Supplement** (linked from ZTF public website under):
  https://www.ztf.caltech.edu/page/technical#science-data-system

- **Science Data System paper**:
  https://iopscience.iop.org/article/10.1088/1538-3873/aae8ac

- **Archive access and services:**
  https://irsa.ipac.caltech.edu/Missions/ztf.html

- **Public alert archive and usage:**
  https://ztf.uw.edu/alerts/public/

Data Access / visibility policy

- Observing time during science operations will be split between three categories:
  - **Public** (NSF-funded MSIP survey: 40%)
  - **Private collaboration** (40%)
  - **Caltech TAC** (20%)
- Managed per exposure (epoch) using a *programID* propagated from scheduler to raw-image metadata

---

- Private/Caltech observers can access their data in near-realtime, soon after archive ingestion. This includes all calibration products and lightcurves from epochs tagged by their respective *programID*s queried via archive GUI.
- Public data will only be available at the public release times for general access by all.
  - raw images, processed epochal images, accompanying source-catalog files, difference images
  - reference images and catalog files
  - lightcurves constructed from public epochal data only
  - calibration data products
- Public alert packets (triggered from events detected in public exposures) will only contain public data. This includes their 30 day event histories.
- Private alert packets (triggered from events detected in private exposures) will contain public data in their 30 day event histories.
- Caltech alert packets (triggered from events detected in Caltech exposures) will contain data from all three programs in their 30 day event histories.
- No restriction on input data used to generate products for Solar System science: streaks & moving-object tracks; selected (human-vetted) products will be delivered to MPC.
- No restriction on input data used to generate reference image (co-add) products.
- No restriction on input data used to generate source match-files (lightcurve files):
  - MOU in place with the only customer of these products: Galactic Marshal
  - only privately-tagged and *already-released* public data therein to be ingested by Marshal

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.

- Improve quality of initial streak candidates prior to ML vetting downstream.

- Sandbox (work space) environment for users to perform analyses close to where the data resides.

Astrometric performance relative to Gaia

- Astrometric precision of bright stars with *r, g* < 18 mag at airmass < 1.2 is <~ 30 milliarcsec (RMS per axis).
- Accuracy for sources with S/N > 10 (*g, r* < 20 mag) at airmass < 2 is <~ 65 millarcsec.