The ZTF Science Data System (ZSDS) Explanatory Supplement

Pipelines, Definitions, Data Products, Access & Usage

Version 5.0, June 10, 2020

Frank J. Masci, Russ R. Laher, Ben Rusholme, David L. Shupe, Steven Groom, Jason Surace, Edward Jackson, Serge Monkewitz, Ron Beck, David Flynn, Scott Terek, Walter Landry, Eugean Hacopians, Vandana Desai, Justin Howell, Tim Brooke, David Imel, Stefanie Wachter, Quan-Zhi Ye, Hsing-Wen Lin, S. Bradley Cenko, Virginia Cunningham, Umaa Rebbapragada, Brian Bue, Adam A. Miller, Ashish Mahabal, Eric C. Bellm, Maria T. Patterson, Mario Jurić, Eran O. Ofek, Richard Walters, Matthew Graham, Mansi M. Kasliwal, Richard G. Dekany, Thomas Kupfer, Kevin Burdge, Christopher B. Cannella, Tom Barlow, Angela Van Sistine, Matteo Giomi, Christoffer Fremling, Nadejda Blagorodnova, David Levitan, Reed Riddle, Roger M. Smith, Andrew Drake, George Helou, Thomas A. Prince, Shrinivas R. Kulkarni

ZWICKY TRANSIENT FACILITY

Revision History

Date	Version	Author	Description
July 13, 2017	1.0	Frank Masci	Initial version
August 21, 2017	1.1	Frank Masci	Included description of ZMODE (moving object) pipeline, products etc.
September 22, 2017	1.2	Frank Masci	- included seeing estimation step in science cal pipeline; - updated PS1 selection cuts and plots for photometric calibration; - updated histograms on #PS1 sources per quadrant; - included step in img-diff pipeline to force noisy PSF pixels in sci and ref to zero.
October 11, 2017	1.3	Frank Masci	Updated archive directory structure description
November 7, 2017	1.4	Frank Masci	- outlined cal-image generation and application process; - added note on recentering of WCS solutions to quad centers - updated bit definitions, thresholds, and fatal bitstring - new sci image product: *_sciimgdaopsfcent.fits - added note that only pipeline status=1 products are archived - outlined
November 10, 2017	1.5	Frank Masci	- added GitHub link to schema description in *matchfiles* - clarified different types of "filterID" codes in filenames
November 20, 2017	1.6	Frank Masci	- included data-access URLs; - additional product for MPC deliveries: XML-formatted ADES files for streaks & tracklets
November 27, 2017	1.7	Frank Masci	- included halo-masking step in instrumental cal. pipeline - included some histograms on astrometric performance

December 11, 2017	1.8	Frank Masci	- included description of alert-packets and object naming - updated DB schema diagrams
December 14, 2017	1.9	Frank Masci	- updated input image criteria for creating reference-images - updated links to access archive products - added more deliverables / services to Table 1
January 8, 2018	2.0	Frank Masci	- updated input image filters for reference image creation - added linearity and halo-masking steps in instrumental cal flowchart with descriptions
January 19, 2018	2.1	Frank Masci	- included more detail for calibrating aperture photometry in sec. 10.2
January 25, 2018	2.2	Frank Masci	- updated section and equation numbers to be in order
February 27, 2018	2.3	Frank Masci	- added link to MOST API documentation - redefined fixed mag ZPs for reference images - added on-sky photometric performance plots (in progress) - updated criteria for selecting science images for reference image creation
March 4, 2018	2.4	Frank Masci	- added plots of astrometric rms versus airmass
March 9, 2018	2.5	Frank Masci	- added model-flux uncertainty plots to performance section - added photometric RMS repeatability plots to performance section
March 13, 2018	2.6	Frank Masci	- updated photometric RMS repeatability and model-based uncertainty plots from PSF-cat.
April 8, 2018	2.7	Frank Masci	- included description for predicting and masking of *counter*-moving ghosts - updated input criteria for reference-image generation - included figures to illustrate ops procedures and DB replication process

April 11, 2018	2.8	Frank Masci	- included description of alert-packet contents and metrics - included description of *matchfile* contents
April 23, 2018	2.9	Frank Masci	updates to contents of alert packets
May 11, 2018	3.0	Frank Masci	- updated processing details for ZMODE pipeline - update RMS-repeatability plots in Fig 19. - included regularization step for diff image PSF & expansion of masked pixels prior to findstreaks execution - minor updates to definitions of some metrics in alert packets
June 3, 2018	3.1	Frank Masci	- updated archive products to include tar files of alert packets - described additional cuts made on alert stream before leaving IPAC as well as proposed further purity cuts - included cautionary notes section for alert packet - updated *matchfile* ICD - restructured ztf-depot directory layout (Fig. 12)
July 3, 2018	3.2	Frank Masci	- updated threshold (*gpctdif: 15 to 80*) for selecting sci-images for reference image generation - updated alert packet cautionary notes (§9.5) - included more archive data links in §7.6 - updated SQL query for inputs to ZMODE pipeline (§6.9) - added 17 new metrics to Avro packets; includes nearest matches to *Gaia* DR1 sources - updated §6.6 to state that ref-image is rescaled, not science - updated Fig. 2 (ZADS flow)

August 31, 2018	3.3	Frank Masci	- added more cautionary notes on alert packet usage in §9.5 - mention use of different LED-exposure combinations when constructing flats in each filter - updated some preamble statements in introduction on intent of this document
January 2, 2019	3.4	Frank Masci	- included description of gain-match factor refinement in §6.6 - summarized new input criteria for reference image generation (§6.7) - updated DB query for events supplied to ZMODE pipeline - included more metrics in *prv_candidates* schema of alert packets (§9.2) - included new section (§10.6) describing metrics and quality flags in photometry catalogs - updated mag-limit maps in Figure 21 - updated column descriptions for *matchfile* products (§18) - added cautionary notes section (§13)
March 14, 2019	3.5	Frank Masci	- updated cautionary notes (§13); all subsections therein - updated *matchfile* column definitions in §18.
April 11, 2019	3.6	Frank Masci	- included *acknowledgments* text for users of products - included caveats in *matchfile* contents (§6.8). - updates to cautionary notes (§13) to conform with final DR1 products - further updates to *matchfile* columns (§18) following refinements to support DR1
April 27, 2019	3.7	Frank Masci	- further updates to cautionary notes (§13) to conform with final DR1 products
May 8, 2019	3.8	Frank Masci	- added URL in §13.4 to ZP thresholds for quality filtering

December 10, 2019	4.0	Frank Masci	Cleanup and numerous updates pertinent to Public DR2.
June 10, 2020	5.0	Frank Masci	Updates to §13 (Cautionary Notes) and §18 (Lightcurve Matchfile contents) to reflect content in Public DR3.

Table of Contents

1 INTRODUCTION

This document describes the algorithms, pipelines, products, their usage, and data-access services of the ZTF Science Data System (ZSDS). It is a "living" document that is progressively updated as advisories and tips on product usage are identified. Cautionary notes and caveats are summarized in §13. Acronyms are summarized in §14.

The ZSDS consists of the data processing pipelines, data archives, infrastructure for long-term curation, and the services for data retrieval and visualization. The ZSDS is housed at IPAC/Caltech.

1.1 Referencing and Acknowledging the ZSDS in Publications

For papers that use science products from the ZSDS, we appreciate you citing the following:

Masci, F. J., Laher, R. R., Rusholme, B., et al. 2018, *The Zwicky Transient Facility: Data Processing, Products, and Archive,* PASP, 131, 995.

This publication is available on Open Access: *https://dx.doi.org/10.1088/1538-3873/aae8ac*

Furthermore, please include the following in your acknowledgements:

2 HELPDESKS AND RESOURCES

For archive and data-access issues, which includes anomalies with any of the GUIs or APIs for data retrieval and analysis, please email:
irsasupport@ipac.caltech.edu

For questions related to data-quality, format, content, processing algorithms, or product usage, including documentation, please email:
ztfdatasystem@gmail.com

For an overview of a specific ZTF Public Data Release with links to data access and supporting documentation, see:
https://www.ztf.caltech.edu

3   DEFINITIONS

ZTF Archive – long-term storage of specific products with user-interfaces for data access.

ZTF Depot – short-term (night-to-night) storage of specific *near-realtime* pipeline products to support QA and Solar System working groups only. Access is via a web portal.
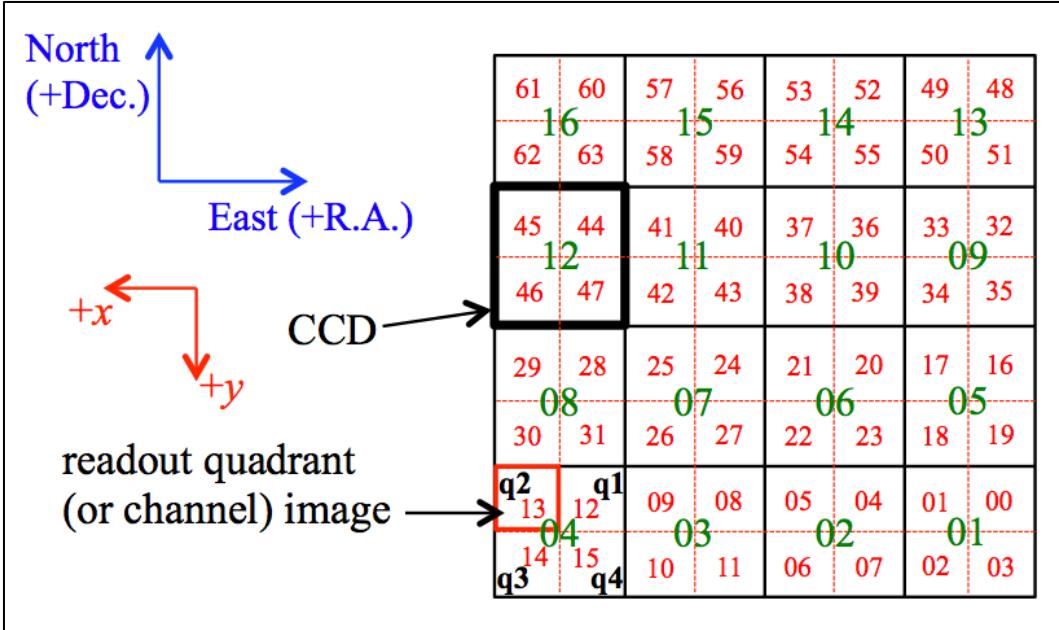
Alert Stream (or simply Alerts) – near-realtime *point-source* event packets derived from difference-image extractions that satisfy specific threshold criteria. Each packet consists of contextual metadata for a single event (e.g., source metrics and historical information) to facilitate filtering and follow-up downstream.

Users Database – stores information, metadata, QA metrics on all ZTF exposures acquired, file products in archive, and a history of all transient candidates with metrics (point source and streaking transients) extracted and posted on either ZTF Depot or through alert-distribution/broker. This is an internal pipeline-operations database and is restricted to specific project personnel.

CCD-based image – raw image data is received at IPAC per CCD; there are sixteen ~6k × 6k CCDs from one exposure of the focal plane; each CCD-based file is received in compressed multi-extension FITS format with each extension containing image data for a readout-quadrant. CCDs are identified by CCDID = 01…16. See Figure 1.

Readout-quadrant (or readout-channel) image – a ~3k × 3k image produced after splitting the CCD-based image file. This is the basic image-unit for pipeline processing and from which all data products are derived. Readout channels on a single exposure of the focal are identified using RCID = $4(\text{CCDID} - 1) + \text{QID} - 1$, where CCDID was defined above and QID = 1…4, the quadrant ID in an individual CCD. See Figure 1.

Realtime – a processing step or task performed close to acquisition of the image data; typically 15 to 25 minutes from time of observation.

Figure 1: The ZTF focal plane showing CCD/readout-channel layout and numbering viewed from the back onto the sky. Green numbers (01..16) are CCD IDs; the labels q1,q2,q3,q4 denote the readout-channel ordering per CCD (same for all CCDs) and packaged in this order in the multi-extension raw-image FITS files; red numbers (00..63) denote the 64 readout-channel IDs (used for internal database bookkeeping). The *x, y* vectors denote the directions of increasing pixel coordinates that all readout quadrants must have when the FOV is orientated on the sky as shown.

4   DELIVERABLES AND SERVICES

Table 1 summarizes the primary products delivered by the ZSDS, formats, timescale of availability, access portal (archive, depot or other), and Sections in this document for further information. Generic names of all file products are given in Section 7.

Product/service	Details	Formats	Availability	Access	Section
raw images with metadata	CCD based; overscans inc.; exposure metadata	fpack'd FITS split into extensions	realtime	GUI or API to archive	6.1, 6.2, 7.1, 7.6
epochal science images and catalog files	processed images; masks; PSFs; difference images; diff-image PSFs; PSF-fit and aperture phot. catalogs	FITS images with metadata; FITS binary tables; ASCII files; diff images are *fpack*-compressed	realtime	GUI or API to archive	6.5, 6.6, 7.2, 7.6
reference images (coadds) and catalog files	coadd image; uncert. image; depth/coverage map; PSF-fit and aperture photom catalogs	FITS images with metada; FITS binary tables	updated weekly/bi-weekly; contingent on quality of inputs & fields visited	GUI or API to archive	6.7, 7.4, 7.6
calibration file products	bias and responsivity maps; uncert. images; some QA	FITS images with metadata	once per day, prior to observing night	GUI or API to archive	6.3, 6.4, 6.10, 7.6
source *matchfiles* (or lightcurve files)	photometrically refined lightcurves & metrics for detected objects in epochal images using PSF-fit catalogs	Pytable (HDF5) formatted files	updated monthly and sometimes bi-monthly; contingent on number of new epochs	Internal (restricted audience). See LCDB GUI service below for generic access	6.8, 7.5, 7.6, 18

Lightcurve DB (LCDB)	spatial/temporal searches and/or metric filtering to retrieve LCs from *matchfiles* using object-based DB queries	Database content	contingent on available data from *matchfiles* (see above)	GUI or API to LCDB	6.8, 7.6, 18
Users DB (replicate of operations DB)	access to all operations metadata; QA metrics; full history on transient candidate metrics	Database content	realtime; refresh rate is close to realtime)	Internal only	11
QA products and survey coverage monitoring	pipeline QA metrics and survey depth-of-coverage maps	ASCII files, FITS and PNG images for coverage maps	realtime; end-of-night for coverage map updates	ZTF-Depot: internal filesystem webserver	8
Solar System related products	For time-critical follow-up: metrics/cutouts for diff-image detected streaks and moving-object tracklets from linked point sources; for MPC delivery	ASCII files and JPEG cutouts	realtime	ZTF-Depot: internal filesystem webserver	6.6, 6.9, 8
Moving-Object Search Tool (MOST)	Return image metadata for an input known Solar System object	ASCII files	realtime	GUI or API to archive	7.6

Alert Stream (AS)	point-source transient event stream pushed to broker network	Avro-formatted packet files containing metadata and image cutouts	realtime	UW and ZTF Alert Brokers	6.6, 8, 9

Table 1: primary deliverables and products

5   OVERVIEW OF PROCESSING FLOW

Below is a summary of the nine pipeline subsystems in the ZSDS. Under each subsystem are the primary output products, their destination if pertinent (archive, depot or alert stream), and frequency of generation. Figure 2 gives a schematic of these subsystems and their inter-dependencies. A high-level design of the overall ZSDS components and interfaces is shown in Figure 3.

A listing of the archive and depot file products with generic filenames are given in Sections 7 and 8 respectively. Details of pipeline processing and algorithms are described in Section 6.

5.1   Raw data ingestion and pre-processing

1. Raw data ingest, archival of raw CCD image files and storage of metadata in database;
   - *raw CCD image files*;
   - *archive*;
   - *realtime*

2. Raw-image decompression, splitting of CCD images into readout-quadrant images, floating bias corrections;
   - *simple QA in database*;
   - *internal only;*
   - *realtime*

5.2   Calibration generation

3. Bias-image derivation from stacking calibration images acquired in afternoon;
   - *image file per readout-quadrant*;
   - *archive*;
   - *made before on-sky operations during day*

4. High-ν flat (pixel-to-pixel responsivity) from stacking calibration images;
   - *image file per readout-quadrant*;
   - *archive*;
   - *made before on-sky operation during day*
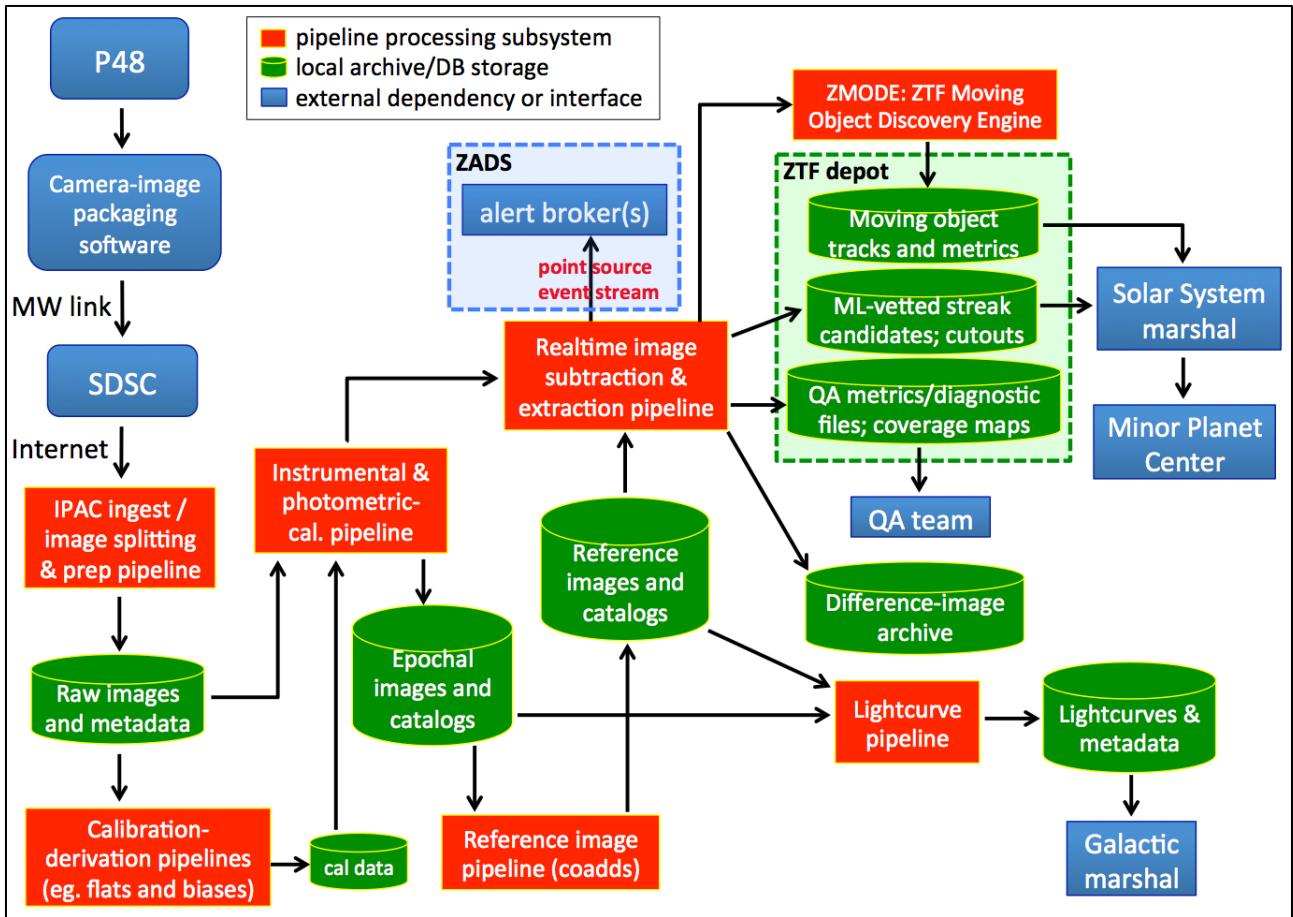
5.3   Real-time science-data processing

5. Instrumental calibration of readout-quadrant images: astrometric and photometric calibration;
   - *epochal science images and source catalogs per readout-quadrant;*
   - *archive;*
   - *realtime*

6. Image subtraction and transient extraction (point-sources and streaks for "fast" moving asteroids); uses outputs from pipelines 5 and 7;
   - *difference images, PSFs, metadata and cutouts;*
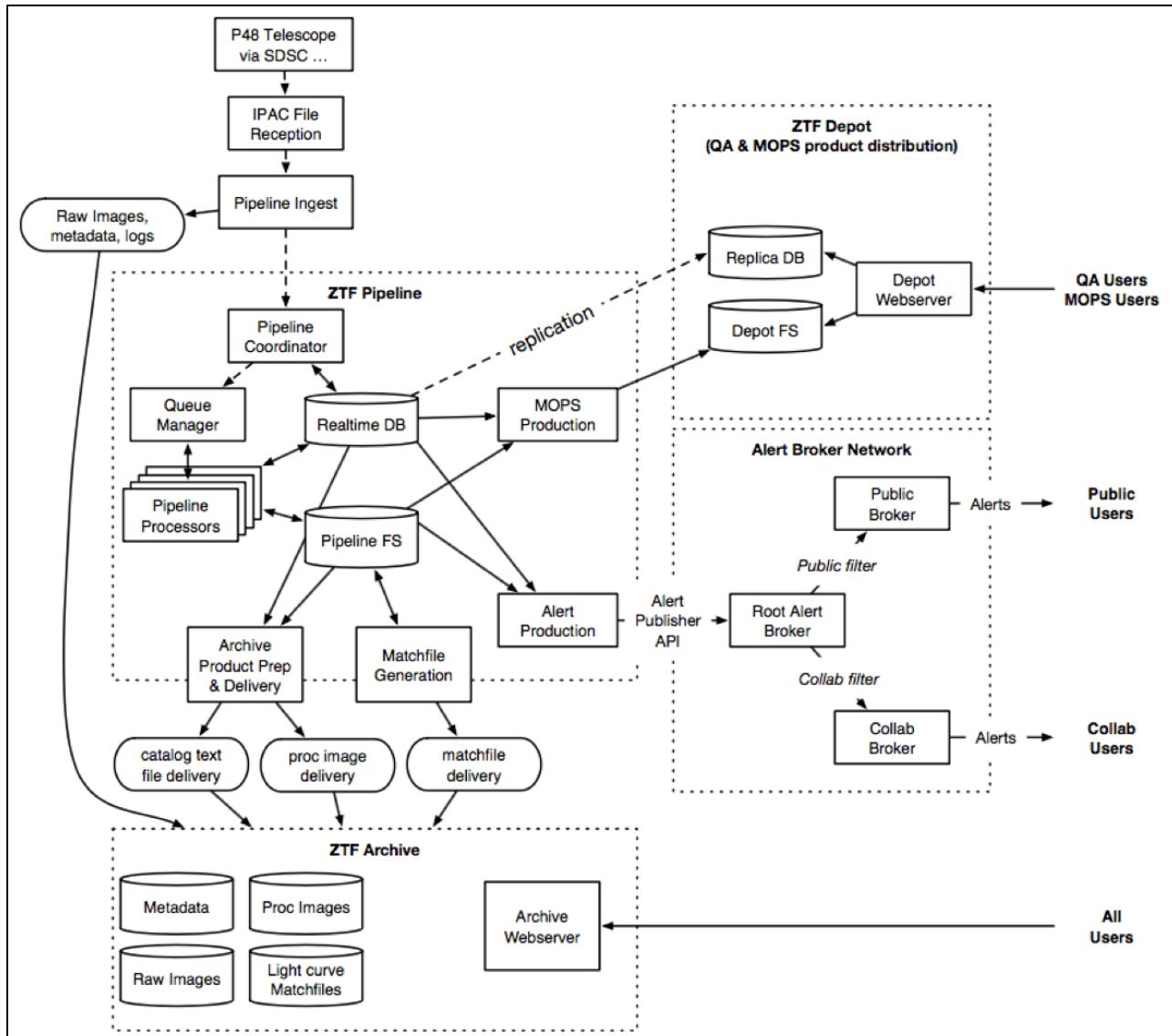
- alert stream for point-source events; depot for streaks; archive for difference images
- realtime

5.4   Ensemble-based processing (combining epochal science products)

7. Reference-image generation (co-addition of epochal images generated by pipeline 5);
   - images and source catalogs per readout-quadrant;
   - archive;
   - as needed until reference-image library complete, when good quality data available

8. Source-matching across epochal source catalogs with relative photometric refinement for lightcurves; inputs are from pipelines 5 and 7;
   - "matchfile" (lightcurve) products per readout-quadrant;
   - archive;
   - updated every month or longer (contingent on number of new epochs available)

9. Moving Object Discovery Engine (ZMODE): construction of moving-object tracks from linking machine-learned vetted point source transients from pipeline 6;
   - tracklet files and orbit-fit metadata with QA metrics;
   - depot;
   - at the end of each observing night using all data from the previous three nights at most if available, otherwise use the most recent one or two night(s)

Figure 2: Pipeline-level data flow in the ZSDS. Green components indicate storage in the form of an archive and/or database residing at IPAC to serve either the public, science marshals (via ZTF-depot) or the collaboration. Specific products are summarized in Section 7. Red components represent the core pipelines. Further details on the pipelines are given in Section 6. Blue components indicate external interfaces or dependencies.

Figure 3: High-level design of the ZSDS, showing databases, file systems, and interfaces. The "ZTF Pipeline" and "ZTF Depot" boxes are expanded in Figure 2.

6   CORE PIPELINE PROCESSING STEPS

Below we outline the algorithms and processing steps in each of the core pipelines summarized in Section 5 and illustrated in Figure 2. These are high-level descriptions; details are available upon request.

All pipeline and database software, and configuration parameter files are available in GitHub:
https://github.com/Caltech-IPAC/ztf/tree/master
Access to this repository is restricted to ZTF project team members and partners.

6.1   Raw data ingestion and preparation

1. Receive raw Rice-compressed multi-extension FITS-formatted camera-image files. One file per CCD is pushed from P48 to *ztfinbox* staging area at IPAC. Each FITS file consists of nine extensions (HDUs): a primary global header with exposure/CCD-specific metadata; four extensions for the light-sensitive pixels with metadata (one per readout-channel image: 3072×3080 pixels); another four extensions for the overscan pixels with metadata (corresponding to each readout-channel image: 30×3080 pixels).

2. Loop over CCD files, check filename format and file type (i.e., has *.fits.fz* suffix); verify CHECKSUM and DATASUM keyword values for all HDUs; set *status* = 0 and raw file *infobits* |= 1 if cannot be verified or if verification check fails. Ensure values of specific telescope / camera telemetry keywords in primary header are non-zero and within range; if not, update *status* and *infobits* flags.

3. Within the same loop, read and verify that IMGTYPE value (e.g., = 'object', 'bias', 'flat', …) is consistent with expected DOMESTAT value (= 'open' or 'close'); if not, update *status* and *infobits* flags. Regardless of the outcome of all checks, copy raw CCD file to long-term archive; store all metadata from primary header; assign and compute additional image/exposure identifiers; load all metadata, identifiers, raw archive filename into operations database for later retrieval. Filename format of raw image file is given in Section 7.1. Database table schemas associated with *ingest* process are shown in Section 11.1. Cleanup *ztfinbox* area.

4. Signal generated by Virtual Pipeline Operator (VPO) script that interfaces with *ingest* to initiate first phase of pipeline processing: CCD-file decompression and splitting into readout-channel based FITS files (Section 6.2). This is contingent on check/quality flags assigned above.

6.2 Raw image decompression, splitting, and floating-bias corrections

1. Query database for raw (prior-ingested) CCD file and metadata; create a field-dependent local working directory if not already created and copy CCD file from archive.

2. Decompress CCD file using *funpack* utility to create multi-extension FITS (MEF) file. Pixels are converted from 16-bit unsigned integers to 32-bit floating point.

3. Split the MEF file and extract the four readout-channel (light-sensitive) images; append primary HDU from CCD file to each readout-channel specific FITS header so can propagate to all products downstream. Consolidate redundant keywords.

4. Extract the four corresponding readout-channel overscan images from MEF file and store in memory. At the time of writing, each overscan strip has $x \times y$ dimensions of 30×3080 pixels.

5. Compute floating bias correction from each readout-channel overscan image using a subset of the columns within each strip. The pixel range used is $x_{min}{:}x_{max}, y_{min}{:}y_{max} = 4{:}27, 1{:}3080$ for all readout-channels. For a given overscan strip, the pixels within this range are first collapsed across the column (x) dimension using a median to create a 1-D array; a polynomial of degree two (quadratic) is then fit to this collapsed 1-D array.

6. Correct the light-sensitive readout-channel pixels by subtracting their corresponding overscan polynomial fits from step 5 on a column-by-column basis. Write overscan-corrected readout-channel images to FITS files with specific headers propagated from above.

7. Compute QA statistics on the raw readout-channel images, accompanying overscan regions, and floating-bias correction fits (residuals) for later loading into the database.

8. Copy overscan-corrected raw readout-channel image files to operations filesystem for later retrieval.

9. Store readout-channel image filenames, specific header keywords, and QA metrics from 7 in the *RawImages* table of the database (schema is given in Section 11.2).

6.3 Bias-image calibration generation

1. Inputs are: a list of overscan-corrected images corresponding to zero-exposure bias-frames acquired prior to observing night for a specific readout-channel. A minimum of 10 *new* bias-frames are required before triggering this pipeline. By *new*, the input frames were not already used to create a previous bias-image calibration product.

2. Stack (collapse) the N bias-frames using a trimmed-average per pixel; the outlier rejection is symmetric and the criterion for rejecting pixel p_i from a stack is $|p_i - median\{p_i\}| > 2.5\sigma$.

21

The σ here is computed using half the '84.13 – 15.86' percentile difference, i.e., a measure that is relatively robust against outliers. Five image products are generated: primary bias-image calibration product (from trimmed stack-average); stacked standard-deviation image *A* (with outliers removed); stacked standard-deviation image *B* (*before* outlier rejection to support additional masking below); 1-σ uncertainty image (= *StdDev image A* / $\sqrt{N_{trimmed}}$); and an image of the number of surviving pixels in the stack following outlier rejection.

3. Compute QA metrics and statistics on products generated in step 2. These include column- and row-collapsed statistics on the primary bias-image product in order to capture spatial variations for later trending.

4. Use stacked standard-deviation image *B* from step 2 (*before* outlier rejection) to find 'noisy' pixels. This is only performed if the number of input frames used was ≥ 4. If so, 'noisy' pixels are identified as those with *StdDev* > 7 × *median*{all *StdDev* values}. Pixels satisfying this criterion are tagged with bit 4 (value $2^4 = 16$) in a mask image. This mask image has filename suffix *_biascmask.fits* and accompanies the bias-image calibration products from step 2. These are later used by the instrumental calibration pipeline (Section 6.5).

5. The bias-image calibration products are copied to the operations filesystem and archived for later retrieval. Filename formats of all archived products are given in Section 7.3. The filenames along with all QA metrics from step 3 are stored in the *CalFiles*, *CalAncilFiles*, and *CalImQA* tables of the database (schema is given in Section 11.2).

6.4 High-ν flat (pixel responsivity map) calibration generation

1. Inputs are:
   a. List of overscan-corrected images corresponding to exposures of the illuminated *flatfield-screen* acquired prior to observing night for a specific readout-channel and filter. A minimum of 20, 20, and 21 *new* flatfield frames in filters *g*, *r*, *i* respectively are required before triggering this pipeline. By *new*, the input frames were not already used to create a previous flatfield-image calibration product. The minimum number of input exposures are determined by the optimum number of different LED-colors (with characteristic peak wavelengths) that need to be combined for each filter. At the time of writing, the breakdown in *g* and *r* is 5 exposures in each of 4 available LEDs, and for *i*, it's 7 exposures in each of 3 LEDs. The LED settings for each exposure are specified in the FITS headers of each delivered raw CCD image file.
   b. Pixel mask (*pmask*) image that records prior-known bad detector pixels for this readout-channel.
   c. Bias calibration image and accompanying *cmask* products from pipeline described in Section 6.3.

2. Loop over the *N* input flatfield frames and perform the following to each:
   a. Subtract bias-image calibration image generated by pipeline in Section 6.3.
   b. Perform linearity correction. Corrected counts (#DN) per pixel are computed using:

$DN_{corrected} = DN / (A_c\,DN^2 + B_c\,DN + 1),$

where DN are the input (observed) counts and A_c, B_c are the coefficients from a quadratic model fit to laboratory data. There is one set of coefficients per CCD quadrant. The correction is typically $<\sim$ 1% over the observable dynamic range.

  c. Normalize the pixel values to a *spatially*-trimmed global mean of one. The trimming is symmetric where pixels p_i satisfying $|p_i - median\{p_i\}| > 3\sigma$ are rejected. The σ here is computed using half the '84.13 – 15.86' percentile difference, i.e., a measure that is relatively robust against outliers. Pixels tagged in the input *pmask* and *bias-cmask* are also excluded when computing the mean.

3. Stack (collapse) the N spatially-unit-normalized flatfield frames using a trimmed-average per pixel; the outlier rejection is symmetric and the criterion for rejecting pixel p_i from a stack is $|p_i - median\{p_i\}| > 2.5\sigma$. The σ here is computed using half the '84.13 – 15.86' percentile difference. The following intermediate image products are generated: trimmed stack-average; stacked standard-deviation image A (with outliers removed); stacked standard-deviation image B (*before* outlier rejection to support additional masking below); 1-σ uncertainty image (= *StdDev image A* / $\sqrt{N_{trimmed}}$); and an image of the number of surviving pixels in the stack following outlier rejection.

4. Use stacked standard-deviation image B from step 3 (*before* outlier rejection) to find additional 'noisy' pixels (i.e., not caught by the bias-image generation pipeline; Section 6.3). This is only performed if the number of input frames used was ≥ 4. If so, 'noisy' pixels are identified as those with $StdDev > 1.85 \times median\{$all $StdDev$ values$\}$. Pixels satisfying this criterion are tagged with bit 4 (value $2^4 = 16$) in a mask image. This mask image has filename suffix _hifreqflatcmask.fits_ and accompanies the flatfield-image calibration products.

5. Divide-out any extrinsic lamp (LED)-induced illumination patterns from the trimmed stack-averaged image generated in step 3, i.e., spatial variations not associated with the detector or optical-system response. Currently, this correction map is an image filled with ones (i.e., no correction is applied). If found to be significant later, any extraneous lamp/LED patterns will need to be fit and removed.

6. Normalize the pixels in the trimmed stack-averaged image generated in step 3 to a *spatially*-trimmed mean of one. The trimming is symmetric where pixels p_i satisfying $|p_i - median\{p_i\}| > 3\sigma$ are rejected. The σ here is computed using half the '84.13 – 15.86' percentile difference. Pixels tagged in the input *pmask* and *bias-cmask* are also excluded when computing the mean. This normalized image then becomes the final high spatial-frequency flatfield image (or pixel-to-pixel responsivity map) to be used downstream.

7. Compute QA metrics and statistics on products generated in steps 3 and 6. These include column- and row-collapsed statistics on the primary flatfield-image product in order to capture spatial variations for later trending.

8. The high-ν flatfield-image calibration products (from steps 3 and 6) are copied to the operations filesystem and archived for later retrieval. Filename formats of all archived products are given in Section 7.3. The filenames along with all QA metrics from step 7 are stored in the *CalFiles*, *CalAncilFiles*, and *CalImQA* tables of the database (schema is given in Section 11.2).

6.5   Instrumental calibration of raw science image data

Unless otherwise stated, all parameters and thresholds mentioned in this section are located in the following configuration file on GitHub:
https://github.com/Caltech-IPAC/ztf/blob/master/cdf/instrphotcal.txt
Access to this repository is restricted to ZTF project team members and partners.

Figure 4 shows an overview of the processing flow.

1. Inputs are:
    a. A single overscan-corrected readout-channel (CCD-quadrant) image for a specific filter.
    b. Pixel mask (*pmask*) that records prior-known bad detector pixels for this readout-channel.
    c. Closest-in-time (most recently acquired/generated) bias calibration image (Section 6.3).
    d. Closest-in-time (most recently acquired/generated) high spatial-frequency flatfield calibration image (Section 6.4).
    e. Calibration mask (*cmask*) accompanying the bias calibration image (containing additional transient/noisy pixels; Section 6.3).
    f. If available, low spatial-frequency flatfield ("*starflat*") responsivity image. This input is pre-generated offline (external to the pipeline) on longer timescales using post-processed archived source catalogs.
    g. Readout-channel specific catalog of preselected calibrator sources to support photometric calibration.
    h. Readout-channel specific catalog of astrometric calibrator sources to support astrometric calibration.
    i. Configuration and processing parameters.

2. Combine ("OR") the two mask image inputs (*cmask* and *pmask*) to initialize the science pixel bit-mask array to propagate and update below. The possible bit-mask values are defined in Section 10.3. Saturated pixels in the *raw* science image are also located and tagged in the mask at this stage. We also initialize the "INFOBITS" string to track quality and processing details at the image level. This string is also updated below. All image INFOBITS are defined in Section 10.4.

3. If image exposure was acquired at MJD ≤ 58779 (2019-10-23T00:00:00 UTC; prior to a camera-waveform update), identify "charge-spillage" ghost artifacts[1] and tag in accompanying mask.

4. Subtract the bias calibration image (input 1c).

5. Divide by the *high spatial-frequency* flatfield calibration image (input 1d). Also locate pixels with a low or high relative response in the flatfield image and tag them in the science pixel mask image.

6. Divide by the *low spatial-frequency* flatfield calibration image (input 1f). Default is an image filled with ones.

7. Perform linearity correction. The corrected counts (#DN) per pixel are computed using:
$DN_{corrected} = DN / (A_c\,DN^2 + B_c\,DN + 1)$,
where DN are the input (observed) counts and A_c, B_c are the coefficients from a quadratic model fit to laboratory data. There is one set of coefficients per CCD quadrant. The correction is typically $<\sim 1\%$ over the observable dynamic range.

8. Compute robust Slowly Varying Background (SVB) and smoothed RMS images from science image. Also compute QA metrics. These are used to support the following processing steps below: (i) aircraft/satellite streak-finding; (ii) derivation of a pixel uncertainty image; and (iii) computation of a signal-to-noise ratio image (for later cosmic-ray detection); (iv) derivation of parameters for PSF estimation, source extraction and photometry. The robust SVB and RMS images are computed using outlier-trimmed medians and standard-deviations in image partitions across a 10×10 grid. Median filtering is used to smooth-out partition boundaries. "Minimal" background and robust pixel-RMS values over all partitions are also computed to support the steps mentioned above. By minimal, we mean relatively uncontaminated by the presence of sources and background structure.

9. Derive true pixel saturation level (in DN) in science image following processing steps above. This uses the pre-masked raw image saturated pixels located in step 2 to find the lowest saturated pixel value (for conservatism). This value is propagated along and used in the source extraction step below.

10. Compute a background-subtracted signal-to-noise ratio image using the outputs from step 8. This image is later used to support cosmic-ray detection and masking.

11. Detect aircraft and satellite streaks (tracks) in the science image and tag their location in the mask. This uses the *createtrackimage* software and is based on finding linear patterns in connected pixels whose flux exceeds a specific threshold relative to the local background (SVB) and RMS computed in step 8. For more details, see Laher et al. (2014).

---
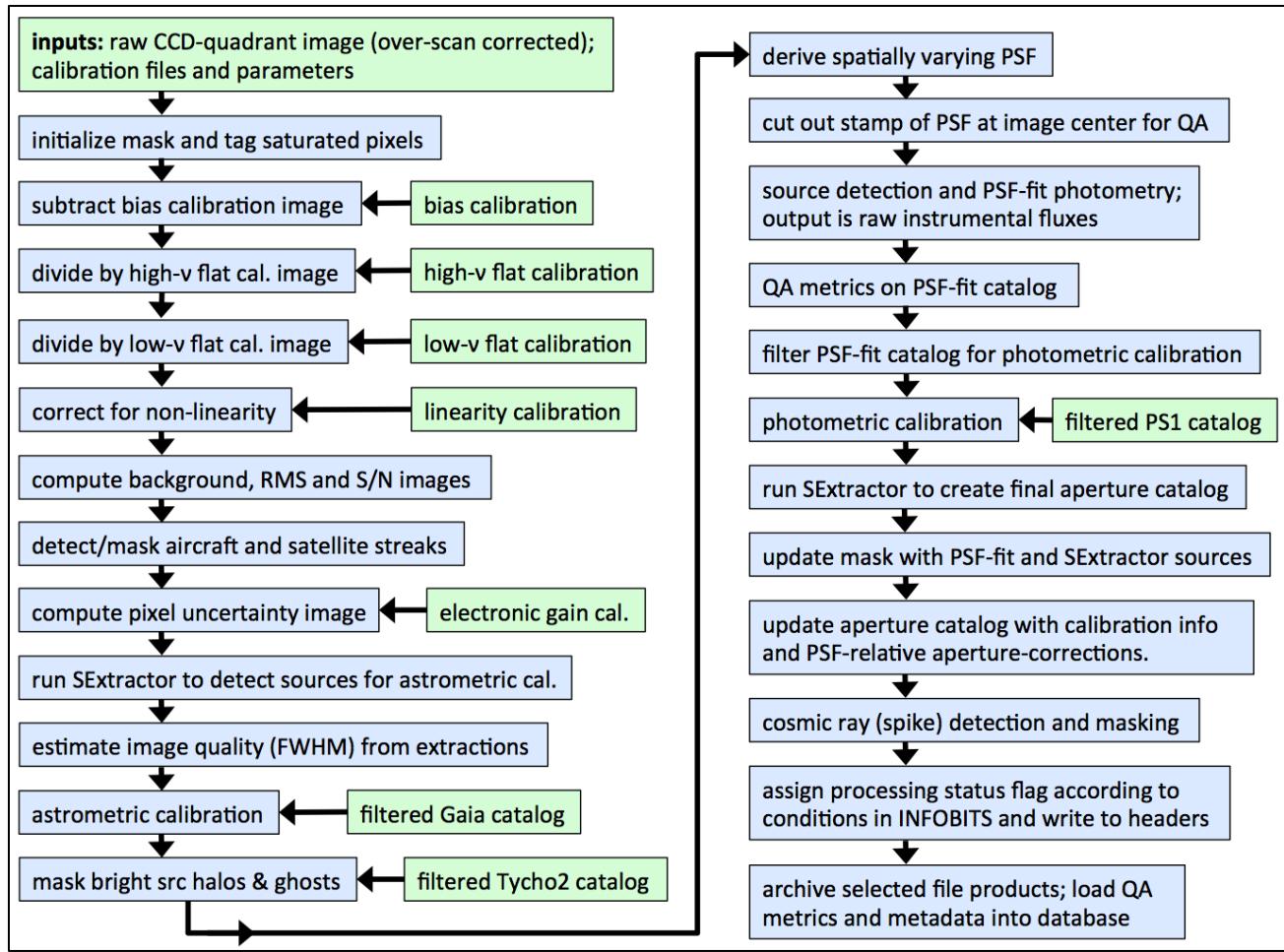
[1] A description can be found in *http://nesssi.cacr.caltech.edu/ZTF/Web/Ghosts.html*

12. Compute pixel uncertainty image using the outputs from step 8. This is used as input to the two *SExtractor* runs downstream (steps 13 and 23).

13. Execute *SExtractor* for the sole purpose of generating an initial source list for input into the astrometric calibration below (sources are detected using a relatively high threshold). The output source file is in FITS_LDAC format, as required by the *SCAMP* software below. Update image INFOBITS string if the number of sources is too low to support astrometric calibration below.

14. Estimate the readout quadrant image quality (FWHM) from the *SExtractor* catalog generated in step 12. This initial estimate is used as input to other processing steps below, e.g., PSF estimation.

15. Perform astrometric calibration using the source list from step 13, external astrometric reference catalog supplied in 1h, prior WCS pointing and distortion parameters. For methodology, see Section 6.5.1. Image INFOBITS are also updated for cases where an astrometric solution could not be derived or is too inaccurate according to pre-defined critiera.

16. Detect and mask halos around bright stars. These halos are also referred to as *co-moving ghosts*. Stars are first located using a subset of the Tycho-2 Catalogue of Bright Stars (Høg et al. 2000). This subset consists of stars with a *V*-band magnitude of ≤ 6.5 – the approximate limit where halos start to appear above the background. These stars are heavily saturated in the nominal 30sec exposures and therefore not detected by the pipeline. The next step involves masking a circular region of radius specified by parameter *bshaloradius*. Characterization of on-sky data shows that this radius is independent of the parent star's *V*-band magnitude. This radius is conservative in the sense that it is padded with additional pixels to account for astrometric error and slightly non-centered halos. Halos from stars falling outside an image but which overlap with the image are also masked. Halo-contaminated pixels are tagged with a special bit in the science image mask (see Section 10.3).

17. Detect and mask ghost artifacts from bright stars falling in other quadrants of the focal plane. These ghosts are also referred to as *counter-moving* ghosts, to distinguish from co-moving ghosts that are responsible for halo artifacts (see step 16). A description of the ghost-location prediction and masking algorithm is given in Appendix B. Ghosts from "parent" stars whose Tycho-2 catalog positions fall outside a focal-plane edge or within a CCD gap are not masked. Therefore, expect to see unmasked ghosts close to CCD edges. Furthermore, due to inaccuracies in the predicted positions of ghosts in exposures observed at high airmass (primarily due to atmospheric refraction effects), there is no ghost-masking at airmass > 2.2. Ghost-contaminated pixels are tagged with a special bit in the science image mask (see Section 10.3).

18. If input image was acquired in *i-filter*, correct for atmospheric fringing.

19. Estimate spatially-varying PSF using an updated and automated version of *DAOPhot / AllStar*. A linearly-varying Gaussian basis is assumed with corrections written to a look-up-table. For details, see Masci et al. (2017). A stamp of the PSF at the center of the readout quadrant is generated in FITS format and archived. Image INFOBITS are also updated for cases where a PSF could not be derived.

20. Detect sources and perform PSF-fit photometry using an updated and automated version of *DAOPhot / AllStar*. Output raw catalog file contains *instrumental* photometry with astrometrically calibrated positions. Also return arrays storing positions of pixel "spikes" (possible cosmic-rays) for later filtering (step 29).

21. Compute QA metrics on raw PSF catalog from 20, including magnitude limits and median chi-square values from PSF-fitting. Image INFOBITS are also updated where pertinent.

22. Filter raw PSF catalog from 20 to retain "clean" sources above a pre-defined S/N threshold and greater than some distance from image boundary. This filtered source list is used to support photometric calibration below.

23. Perform photometric calibration using the filtered PSF catalog from 22 and external calibrator catalog from 1g. For methodology, see Section 6.5.2. Image INFOBITS are also updated for cases where a photometric solution could not be derived or is too inaccurate according to pre-defined criteria.

24. Update header of initial PSF-fit catalog file from 20 with photometric calibration information and metrics. Convert to FITS binary table format. Details on how to use photometric calibration solutions are given in Section 10.1.

25. Execute *SExtractor* to generate aperture photometry catalog containing instrumental photometry. Also compute QA metrics. Calibration information from step 23 and relative PSF-to-aperture corrections are computed and added to the catalog header below.

26. Update science image pixel mask with locations of source detections from PSF-fitting (step 20) and *SExtractor* (step 25).

27. Compute *SExtractor* aperture magnitude offset corrections relative to PSF-fit magnitudes using a filtered subset of the extractions from each catalog. Corrections are computed for all aperture sizes in the *SExtractor* catalog. For details on how to interpret and apply these corrections, see Section 10.2.

28. Write photometric calibration keywords (step 23) and aperture correction information (step 27) to *SExtractor* catalog FITS catalog header.

29. Record pixel "spikes" (possible cosmic-rays) in science image mask that were detected by *DAOPhot* in step 20. These were initially detected using the *sharp* metric, defined as the ratio of heights of the best fitting 2D delta function to the best fitting 2D Gaussian function

(with prior FWHM from the derived PSF). Values satisfying $1 < sharp \leq 2$ represent profiles significantly narrower than the PSF and are candidates for cosmic-rays. The S/N values for all pixels within a 3×3 square region centered on each candidate spike are then computed and thresholded. Only those 3×3 regions with ≤ 3 pixels above the S/N threshold will have such pixels tagged in the science mask. The reason for this is to avoid tagging pixels associated with real source signal and/or neighboring sources.

30. Assign an overall processing quality status flag according to the INFOBITS string. This string is checked for the existence of specific "fatal" conditions (e.g., inability to perform astrometric and/or photometric calibration) as specified by *fatalinfobits*. This flag is written as the STATUS keyword to the science image FITS header. For details on how this is set, see Section 10.4.

31. Write the image INFOBITS definitions (Section 10.4) with their specific thresholds and all photometric calibration information to the science image FITS header. Also write the pixel bit-mask definitions (Section 10.3) to the science mask FITS header.

32. Selected science-image products (with filenames summarized in Section 7.2) are then archived. The product filenames along with all QA metrics computed above are stored in the *ProcImages*, *AncilFiles*, and *ImQA* tables of the database (schema is given in Section 11.2).

Figure 4: High-level processing flow in the instrumental calibration pipeline. This represents the first phase of the real-time pipeline. For details, see Section 6.5.
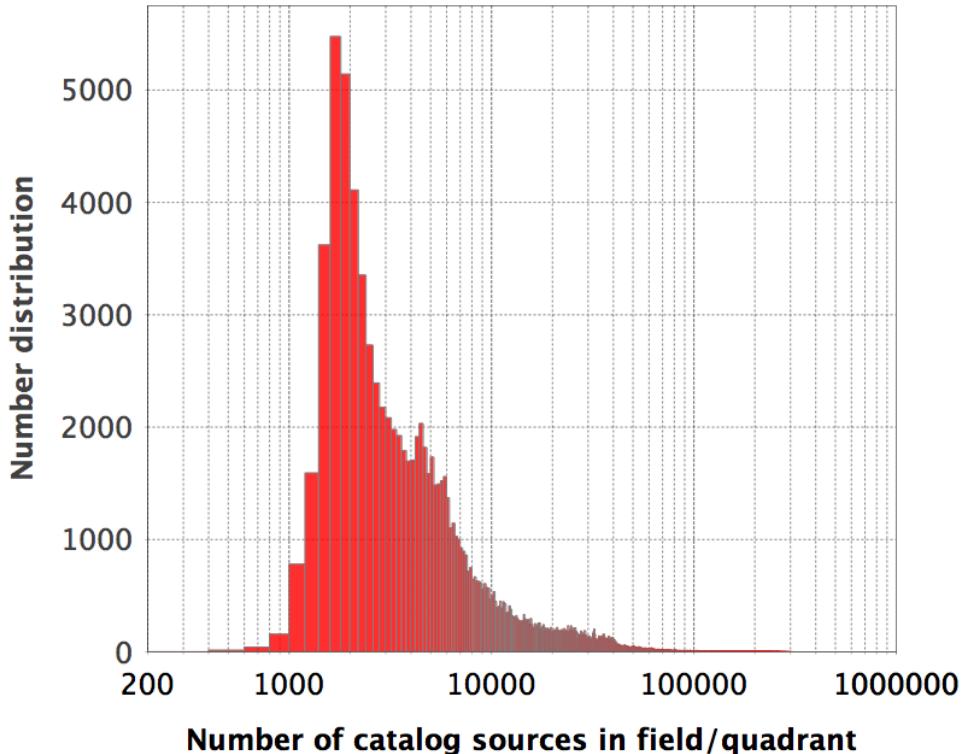
6.5.1 Astrometric calibration

The astrometric solution uses the *SCAMP* software from the *Astromatic* suite. This is executed in two passes as described below. The solution is calculated on a per-readout-channel basis using a pre-calibrated distortion prior and initial pointing derived from the Telescope Control System (TCS). This prior pointing is present in the input image FITS header. The steps are as follows.

1. Read and store inputs:
   a. Path/filename to the pre-processed science image (following step 11 above).
   b. Path/filename to the FITS_LDAC-formatted file containing sources extracted from the image in 1a (from step 12 above).
   c. Path/filename of the *Gaia* reference source table, also in FITS_LDAC format. This table pertains exclusively to the field/readout-channel image being processed and was pre-generated by partitioning the entire *Gaia* DR1 catalog according to the known all-sky field tiling with readout-channel boundaries therein. These partitioned *Gaia* tables are distributed across multiple fileservers for fast access during parallel processing. A histogram of the number of *Gaia* astrometric reference sources per field/readout-channel in shown in Figure 5.

2. Set up the image WCS *prior* for the first run of *SCAMP*, as an ASCII header file. This file is constructed from WCS header keywords from the input image, and a carefully constructed distortion prior specifying the optical distortion and alignment of the CCDs.

3. Check that the WCS prior and reference (*Gaia*) catalog overlap as expected.

4. Set the position angle tolerances for *SCAMP*:
   a. if Dec < 50 degrees, tolerance is 0.1 deg.
   b. if 50 deg < Dec < 70 deg, tolerance is 1.0 deg.
   c. if Dec > 70 deg, tolerance is 3.0 deg.

5. Run the first pass of *SCAMP*. This run is only for pattern-matching with parameter settings:
   a. STABILITY_TYPE is set to PRE-DISTORTED, meaning that *SCAMP* will only use the distortion terms in the WCS prior.
   b. The maximum change in pixel scale is set (currently 2%).
   c. A maximum positional uncertainty of 15 arcminutes is allowed.

6. Parse the *SCAMP* output to store some QA metrics, including the number of input source detections and reference sources used, the chi-squared and the uncertainties along RA and Dec.

7. Run the second pass of *SCAMP*. This second run disables any pattern matching, and computes a 4th-order polynomial distortion solution.

8. The *raw* WCS solution from above is centered in the FOV of the optical system (telescope boresight). The WCS solution is now transferred to the center of the readout-channel image,

30

i.e., with reference pixel coordinate: CRPIX1,CRPIX2 = 1536.5,1540.5. This includes re-centering the coordinate-system origin representing the distortion polynomial. This new WCS solution is written to the output science image header.

9. Parse the *SCAMP* output for QA metrics.

10. Rematch the input detected and reference catalogs based on the final WCS solution, and compute additional QA metrics from the comparison. Processing now continues with step 15 of the instrumental calibration pipeline (Section 6.5).

Figure 5: distribution of the number of *Gaia* sources per readout-channel-based catalog to support astrometric calibration (Section 6.5.1). Sources were queried from *Gaia* DR1 with no filtering.

6.5.2　Photometric calibration

We have selected a subset of point sources from the Pan STARRS1 (PS1) DR1 catalog to support "basic" photometric calibration of the individual ZTF readout-channel images in *realtime.* That is, each of the 64 individual readout-channels from each camera exposure, observed through filter *f*, is assigned its own zeropoint (ZP_f) and color coefficient term (c_f), where $f = g$, r, or i.

Averaged PSF-fit photometry from the PS1 *MeanObject* table (Flewelling et al. 2017) is used to select photometric calibrators. Our goal here is to select predominately stable stars, i.e., showing no variability across multiple PS1 epochs. PS1 does not include any star-galaxy score of flags. We therefore derived a conservative *star-only* retention filter using the prescription in Rest (2017). This is reflected in the cuts below. To minimize instances of temporal variability, the PS1 multi-epoch standard-deviation measure (*gmeanpsfmagstd*) was fit as a function of magnitude (*gmeanpsfmag*) and only PS1 sources within an upper envelope of the distribution were selected (see below). To further mitigate temporal variability, we included color and magnitude cuts to select stars away from the instability strip that crosses the Main Sequence.

The criteria used to select photometric calibrators from PS1 are below. Criteria (1) to (9) are combined in the **AND**'d sense. Figures 6-8 illustrate some of the criteria.

1. *imeanpsfmag* ≥ 14.0

2. $-1.5 \leq$ *imeanpsfmag – imeankronmag* ≤ -0.04

3. *gmeanpsfmag – imeanpsfmag* ≤ 3.0

4. *gmeanpsfmag – imeanpsfmag* ≥ 0.2
   for $18.0 \leq$ *gmeanpsfmag* ≤ 21.0
   OR
   gmeanpsfmag – imeanpsfmag $\geq (2.771429 - 0.142857$*gmeanpsfmag*$)$
   for $14.5 \leq$ *gmeanpsfmag* < 18.0

5. signal-to-noise ratio of ≥ 2 in all relevant PS1 filters (provides security against bad measurements):
   $0 <$ *gmeanpsfmagerr* $\leq 1.086/2$
   $0 <$ *rmeanpsfmagerr* $\leq 1.086/2$
   $0 <$ *imeanpsfmagerr* $\leq 1.086/2$
   $0 <$ *zmeanpsfmagerr* $\leq 1.086/2$

6. The number of single-epoch PS1 measurements used to compute the mean magnitudes is ≥ 6, i.e:
   gmeanpsfmagnpt ≥ 6
   rmeanpsfmagnpt ≥ 6
   imeanpsfmagnpt ≥ 6
   zmeanpsfmagnpt ≥ 6

7. The *g*-filter standard-deviation in single-epoch magnitude measurements satisfies the following bound (dependent on *gmeanpsfmag*). This mitigates source variability and is derived from a polynomial fit (see Figure 8 below).

$gmeanpsfmagstd \leq$

$-59562.88132240652 +$

$26033.76397640325 gmeanpsfmag -$

$4904.584447388286 gmeanpsfmag^2 +$

$516.1652478744579 gmeanpsfmag^3 -$

$32.65924784026809 gmeanpsfmag^4 +$
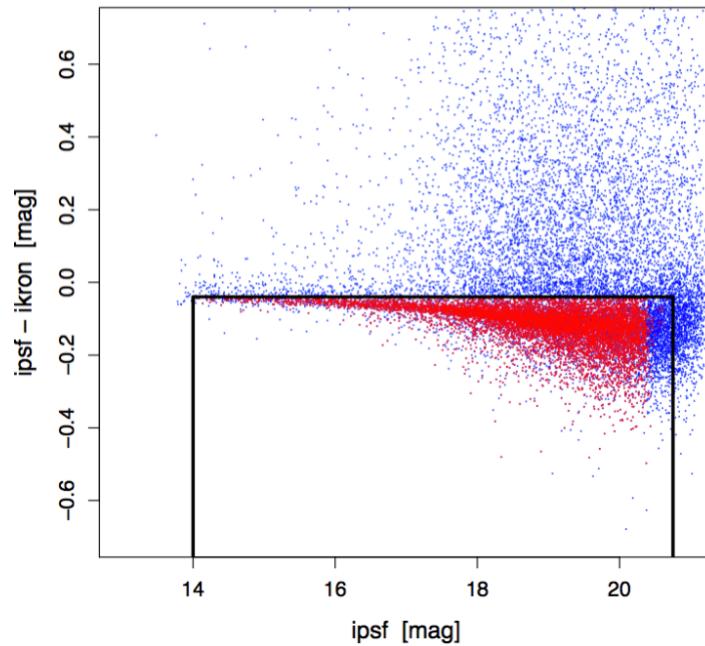
$1.221620051175976 gmeanpsfmag^5 -$

$0.02279306506024598 gmeanpsfmag^6 +$
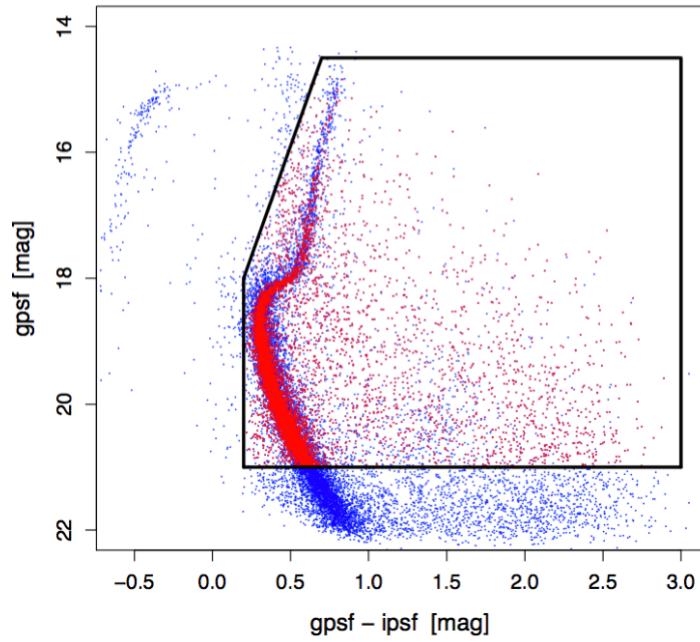
$7.5975787200012634e\text{-}06 gmeanpsfmag^8 -$

$9.384153316363922e\text{-}08 gmeanpsfmag^9.$
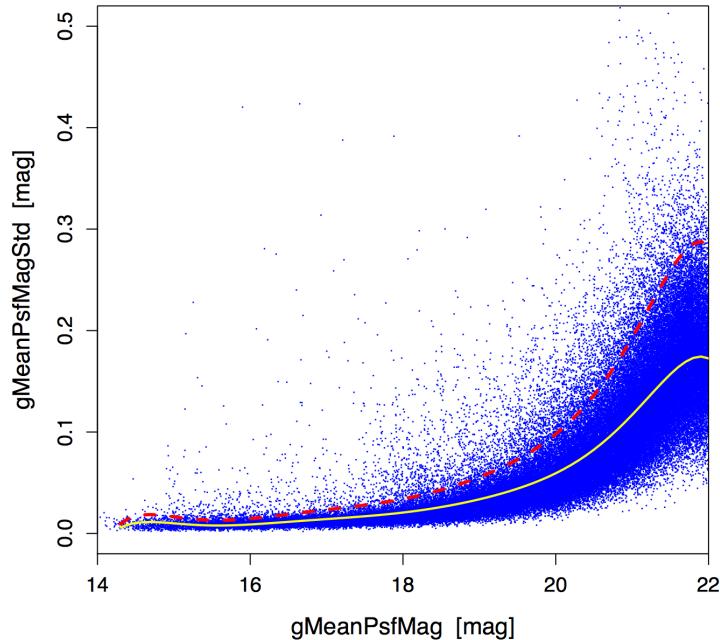
8. No other PS1 source with *gmeanpsfmag* $\leq$ 22.0 within 3.5 arcsec of its position and with no other cuts imposed. This mitigates confusion / contamination. This magnitude limit is $\approx$ 1.5 mag deeper than the 5σ limit expected for ZTF in *g*.

9. To speed up the confusion-avoidance processing (criterion 8), retain the 3000 brightest calibrators with *gmeanpsfmag* >= 14.5 following cuts (1) to (7) within each ($\approx$ 51arcmin x 51arcmin) readout-channel image. This is crucial in/near the galactic plane. Following the confusion processing (8), the number of calibrators remaining within the densest readout-channel images is $\approx$ 350 to 400.

Figure 6: *imeanpsfmag – imeankronmag* **versus** *imeanpsfmag* **for a 30 arcmin radius centered on the M13 globular cluster. Photometric calibrators are indicated by the red points and bounded by the black lines using the criteria in Section 6.5.2. All photometry is from the PS1** *MeanObject* **table.**
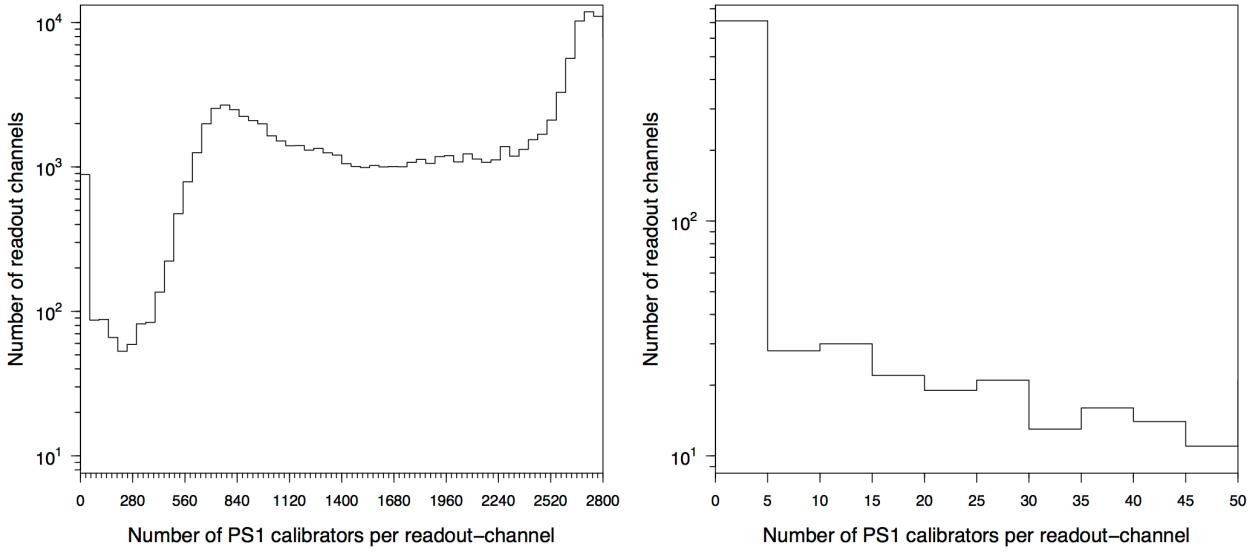
Figure 7: g*meanpsfmag* **versus** g*meanpsfmag – imeanpsfmag* **for a 30 arcmin radius centered on the M13 globular cluster. Photometric calibrators are indicated by the red points and bounded by the black lines using the criteria in Section 6.5.2. All photometry is from the PS1** *MeanObject* **table.**

Figure 8: Multi-epoch standard-deviation in PSF-fitted g-band magnitude (*gmeanpsfmagstd*) versus g*meanpsfmag* for a 2 degree radius centered on the M13 globular cluster. The yellow solid line is a robust polynomial fit and red dotted line is an upper bound. This bound is one of the criteria used for selecting photometric calibrators from the PS1 *MeanObject* table (criterion 7 in Section 6.5.2).

The above selection criteria are applied to a PS1 working database to create "static" catalog table files for use in the production pipeline. These table files are distributed across multiple fileservers for fast retrieval in the parallel processing environment. This pre-generation method enormously reduces runtime and avoids loads on the database.

Figure 9 shows a histogram of the number of PS1 calibrators per readout-channel (all *g, r, i* filters) according to the above criteria. 111,745 ZTF readout-channels from 2,194 science and engineering (test) fields covering the PS1 3π survey footprint had ≥ 5 photometric calibrators per readout-channel. 703 readout-channels from 93 distinct fields had < 5 photometric calibrators and hence, all images associated with these fields/readout-channels will not receive a calibration solution. This is because the minimum number of calibrators required per readout-channel is 5. The southern-most declination limit of a readout-channel image (center) that can be calibrated appears to be ~ − 31.62046°.

Figure 9: *Left:* **full distribution of the number of PS1 calibrators per readout-channel extracted from the *MeanObject* table using the criteria in Section 6.5.2.** *Right:* **a zoom-in on the range 0 to 50. These numbers are based on the nominal ZTF sky-tiling grid.**

Denoting the PS1 photometry as g^{PS1}, r^{PS1}, i^{PS1} and the ZTF instrumental photometry as g, r, i, the equations that are fit to solve for ZP_f, c_f per readout-channel are as follows:

$$g^{PS1} - g = ZP_g + c_g(g^{PS1} - r^{PS1})$$
$$r^{PS1} - r = ZP_r + c_r(g^{PS1} - r^{PS1})$$
$$i^{PS1} - i = ZP_i + c_i(r^{PS1} - i^{PS1})$$

(Eq. 1)

At the time of writing, it is not known whether $i^{PS1} - z^{PS1}$ color from PS1 is more optimal for calibrating the *i*-filter color term.

Prior to cross-matching between PS1 and ZTF extractions, we only match to calibrators falling at > 30 pixels from an image edge and within the expected unsaturated PS1 *r*-magnitude range: 14.5 < *rmeanpsfmag* < 20.0. Furthermore, we only match to ZTF extractions with S/N > 10 in the respective filter. Following matching, the input quantities $f^{PS1} - f$ (on the left) and $m_1^{PS1} - m_2^{PS1}$ (on the right) are symmetrically trimmed to guard against outliers before performing the linear regression. The minimum required number of source matches following trimming for the linear regression is three. Less than this, a condition is recorded in INFOBITS (Section 10.4) warning that photometric calibration was not possible for that readout-channel.

36

Information and metrics relating to the photometric calibration solution are written to the epochal-based image and catalog FITS headers. For example:

```
CALNOTE1= '*** photometric-calibration keywords below are from PSF-fitting:'
CALNOTE2= '*** will also be applicable to aperture measurements on point'
CALNOTE3= '*** sources using psf_mag-aper_mag corrections below (if available)'
NMATCHES=                 1802 / Number of photometric calibrators used
MAGZP   =            27.492246 / Magnitude zero point for PSF-photometry [mag]
MAGZPUNC=          0.0000002992 / Magnitude zero point uncertainty [mag]
MAGZPRMS=             0.005527 / Magnitude zero point RMS deviation [mag]
CLRCOEFF=             0.001152 / Color coefficient from linear fit
CLRCOUNC=          0.0000005617 / Color coefficient uncertainty from linear fit
ZPCLRCOV=           -0.00000040 / Covariance in MAGZP and CLRCOEFF [mag^2]
PCOLOR  = 'g - R   '           / Calibrator color used for PSF-fit calibration
```

For details on how to apply the photometric calibration solutions to the instrumental photometry present in the ZTF catalog products (or on any new photometry extracted from the images), see Section 10.1. Other metrics are also written to the headers. These are statistics that collapse the PS1 – instrumental measurements along the individual magnitude difference and color axes, i.e., ZPMED, ZPAVG, ZPRMSALL, CLRMED, CLRAVG, CLRRMS.

6.6 Image subtraction and transient (event) extraction

Unless otherwise stated, all parameters and thresholds mentioned in this section are located in the following configuration file on GitHub:
https://github.com/Caltech-IPAC/ztf/blob/master/cdf/imgdiffextract.txt
Access to this repository is restricted to ZTF project team members and partners.

Figure 5 shows an overview of the processing flow.

1. Inputs are:
    a. An instrumentally calibrated readout-channel science image (with astrometric and photometric calibration solutions in FITS header); from pipeline in Section 6.5.
    b. Pixel bit-mask image corresponding to science image, also from Section 6.5.
    c. PSF-fit source catalog corresponding to this science image, also from Section 6.5.
    d. Reference (*co-add template*) image from pipeline described in Section 6.7.
    e. PSF-fit source catalog corresponding to reference image, also from Section 6.7.
    f. Configuration and processing parameters, paths to MPC asteroid ephemeris files and machine-learned training-model data files for point-sources and streaks.

2. Store pixel and FITS header metadata from image inputs: 1a, 1b, and 1d. Use the input specified fatal-bits (*fatbits*) to mask against and convert input science mask to a new mask with NaNs for bad-pixels and ones for good (usable) pixels. At the time of writing, saturated pixels are not masked. This will allow saturated transients to be detected from difference images downstream. This regularized mask image will be multiplied with final products at end of processing.

3. Store data from reference image PSF-fit catalog (input 1e) and map the *x, y* positions of sources therein to *x, y* positions in the science image coordinate frame. This frame will later

become the difference-image coordinate frame. These remapped reference-source positions (and associated metrics) will be used to (i) compute the relative gain-match factor between science and reference images, and (ii) cross-match reference-image sources to transient candidates extracted later on.

4. Filter the remapped reference-image sources from step 3 to retain a subset of clean *point-sources* within a specified magnitude, PSF-fit chi-square, and *sharpness* range. We also only retain reference-image catalog sources that have no other source within a square region of side length 2×*refexclbox* pixels and that are at a distance > *edgebuffer* from an image edge. This subset of sources will be used to derive the relative gain-match factor below. We also store position and flux information from the science image PSF-fit catalog (input 1c).

5. Compute the gain-match factor f_g to rescale the reference image pixel values to those in the science image. This is computed from the median ratio of science image PSF-fit source flux to reference image PSF-fit flux, following source matching. We require a minimum of 30 matches to proceed. As a fallback, if there are insufficient matches satisfying the filters above, the input image photometric zeropoints (ZP) are used to derive the gain-match factor via $f_g = 10^{0.4(ZPsci - ZPref)}$. This estimate is not as accurate as the relative source-match method since it relies (indirectly) on the absolute accuracy of the photometric calibration upstream. This initial value represents a coarse estimate of the gain-match factor. This is later refined by searching for trial gain values within an interval $1 \pm dsearch$ in increments of *dscl* and retaining the correction factor (δf_g) that gives a minimum in the median sum of pixel values within cutouts centered on source positions from the initial gain matching above. We use the brightest *nusesrcmax* sources and pixels are summed within square cutouts of half-side length *cuthalflen* pixels. The refined gain-match factor is $f_g \times \delta f_g$.

6. Multiply the reference image pixel values by the refined relative gain-match factor f_g from 5. Given the reference image was rescaled to match the effective photometric throughput in the science image, the difference-image derived therefrom will take on the *science* image `MAGZP` value.

7. Resample and interpolate the reference image pixels onto the *distorted* science image coordinate frame. This uses the WCS (and science-image distortion) information present in the image headers. The pixel interpolation uses a 2D Lanczos kernel of window size three, as implemented in *SWarp* (Bertin et al. 2002). The outputs are a new reference-image (now in the science-image frame) and a weight-image that is later converted to an effective mask for use downstream.

8. Convert the weight-image corresponding to the resampled reference image from step 7 into a mask image consisting of NaNs for bad-pixels and ones for good (usable) pixels. At this stage, we combine (multiply) the pre-filtered science mask image from step 2 and this reference-image mask to generate an effective difference-image mask for use below. For conservatism, this effective mask is also convolved with a 5×5 constant kernel for more complete blanketing of possible neighboring bad pixels (e.g., from bright/saturated source

bleeds and local artifacts). This implies that an additional $5 \times 5 - 1$ pixels per input bad pixel are masked.

9. Compute a robust *differential* Slowly Varying Background (SVB) image from a preliminary "science – reference" difference image, where the science image was gain-matched to the reference image (step 6) and the latter was resampled onto science frame (step 7). I.e., PSF-matching has not yet occurred. Bad pixels are masked using the effective mask generated in step 8. This *differential* SVB is later used to match relative background-level variations between the science and reference images. These variations will then be mostly absent in the final difference image. In brief, the SVB is computed using outlier-trimmed medians in image partitions across a 10×10 grid of the preliminary difference image. Median filtering is used to smooth-out partition boundaries. For details, see Masci et al. (2017).

10. Subtract the *differential* SVB image computed in step 9 from the regularized science image (output from step 6).

11. Compute robust background-level and pixel noise-sigma in regularized science image (from step 10) using *minimally-contaminated* image partition from SVB computation above. Use these to compute a pixel-uncertainty image for the regularized science image.

12. Generate a PSF template image from the regularized (background-subtracted) science image using an updated and automated version of *DAOPhot / AllStar*. To mitigate spurious signals in the final difference image due to noise in the PSF image, we force all pixels at radius $\geq$ *keeppsffac* $\times$ FWHM$_{sci}$ to zero, where *keeppsffac* is an inflation factor. The PSF image is then normalized to a sum of unity. QA metrics are also computed.

13. Compute robust background-level and pixel noise-sigma in resampled reference image (from step 7) using *minimally-contaminated* image partition from SVB computation above. Use these to compute a pixel-uncertainty image for the resampled reference image.

14. Generate a PSF template image from the resampled (background-subtracted) reference image using an updated and automated version of *DAOPhot / AllStar*. To mitigate spurious signals in the final difference image due to noise in the PSF image, we force all pixels at radius $\geq$ *keeppsffac* $\times$ FWHM$_{ref}$ to zero, where *keeppsffac* is an inflation factor. The PSF image is then normalized to a sum of unity. QA metrics are also computed.

15. Using the regularized science and reference images (from steps 10 and 7 respectively), corresponding pixel uncertainty images (steps 11 and 13 respectively), and corresponding PSF template images (steps 12 and 14 respectively), perform PSF-matching and image-differencing using the ZOGY method (Zackay et al. 2016). There are five primary outputs:
    a. "*science – reference*" difference image corrected for correlated noise;
    b. effective PSF image for difference image in (a);
    c. "*science – reference*" difference image *not corrected* for correlated noise;
    d. effective PSF image for difference image in (c);

      e. point-source match-filtered S/N (*Scorr*) image pertaining to difference images from (a) and (c).

16. The difference image from 15a is rescaled to restore the effective pixel-noise and photometric throughput (ZP) propagated from the input science and reference images. The scale factor is $\sqrt{(\sigma^2_{sci} + \sigma^2_{ref})}$, where the sigmas are robust estimates from steps 11 and 13 above. We also generate negative (*reference – science*) difference images by multiplying the outputs from 15a and 15c by –1 with a corresponding negative *Scorr* image (equivalent to 15e). We then apply (multiply) each of these image products with the effective mask generated in step 8. The difference image PSF (15b) is regularized by replacing negative and outlier pixels therein with those from the science image PSF (output from 12). It is then renormalized to a sum of unity. The difference image from 15a is used to compute quality metrics (including features for the extracted events), while the difference image from 15c is only used to perform PSF-fit photometry on the extracted events using the PSF in 15d (see below). The difference image from 15c was found to produce more stable photometry several months into the survey.

17. Compute QA metrics on the final difference image from 15a (following rescaling in step 16). Also compute a pixel-uncertainty image for this difference image using a combination of the inputs used to compute the science and reference image pixel uncertainties (steps 11 and 13). This pixel-uncertainty image will be used to compute photometric uncertainties for extracted transient candidates below.

18. Assign a status-quality flag for the final difference image: *status* = 0 (bad), otherwise = 1, implying difference image is "good" or usable for transient extraction downstream. *Status* = 0 is set if (i) the input science and/or reference images were of bad/low quality according to specific fatal bits (*fatalinfobitssci* and *fatalinfobitsref*) in their INFOBITS strings (from pipelines 6.5 and 6.7), **OR** (ii) if the actual difference image does not satisfy a number of quality criteria determined by thresholding specific metrics before (*bef*) and after (*aft*) PSF-matching. The criteria used for (ii) are:
      a. The mean-squared value in difference image pixels/overall variance, i.e., $mean\{d^2_i\}_{aft}/\sigma^2_i > diffavgchisqaftthres$ **AND**
      b. The median-squared value in difference image pixels: $median\{d^2_i\}_{aft} > diffmedsqaftthres$ **OR** its relative change before/after PSF-matching: $(median\{d^2_i\}_{aft} / median\{d^2_i\}_{bef}) – 1 > diffmedsqchgthres$.
If *status* = 0, processing stops and no transient candidates are extracted from the suspected "bad" difference images (both positive and negative counterpart). This avoids computational bottlenecks downstream when attempting to extract and disseminate the inevitable large number of false-positives. The *status* flag value is written to the difference image FITS header as the STATUS keyword.

19. If *status* = 1, extract and filter point-source transient candidates from the positive "science – reference" difference image. The process consists of:
      a. The point-source match-filtered S/N (*Scorr*) image from 15e is used in *SExtractor* to first detect sources with S/N > 5; then within the same *SExtractor* run, measure

aperture fluxes (in DN) on these detected positions in the difference flux-image (from 15a only). The difference image pixel-uncertainty image (from 17) is used to compute photometric uncertainties.

b. Store contents of the output *SExtractor* table and omit bad measurements according to specific bits (*omitsexflags*) in the output FLAGS values. At this stage, we have N_{raw} extracted candidates. If N_{raw} exceeds some input (typically large) threshold, *ncandrawthres*, we abort and do not proceed to filter candidates for later database loading. This check is also intended to catch a possible "bad" difference image, i.e., not caught by the image metrics in step 18. This avoids computational burdens downstream since a large N_{raw} will increase processing runtimes in a non-linear manner. This includes automated vetting.

c. For each raw transient candidate from 19b, use the following *SExtractor* output metrics (computed on the difference image from 15a) to retain possible "good" transient candidates:

(*mindedge* > *diffimedgetol*) **and**
(*sexsnr* > *sexsnthres*) **and**
(*elong* ≤ *elongthres*) **and**
(0 < *fluxapratio* ≤ *fluxapratiothres*).

Each line above corresponds to respectively: the minimum distance to any image edge; the S/N ratio using an 8-pixel diameter fixed concentric aperture (nominally > 5); the source elongation (A/B); and the flux ratio using two aperture sizes: *fluxapratio* = 8-pixel diameter aperture flux / 18-pixel diameter aperture flux.

d. For each transient candidate that passes the criteria in 19c, count the local number of negative (*nneg*) and bad (*nbad*) pixels within a 5 × 5 pixel cutout centered on each, and a new metric, *sumrat* (see below). Then threshold these three metrics to further retain possible "good" candidates:

(*nneg* ≤ *nnegthres*) **and**
(*nbad* ≤ *nbadthres*) **and**
(*sumrat* > *sumratthres*).

The *sumrat* metric is defined as the ratio of pixel sums (values p_j) in a 3 × 3 median-filtered image of an input 5 × 5 pixel cutout on the difference-image with input values p_i:

$$sumrat = \frac{\sum_j p_j \ \left(= med\{p_i\}_{3\times3}\right)}{\sum_j |p_j| \ \left(= |med\{p_i\}_{3\times3}|\right)}.$$

The denominator differs from the numerator by summing the absolute values instead. Under the null hypothesis of pure Gaussian noise, the ±1σ range expected for *sumrat* is −0.25 <~ *sumrat* <~ 0.25. This approaches a maximum of one when real signal (significantly above the noise) is *exclusively* present, for example, when no positive/negative "cancelling" residuals in a yin-yang pattern are present.

e. For each transient candidate that passes the criteria in 19d, proceed with PSF-fit photometry using ZOGY outputs 15c and 15e, and compute additional quality metrics therefrom for thresholding. PSF-fit photometry is performed within 25 × 25 pixel cutouts on the difference-image (values p_i) using the unit-normalized PSF-

template from 15e with pixel values q_i. The maximum-likelihood estimate for the PSF-fit flux is given by:

$$F_{psf} = \frac{\sum_i p_i q_i}{\sum_i q_i^2},$$

with uncertainty:

$$\sigma_F = \frac{\sqrt{\sum_i q_i^2 \sigma_i^2}}{\sum_i q_i^2},$$

where σ_i^2 is the pixel variance from the difference uncertainty image (step 17). The PSF-fit flux (F_{psf}) is later converted to a magnitude using the difference (= science) image ZP from step 6. The associated reduced chi-square to determine goodness of fit or plausibility of the input pixel uncertainties is given by:

$$\chi^2 = \frac{1}{\nu} \sum_i \frac{\left(p_i - q_i F_{psf}\right)^2}{\sigma_i^2},$$

where ν is the number of degrees of freedom (= number of pixels in fit − 1). The above metrics are then used to further retain "good" candidates according to the following (final) two criteria:

$(0 < \chi^2 \leq chipsfthres)$ **and**

$|Mag_{(8\text{pix. diam. aperture})} - Mag_{psf}| \leq magdiffthres$.
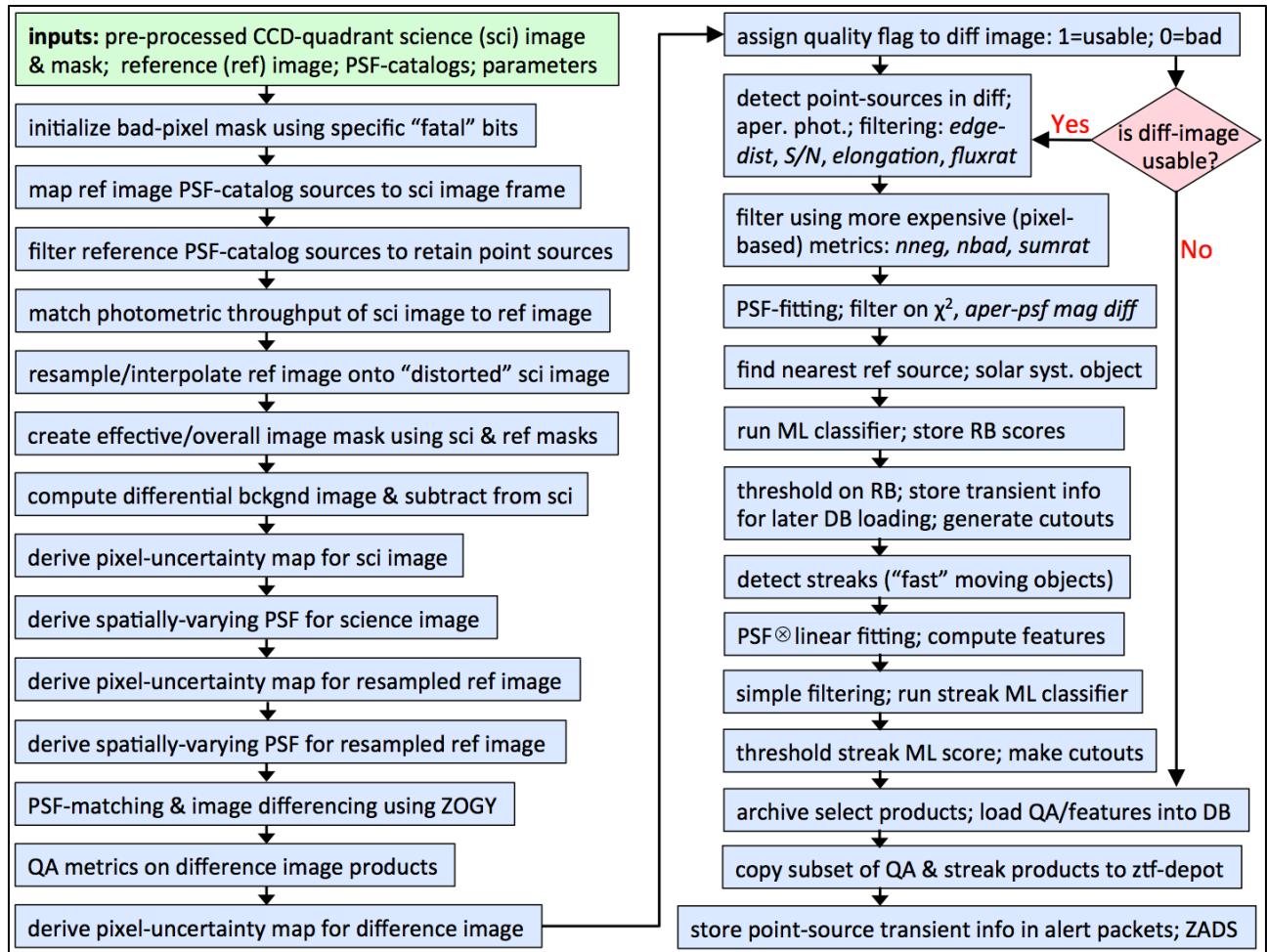
The reason for the above tiered filtering approach (i.e., 19b, 19c, 19d, and 19e) is that the metric computations in 19d and 19e are progressively more expensive in terms of runtime. It is important to eliminate the bulk of "bad" raw candidates (extracted in 19a) early on using simple metrics before proceeding. Otherwise, precious resources will be spent operating on large numbers of worthless candidates. This includes the candidate-association processing, machine-learned vetting, and image-cutout steps further below.

f. For transient candidates surviving criteria 19b-e, positionally cross-match to the input reference-image source catalog (input 1e) to find the nearest reference source falling within a radius *radrefmatch*. Associated reference-image catalog metrics are stored for each candidate.

g. If input was a positive (science − reference) difference image, find the nearest Solar-System object to each input transient candidate within a radius *ssradtol* and object (usually V-band) magnitude limit. A daily-updated MPC ephemeris file listing data for all known solar-system bodies is used with the *astcheck* utility. The associated magnitude, name, and distance to the nearest Solar-System object (if found within tolerance *ssradtol*) are stored.

h. For each transient candidate, we also positionally cross-match to the *Gaia* DR1 catalog to find the nearest *Gaia* source irrespective of magnitude, as well as the nearest *Gaia* source brighter than some magnitude limit (typically brighter than the ZTF saturation limit). Distances and Gaia *G*-band magnitudes are stored.

i. For each filtered point-source transient candidate, use all relevant *features* (metrics) assigned above to compute a *RealBogus* (RB) quality score using a machine-learned classification model. This model was trained on known *real* and *bogus* exemplars

following the filtering criteria (19b-e) and source-association metadata (19f-g) above. The RB score is a value between 0 and 1 where values closer to 1 imply a more reliable candidate. This score is thresholded using the *rbpntsrcthres* parameter. All candidates (with their features) satisfying this threshold are stored in a text file for later loading into the operations DB.

j. For each point-source transient satisfying the RB threshold in 19i, generate FITS-formatted cutouts from the relevant difference image (positive or negative) and corresponding regularized science and reference images used to generate these differences.

20. If difference image *status* = 1 (according to criteria in step 18), extract and filter *streak*-transient candidates from the positive (science – reference) difference image. These streaks are candidates for fast-moving objects, i.e., fast enough to create tracks in the individual exposures. These would be significantly extended than the PSF, along one axis. The streak processing entails:

a. Executing the *findstreaks* module to generate an initial list of streak candidates. This is based on finding linear patterns in the difference image. Masked regions due to saturation are first expanded to mitigate bright-star edge artifacts from contaminating the streak candidates. The method consists of finding connected pixels whose flux exceeds a specific threshold relative to the local noise-sigma computed in step 17. For more details, see Waszczak et al. (2017).

b. For each candidate streak from 20a, compute additional features and metrics (approximate start/end/mid-point positions, angles, and lengths) for input into the streak-PSF-fitting step below.

c. Execute the streak-PSF-fitting code on the list of raw candidates with metrics from 20b. This involves fitting a Gaussian representation for the difference image PSF convolved with a line. This is a non-linear fit that returns refined estimates for the streak endpoint positions, PSF-fitted and aperture fluxes, and uncertainties. Goodness-of-fit metrics and features are also returned for use in the machine-learned vetting step below.

d. For the list candidates with features from 20c, execute the machine-learned (ML) streak classifier using a pre-trained classification model. Prior to ML-classification, the output features from the streak-PSF-fitting in 20c are first filtered to remove "obvious" false-positives. The ML-classifier returns a probability that quantifies the relative reliability of each candidate streak.

e. The ML-predicted probabilities from 20d are thresholded to generate a final list of "best" candidate streaks. JPEG-formatted cutouts on each thresholded candidate are also generated from the difference image.

f. For each streak, generate metadata file in the XML ADES format for later submission to the MPC (following human vetting).

21. Following image differencing and extraction of *reliable* point-source events and moving-object streaks, selected file products from steps 15 and 16 (with filenames summarized in Section 7.2) are archived. Difference image QA metrics are stored in the *Subtractions* table,

point-source transient features in the *Candidates* table, and streak features in the *Streaks* table of the database (schema is given in Section 11.3).

22. As each point-source transient is stored in the DB (*Candidates* table), additional metadata from the PS1 DR1 catalog is assigned. This catalog, which includes pre-assigned star/galaxy classification scores, resides in the static *Stars* DB table. The ancillary *CandMatches* table is used to store the *Stars* table indices of the nearest PS1-sources falling within 1 arcsec of each record in the *Candidates* table.

23. Selected image-based QA metrics, including those from the science-image processing pipeline in Section 6.5 are written to text files and deposited on ZTF-Depot. Furthermore, all products related to ML-vetted moving-object streaks (extracted features, cutouts, and ADES-metadata files from 20e,f) are copied to ZTF-Depot in *realtime* (soon after generation). See Section 8 for details.

24. Following archiving and DB updates (steps 21 and 22), point-source transient (event) metadata from the DB, historical associations (if any), and cutouts are collected to generate alert packets in *Avro* format. These alerts are later fed into *Kafka*, to support the ZTF Alert Distribution System (ZADS). The contents of the alert packets are described in Section 9.

Figure 10: High-level processing flow in the image-subtraction and transient (point-source and streak) extraction pipeline. This represents the second phase of the real-time pipeline (the first was shown in Figure 4). For details, see Section 6.6.

6.7 Reference image (co-add) generation

Unless otherwise stated, all parameters and thresholds mentioned in this section are located in the following configuration file on GitHub:
https://github.com/Caltech-IPAC/ztf/blob/master/cdf/refimage.txt
Access to this repository is restricted to ZTF project team members and partners.

Figure 6 shows an overview of the processing flow.

1. Inputs are:
   a. A filtered list of instrumentally calibrated spatially-overlapping science images corresponding to a given filter, field, and readout-channel, all with astrometric and photometric calibration solutions in their FITS headers. These are from the pipeline described in Section 6.5. The criteria for the science image inputs before triggering the reference image pipeline are:
      > **(i)** Image quality falling in range $2.0 \leq \text{FWHM} \leq 3.5$ arcsec for any filter.
      > **(ii)** Overall quality *status* = 1 where the criteria used to set *status* = 1 (or equivalently, none of the bad INFOBITS) are defined in Section 10.4.
      > **(iii)** $25.5 \leq \text{MAGZP}(g) \leq 26.5$ or $25.5 \leq \text{MAGZP}(R) \leq 26.5$ or $25.35 \leq \text{MAGZP}(i) \leq 25.85$ for filters *g, R, i* respectively.
      > **(iv)** $-0.12 \leq \text{CLRCOEFF}(g) \leq 0.06$ or $0.04 \leq \text{CLRCOEFF}(R) \leq 0.135$ or $0.17 \leq \text{CLRCOEFF}(i) \leq 0.202$ for filters *g, R, i* respectively.
      > **(v)** $\text{MAGLIM}(g) \geq 19.6$ or $\text{MAGLIM}(R) \geq 19.6$ or $\text{MAGLIM}(i) \geq 19.2$ for filters *g, R, i* respectively.
      > **(vi)** Global pixel median (*gmedian*) ≤ 500 DN for any filter.
      > **(vii)** Global robust pixel RMS (*gpctdif*) ≤ 80 DN for any filter.
      > **(viii)** All science exposures acquired on or after UT night-date February 5, 2018. This is when the camera was reinstalled on the telescope.
      > **(ix)** A minimum of 15 overlapping science images satisfying (i) to (viii).
      > **(x)** Following criteria **(i)** to **(ix)**, the resulting science image list is sorted in order of *increasing* FWHM after which the first N_{max} images are retained. N_{max} therefore defines the desired depth. Currently, $N_{max} = 40$.

      As a further note, "survey-ready" reference images, i.e., that satisfy the above criteria started to be generated on 2018-03-24 at 09:00, Pacific Time. Therefore, any reference image created before this date that you may encounter in the archive is not survey quality. These older references *are not used* in the image subtraction or lightcurve production pipelines. These will eventually be replaced as more data is collected over the course of the survey.

      Automated reference image generation according to the above criteria continued for several months when it was realized that the sky-coverage in reference images was below expectations according to the fraction and depth of sky covered by individual exposures. This is because the above input criteria were too stringent. Analyses showed that some of the criteria could be relaxed without significantly affecting reference image quality. Therefore, starting at approximately 2018-09-19 07:00 PT

for *i*-filter and 2018-11-27 07:28 PT for the *g* and *R*-filters, the following new criteria were adopted in order to fill in more of the observable sky:

> **(i)** Image quality falling in range $1.7 \leq \text{FWHM} \leq 5.0$ arcsec for the *g* and *R* filters, and $1.7 \leq \text{FWHM} \leq 4.5$ arcsec for *i* filter.
>
> **(ii)** Overall quality *status* = 1 where the criteria used to set *status* = 1 (or equivalently, none of the bad INFOBITS) are defined in Section 10.4.
>
> **(iii)** $25.3 \leq \text{MAGZP}(g) \leq 26.5$ or $25.3 \leq \text{MAGZP}(R) \leq 26.5$ or $25.25 \leq \text{MAGZP}(i) \leq 25.85$ for filters *g, R, i* respectively. For *i*-filter only, this criterion was updated on 2020-04-27 to $25.21 \leq \text{MAGZP}(i) \leq 27.24$ to account for science programs with longer exposure times.
>
> **(iv)** $-0.20 \leq \text{CLRCOEFF}(g) \leq 0.15$ or $-0.05 \leq \text{CLRCOEFF}(R) \leq 0.22$ or $0.05 \leq \text{CLRCOEFF}(i) \leq 0.30$ for filters *g, R, i* respectively.
>
> **(v)** $\text{MAGLIM}(g) \geq 19.0$ or $\text{MAGLIM}(R) \geq 19.0$ or $\text{MAGLIM}(i) \geq 18.0$ for filters *g, R, i* respectively.
>
> **(vi)** Global pixel median: $gmedian(g) \leq 1900$ DN or $gmedian(R) \leq 1600$ DN or $gmedian(i) \leq 1200$ DN for filters *g, R, i* respectively. For *i*-filter only, this criterion was updated on 2020-04-27 to $gmedian(i) \leq 5000$ DN.
>
> **(vii)** Global robust pixel RMS: $gpctdif(g) \leq 100$ DN or $gpctdif(R) \leq 100$ DN or $gpctdif(i) \leq 80$ DN for filters *g, R, i* respectively.
>
> **(viii)** All science exposures acquired on or after UT night-date February 5, 2018. This is when the camera was reinstalled on the telescope.
>
> **(ix)** A minimum of 15 overlapping science images satisfying (i) to (viii).
>
> **(x)** Following criteria **(i)** to **(ix)**, the resulting science image list is sorted in order of *increasing* FWHM after which the first N_{max} images are retained. N_{max} therefore defines the desired depth. Currently, $N_{max} = 40$.

An important revision is to criterion **(i)** - the input FWHM cuts. The lower limit of 1.7 arcsec is now below the Nyquist limit of ~ 2.0 arcsec. The use of *some* slightly under-sampled input data (as indicated by the median derived image quality) does not significantly lead to artifacts when interpolating pixels onto the output co-add grid (see below) during co-addition using trimmed-averaging.

Commencing at approximately 2019-11-22 11:00 PT a "throughput flatness" criterion on the input science frames was included. This is defined such that the spatial distribution in photometric throughput in a science image using residuals with respect to the PS1 catalog binned over a 3 x 3 spatial grid has a maximum deviation (Δ_{max}) that is < 0.05 magnitudes. More specifically, $\Delta_{max} = \max\{\text{dMag}\} - \min\{\text{dMag}\}$, where $\{\text{dMag}\} = \text{median}(\text{PS1} - \text{ZTF}_{mag})$ in 9 (3 x 3) spatial bins using only sources with $13.5 \leq \text{ZTF}_{mag} \leq 18.5$. This flatness criterion significantly improved the quality of the resulting reference images moving forward. A regeneration of all survey reference images that include this criterion is TBD.

b. List of bit-mask images corresponding to the filtered list in 1a (also generated by the pipeline described in Section 6.5).

c. Configuration and processing parameters.
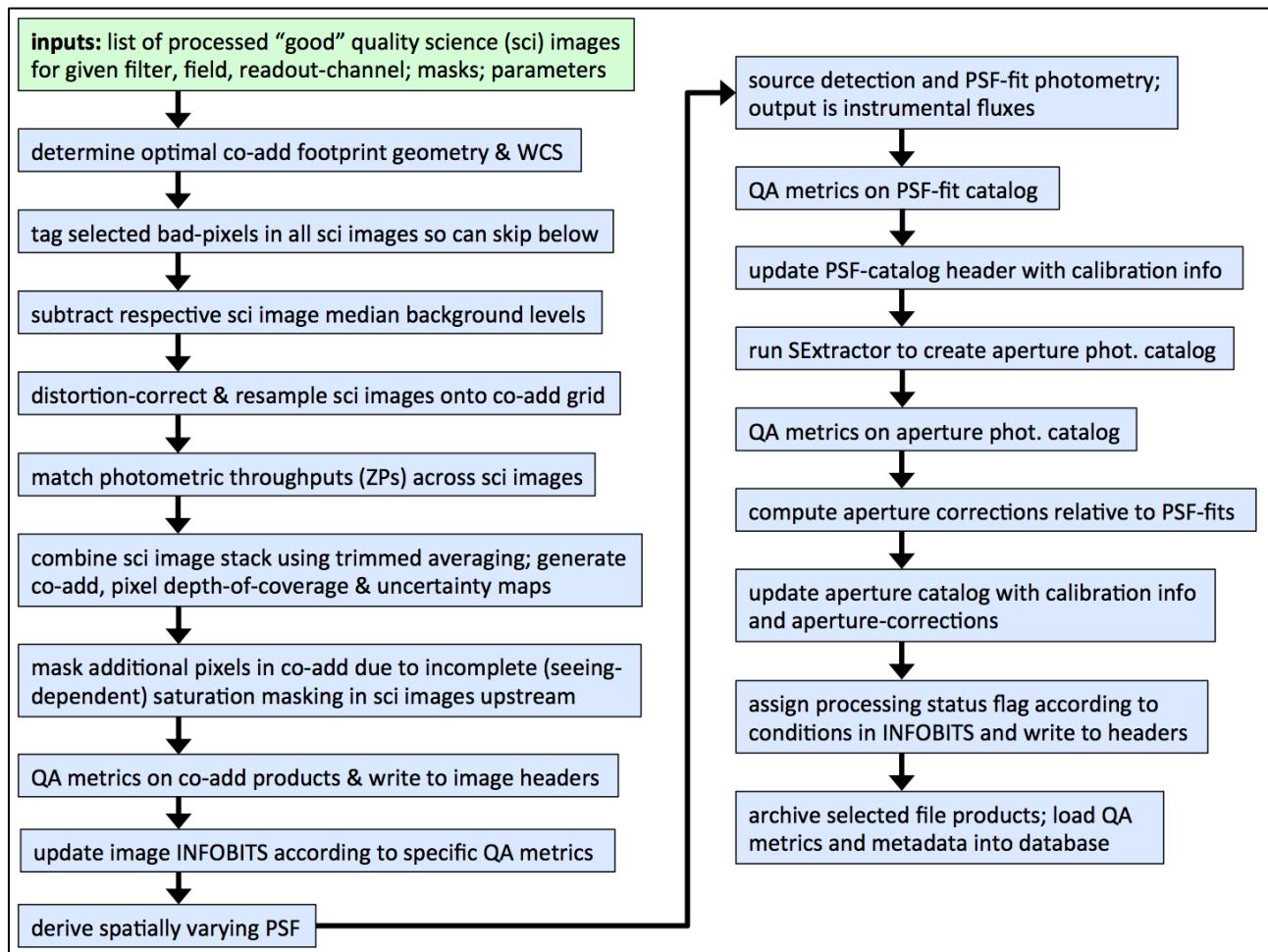
2. The WCS (pointing) geometry of the co-add footprint is optimally determined using WCS metadata from all the input image headers. The co-add pixel scale is currently fixed at 1.0 arcsec and co-add X,Y dimensions are fixed at 3200×3200 pixels. These dimensions are generous and can accommodate telescope pointing errors as well as small offsets (intentional dithering) relative to the predefined sky-tiling grid.

3. Bad and saturated pixels are internally set to NaN in each input image using their accompanying masks and a list of input fatal bits (parameter *fat*). This facilitates easier omission and tracking of bad pixels downstream.

4. Respective input image median levels are subtracted. This stabilizes (or homogenizes) the images against temporally-varying backgrounds before they are combined (below). These backgrounds are not always astrophysical, for example, there may be contamination from scattered moonlight or other bright objects whose line-of-sight may not directly fall on the image or the focal plane. The individual image background levels are stored for later use.

5. Each input image is then "de-warped" (distortion-corrected) and interpolated onto the output co-add grid using its astrometric and distortion solution. The pixel interpolation uses a 2D Lanczos kernel of window size three, as implemented in *SWarp* (Bertin et al. 2002). This kernel provides a relatively "compact support", minimizing aliasing and the spreading of bad pixels on output.

6. Since the input images would have been observed at different atmospheric transparencies, their photometric throughput (or effective photon-to-DN gain factors) will be different. Throughput-matching the images to a fixed photometric gain or zero point (*ZPref*) is therefore necessary before combining them. This is accomplished using the image photometric zero points (*ZPsci*) to first derive gain-match factors, i.e., $f_g = 10^{0.4(ZPref - ZPsci)}$, and then multiplying the input image pixels by their respective factors.

7. The pixel-masked, background-stabilized, resampled, and throughput-matched science images are combined using a trimmed-average per pixel; the outlier rejection is symmetric and the criterion for rejecting pixel p_i from a stack is $|p_i - median\{p_i\}| > 5\sigma$. The σ here is computed using half the '84.13 – 15.86' percentile difference, i.e., a measure that is relatively robust against outliers. The background levels computed in step 4 are then rescaled using the respective gain factors from step 6 and medianed. This constant serves as a fiducial background for the final co-add and is added to the trimmed-averaged image to create the final reference image product. Ancillary products include a pixel depth-of-coverage map, i.e., an image of the number of surviving samples per pixel following outlier rejection ($N_{trimmed}$), and a 1-σ pixel uncertainty image (= *pixel-stack StdDev* / $\sqrt{N_{trimmed}}$).

8. Due to temporal variations in seeing (i.e., across the input image stack), it's possible that saturated pixels on bright sources may not have been masked (in step 3) in some overlapping images where the seeing was "bad" while they were masked in images where the seeing was "good". This incomplete masking across the stack will lead to biased co-add intensity estimates on bright sources since any remaining (unmasked and non-outlier) pixels

are still averaged. To mitigate this effect, we additionally mask pixels in the reference image (with NaNs), if any corresponding pixels in the depth-of-coverage map have values (sample counts) $\leq fmsk \times N$, where *fmsk* is a parameter (currently 0.7) and *N* the number of input images.

9. Compute QA metrics on the image products generated in step 7; this includes computing the medians, minima and maxima of specific calibration metrics from the input-images. Also write selected metrics and metadata to header of reference image.

10. Initialize the reference image INFOBITS string and update according to specific thresholds for some of the metrics computed in step 9 (e.g., background level, noise-level, fraction of masked pixels). All reference image INFOBITS are defined in section 10.5.

11. Estimate spatially-varying PSF using an updated and automated version of *DAOPhot / AllStar*. A linearly-varying Gaussian basis is assumed with corrections written to a look-up-table. For details, see Masci et al. (2017). Image INFOBITS are also updated for cases where a PSF could not be derived.

12. Detect sources and perform PSF-fit photometry using an updated and automated version of *DAOPhot / AllStar*. Output catalog file contains calibrated photometry and astrometrically calibrated positions.

13. Compute QA metrics on PSF catalog from 12, including magnitude limits and median chi-square values from PSF-fitting. Image INFOBITS are also updated where pertinent.

14. Update header of PSF-fit catalog file from 12 with photometric calibration information and metrics. Convert to FITS binary table format. See Section 10.1.2 for details on how to apply photometric calibration solutions to the reference image catalog products.

15. Execute *SExtractor* to generate aperture photometry catalog containing instrumental photometry. Also compute QA metrics. Calibration information and relative PSF-to-aperture corrections are computed and added to the catalog header below.

16. Compute *SExtractor* aperture magnitude offset corrections relative to PSF-fit magnitudes using a filtered subset of the extractions from each catalog. Corrections are computed for all aperture sizes in the *SExtractor* catalog. For details on how to interpret and apply these corrections, see Section 10.2.

17. Write photometric calibration keywords and aperture correction information (step 16) to *SExtractor* catalog FITS catalog header.

18. Assign an overall processing quality status flag according to the INFOBITS string. This string is checked for the existence of specific "fatal" conditions (e.g., inability to perform astrometric and/or photometric calibration) as specified by *fatalinfobits*. This flag is written

as the STATUS keyword to the science image FITS header. For details on how this is set, see Section 10.5.

19. Write the INFOBITS definitions (Section 10.5) with their specific thresholds and other metrics to the reference image FITS header.

20. Selected reference image and catalog products (with filenames summarized in Section 7.4) are then archived. The product filenames along with all QA metrics are stored in the *RefImages*, *RefImAncilFiles*, *RefImCatalogs*, and *RefImQA* tables of the database (schema is given in Section 11.2).

Figure 11: High-level processing flow in the reference image (co-add) generation pipeline. For details, see Section 6.7.

6.8 Source-matching (lightcurve) generation

Lightcurves are constructed by cross-matching single-epoch science image extractions derived from PSF-fit photometry (catalog files from the pipeline described in Section 6.5). Lightcurves and associated statistics are stored in files called *matchfiles*, in HDF5 format, specifically a format used by the *pytables* package. Each matchfile contains lightcurves corresponding to a specific filter, field, CCD, and readout-channel combination. A description of the contents of a *matchfile* is given in Appendix C.

Matchfile creation (and updating of existing versions as data is accumulated) is performed on longer timescales than the real-time pipelines (Sections 6.5 & 6.6), and is contingent on having enough (new) data in a specific filter, field, and readout-channel. The source-matching pipeline is scheduled at times when the real-time pipelines are not running. An overview of the process is as follows:

1. Inputs:
    a. If a matchfile is to be created (for the first time), a PSF-fit source catalog corresponding to the reference image for the given filter, field, readout-channel is needed. This catalog is used to seed the creation of a *sources* table in the matchfile.
    b. If a pre-existing matchfile is to be updated, its path/filename is required.
    c. Optionally, single-epoch PSF-fit photometry catalog files from the pipeline described in Section 6.5, for the specific filter, field, readout-channel under consideration.

2. If a matchfile is to be created (for the first time), the following empty HDF5 tables are created in the matchfile: *exposures*, *sourcedata*, *transients*, and *transientdata*.

3. If single-epoch PSF-photometry catalogs are provided, a positional match to the coordinates in the *sources* table is performed. The default matching radius is 1.5 arcsec. For each catalog, a single entry is made in the *exposures* table.

4. Each source that matches a reference source is assigned the *matchid* from the *sources* table. Photometry for that source is entered in the *sourcedata* table.

5. Any input detection that is not matched is considered a transient. A position match is made against coordinates in the *transients* table. Transients receive their own *matchid* and their photometry is recorded in the *transientdata* table.

6. The *sourcedata* and *transientdata* tables are sorted by *matchid* to enable quick lookup of lightcurve data for a given source or transient.

7. Regardless of whether or not single-epoch PSF-photometry catalogs were supplied, "collapsed-lightcurve" statistics for sources and transients are computed, and stored in the *sources* and *transients* tables.

8. If specified, relative photometry corrections are computed based on a maximum of 400 "stable" (least-variable) sources in the given field, filter, readout-channel combination. These corrections are computed using the algorithm described in Appendix A of Ofek et al. (2011); see also Levitan (2013). Both corrected and uncorrected magnitudes are stored for each source. The delta-magnitude correction computed for each single-epoch exposure is stored in the *exposures* table.

9. Following creation and/or updating, matchfiles are copied to the archive. Only selected users on the project can access matchfiles in their entirety. Additionally, the archive ingestion process reads the *matchid*s, associated time-collapsed statistics, metadata, and position information for all unique objects in each matchfile and stores these in a database table. Lightcurves (the single-epoch photometry points) are not stored in this database. The object-based table interfaces with a GUI to enable spatial (cone or polygon) searches and range-queries on the pre-computed statistics. The returned *matchid*s (or *objectid*s) are then used to retrieve entire lightcurves from the associated matchfiles for packaging and delivery to the user. For details on how to access the lightcurve GUI, see Section 7.6.

6.8.1 Advisories and Caveats

Following the analysis of a large number of lightcurves, here are some cautionary notes.

- The relative photometric refinement step can to lead to erratic zeropoint corrections and hence unusable lightcurves. This is caused by occasional bad (unflagged) data going into the solver, which is difficult to track. As of April 8, 2019, we have decided to turn off relative photometric refinement when generating matchfiles.

- A large fraction of the records in the *transients* table (with photometry in the *transientsdata* table) appear to be associated with image artifacts, detections falling off the overlapping reference image footprint, and blended detections with ambiguous (non-unique) counterparts in the deeper, better quality reference images. To avoid too many spurious detections from dominating the sizes of matchfiles, the *transients* and *transientsdata* tables are no longer populated. The alert packets (§9 and §13.8) are the primary products from which transients should be mined.

- Photometric uncertainties imported from the raw PSF-fit photometry catalogs appear overestimated by factors of 1.5 to 2.5 relative to estimates of the RMS in photometric repeatability. These are stored as *psfmagerr* in the matchfiles (with accompanying photometry *psfmag*). The overestimation in *psfmagerr* is a function of magnitude. To mitigate this, we also provide values *magerr* (with accompanying *mag*) that are estimated from a robust fit to $y = 1.4826*medianabsdev$ versus $x = medianmag$. These two metrics are from the *Sources* table in a matchfile. The 1.4826 scaling reproduces the σ for Gaussian noise. The fit is piecewise and of the form:

$y = A + Bx + C10^{0.4x} + D10^{0.8x}$ for $x \leq 20.5$, and
$y = bx + c$ for $x > 20.5$.

For robustness, this fit is performed on 40th-percentile-binned values of $y = 1.4826*medianabsdev,$ primarily to catch the lower envelope of the distribution as a function of $x = medianmag$ (i.e., the non-variable/noisy population). The linear fit is used for extrapolating to fainter magnitudes to avoid a possible turnover and hence biases in the four-parameter fit. This linear fit is derived from the faintest two magnitude bins. This extrapolation provides more reliable *magerr* estimates at faint magnitudes. The coefficients *A, B, C, D*, and *b, c* are written to the *attributes* section of a matchfile. For a given lightcurve measurement $x = mag$, the 1-σ magnitude uncertainty *magerr* is predicted from the above fit and written to the *sourcedata* photometry table in the matchfile. This defaults to *psfmagerr* (the raw input catalog uncertainty) if issues were encountered in the fit. Currently, ~ 1.17% of the 125,502 matchfiles had fitting failures, primarily due to lack of enough good data in the input metrics following binning. For these matchfiles, *magerr* = *psfmagerr*.

A summary of the tables and metrics in the matchfile products are given in Appendix C (Section 18).

6.9   Moving Object Discovery Engine (ZMODE): point-source transient linking

The primary software used to construct candidate moving-object tracks is called **ZMODE**: ZTF's Moving Object Discovery Engine. Unless otherwise stated, all parameters and thresholds mentioned in this section are located in the following configuration files on GitHub. Access to this repository is restricted to ZTF project team members and partners.

- "High-level" processing parameters used when running the overall ZMODE executive script (e.g., initial DB query, input transient point-source quality filters etc): *https://github.com/Caltech-IPAC/ztf/blob/master/cdf/zmodeexec.txt*

- Parameters used when executing the primary track-construction module (ZMODE): *https://github.com/Caltech-IPAC/ztf/blob/master/cdf/zmode.txt*

- Parameters used when generating provisional MPC report, cross-matches to known (MPC-archived) objects, image-cutouts on new candidate objects, and *digest2* report: *https://github.com/Caltech-IPAC/ztf/blob/master/cdf/makeMPCreport.txt*

Below are the steps performed by the ZMODE executive script. Execution typically occurs at the end of each observing night, using point-source transient candidate detections spanning *at most* the last four *continuous* observing nights if available. If four continuous nights are not available, detections from the recent three or two night(s) are used as input. A list of the final product files from each processing step is given below.

1. Primary inputs to the executive are a comma-separated list of "*nightdate*" strings specifying the nights from which the difference-image point source transient candidates should be used; a configuration file listing thresholds for the transient detections (*RealBogus* score,

53

S/N, and magnitude limit), as well as quality thresholds for the orbit-fit metrics from the *find_orb* software. The output directory for all ZMODE products is also specified.

2. Execute SQL query to retrieve point-source transients from the *candidates* DB table using specific thresholds (schema is in Section 11.3). At the time of writing, the SQL query is as follows. Input thresholds/parameters are italicized.

```
select c.nid,i.nightdate,s.jd,c.pid,c.candid,c.ra,c.dec,
       c.sigmara,c.sigmadec,c.covradec,c.distnr,c.magnr,c.magpsf,
       (1.085736/c.sigmapsf) as snr,s.fid,c.ssnamenr,c.ssdistnr
       from candidates c, subtractions s, immeta i
       where c.pid=s.pid and s.pid=i.pid and c.isdiffpos='t' and
       s.ncandscimreffilt <= ncandfiltthres and
       c.rb >= rbthres and
       (1.085736/c.sigmapsf) >= snrthres and
       c.magpsf <= magthres and
       c.nbad <= nbadthres and
       c.fwhm <= fwhmthres and
       c.elong <= elongthres and
       c.magdiff >= magdiffthreslo and
       c.magdiff <= magdiffthreshi and
       not (c.distnr <= distnrthres and c.magnr <= magnrthres) and
       i.nightdate in (nightdates)
       order by c.nid,i.nightdate,s.jd limit limitnum
```

3. Execute the ZMODE software on the transient detection list from step 2 to construct candidate moving-object tracks. A description of the ZMODE algorithm is given in Appendix A.

4. Perform orbit fitting on each candidate ZMODE track found in step 3. The *find_orb* software is used. The output metrics from find_orb are then thresholded to construct a list of "good" likely real moving-objects, i.e., worthy for downstream human scanning. An orbital elements file is also generated. A description of this process is also given in Appendix A.

5. Following *find_orb* and thresholding, generate:
    (i)   A *preliminary* (or *provisional*) MPC report in the **old** 80-character format for downstream human scanning that contains all previously known objects and *new* potential objects not found from cross-matching to the MPC archive;
    (ii) Same as (i), but in the **new** XML ADES format.
    (iii) A *digest2* report containing scores for probable orbit classes; and
    (iv) Science (epochal) image and reference (co-add) and image cutouts centered on every transient detection that was used to construct each *new* track (i.e., not associated with a known object).

6. Select file products are copied from the operations processing directory to ztf-depot (directory structure given in Section 8.1). Example files currently written to ztf-depot are below. Beside each one […] is its origin from the specific processing step above. The *run*<ID> in each file product refers to the most recent nightly run.

```
/ztf/depot/20160830/MOPSprods/zmodeinpdetections_run001.csv  [output from step 2]
/ztf/depot/20160830/MOPSprods/zmodetracks_run001.tbl    [raw tracks from step 3]
/ztf/depot/20160830/MOPSprods/zmodelog_run001.txt           [log from step 3]
/ztf/depot/20160830/MOPSprods/zmodetracks_run001_obsinfo.txt    [from step 4]
/ztf/depot/20160830/MOPSprods/zmodetracks_run001_good.tbl       [from step 4]
/ztf/depot/20160830/MOPSprods/zmodetracks_run001_elements.txt     [from step 4]
/ztf/depot/20160830/MOPSprods/zmodefolog_run001.txt          [log from step 4]
/ztf/depot/20160830/MOPSprods/zmodempcreplog_run001.txt          [from step5]
/ztf/depot/20160830/MOPSprods/zmodetracks_run001_digest2.txt        [from step 5]
/ztf/depot/20160830/MOPSprods/zmodetracks_run001_mpcreport.txt     [from step 5]
/ztf/depot/20160830/MOPSprods/zmodetracks_run001_mpcreport.xml     [from step 5]
/ztf/depot/20160830/MOPSprods/zmodetracks_run001_stamps/    [stamp image cutouts]
```

6.10 Calibration image generation, application to science data & association in archive

An automated process (executed by the Virtual Pipeline Operator or VPO) manages the creation of calibration image products and their application to the science image data. The bias and high-frequency flat calibration generation pipelines were outlined in Sections 6.3 and 6.4 respectively. Products therefrom are then used to process the science image data (Section 6.5). Below we review the daily orchestration of these processes.

The VPO *starts* to query the pipeline operations database at T_{start} = 4PM (PST – Autumn/Winter) or 5PM (PDT – Spring/Summer) for any *new* exposures received with IMGTYPE = "flat" and "bias". If $N \geq 10$ exposures of any of these types already exist at T_{start} with observation timestamps going back 12 hours, i.e., back to 4AM (PST) or 5AM (PDT) the same day, the pipelines are triggered to create a stacked bias and/or flat calibration product. These products are archived and registered in the *CalFiles* table of the database (Section 11.2). These calibration products are automatically used to process the forthcoming night's science data (if any). The real-time science pipeline always searches for the most recent calibration product in the archive, i.e., nearest in time going back from the observation time of a science exposure.

If however the database returns $N < 10$ bias and/or flat exposures at T_{start}, the VPO keeps querying the database throughout the night for additional bias and/or flat exposures until $N \geq 10$ is satisfied. Once a new bias and/or flat calibration is made for that night, the database querying for bias and/or flat IMGTYPEs stops, otherwise the querying stops automatically at 4AM (PST) or 5AM (PDT) the next day. Any calibration exposures acquired after these times (following an observing night) will be considered for calibration generation the next day, where the cycle repeats.

Related to the above calibration generation and archiving is the association of calibration products with the archived science data (see archive layout in Section 7). The date-time stamps in the archive paths and filenames defined in Section 7 are all **UT based**. That is, the definition of a day/night cycle for all ZTF Data System operations is:
```
00:00:00.000 to 23:59:59.999 (UT)
```

This is equivalent to the following local (Pacific) times, depending on the season:
```
16:00:00 to 04:00:00 (PST — Autumn/Winter)
17:00:00 to 05:00:00 (PDT — Spring/Summer)
```

During survey operations, raw calibration data (and products derived therefrom) will be acquired/generated at the start of each night, i.e., prior to on-sky observing. As a detail, the date-timestamp used to construct calibration product filenames are forced to that of the upcoming observing night. This facilitates unambiguous association of calibration and science data in the archive based on a single UT date, i.e., defining each observing night.

7 SUMMARY OF FILE PRODUCTS IN ARCHIVE

Table 1 gave a summary of the data access details and availability timescales. The archive directory structure is as follows. Identifiers in the paths are defined below.

```
/ztf/archive
   | | | | |
   | | | | |___/raw/<YYYY>/<MMDD>/<dddddd>/
   | | | |         ztf_<YYYYMMDDdddddd>_<fieldID>_<filterID>_c<ccdID>_<imgtype>.fits.fz
   | | | |
   | | | |
   | | | |_____/cal/<YYYY>/<MMDD>/<caltype>/<filterID>/ccd<ccdID>/q<quadID>/
   | | |           ztf_<YYYYMMDD>_<filterID>_c<ccdID>_q<quadID>_<ptype>
   | | |
   | | |
   | | |_____/sci/<YYYY>/<MMDD>/<dddddd>/
   | |             ztf_<YYYYMMDDdddddd>_<fieldID>_<filterID>_c<ccdID>_<imgtype>_q<quadID>_<ptype>
   | |
   | |
   | |_____/ref/<fff>/field<fieldID>/<filterID>/ccd<ccdID>/q<quadID>/
   |               ztf_<fieldID>_<filterID>_c<ccdID>_q<quadID>_<ptype>
   |
   |
   |_____/mtc/<fff>/field<fieldID>/<filterID>/ccd<ccdID>/q<quadID>/
                   ztf_<fieldID>_<filterID>_c<ccdID>_q<quadID>_<ptype>


where:
 /raw - raw image data files
 /cal - calibration product files
 /sci - epochal science product files and difference images
 /ref - reference images (co-adds) and catalog files
 /mtc - source matchfiles (lightcurve files) and logs


 <fff> = first three (leftmost) digits of <fieldID> (see below)
 <caltype> = calibration type, e.g., "bias", "hifreqflat"
 <imgtype> = single character label in raw camera image file (see below)
 <ptype> = product type suffix string from pipeline (see product list below)
```

7.1 Raw camera image files

One *fpack*'ed (compressed) multi-extension FITS-formatted file per CCD.

Generically:
ztf_<YYYYMMDDdddddd>_<fieldID>_<filterID>_c<ccdID>_<imgtype>.fits.fz

where:
  - YYYY = year
  - MMDD = month and day (all UT based)
  - dddddd = fractional time of day of exposure (all UT based)
  - fieldID = 6-digit survey field ID if targeted science (on-sky) exposure, otherwise "000000"
          for calibrations (see imgtype below).
  - filterID = 2-letter filter code:
          = **zg, zr, zi** for exposure acquired in *g, R,* or *i* respectively (for on-sky & flat-fields)
          = **bi, dk** for bias and dark images respectively (filter neutral)
  - ccdID = 2-digit detector chip ID: 01, 02, … 16 (see Figure 1 for ordering)
  - imgtype = o: on-sky object observation or science exposure
          = b: bias calibration image
          = d: dark calibration image
          = f: dome/screen flatfield calibration image
          = c: focus image
          = g: guider image

Examples:
  ztf_20171021128889_001614_zg_c03_o.fits.fz (*g*-filter raw on-sky science image CCD file)
  ztf_20171022116732_000000_zr_c16_f.fits.fz   (*r*-filter raw flat-field image CCD file)
  ztf_20171112116739_000000_bi_c02_b.fits.fz (raw *bias* image CCD file)
  ztf_20171119316841_000000_dk_c02_d.fits.fz (raw *dark* image CCD file)

7.2 Science (single exposure-epoch based) product files

Files are CCD-quadrant (readout-amplifier) based.

Generically:
ztf_<YYYYMMDDddddd>_<fieldID>_<filterID>_c<ccdID>_<imgtype>_q<quadID>_<ptype>

where:
  - first five quantities in angled brackets are the same as in the raw-file root name (see §7.1)
  - quadID = 1-digit quadrant ID in ccd: 1, 2, 3, or 4 (see Figure 1 for ordering)
  - ptype = sciimg.fits: primary science image (see header for "info-bit" QA summary)
          = mskimg.fits: bit-mask image (see header for bit-definitions)
          = psfcat.fits: PSF-fit photometry catalog from DAOPhot
          = sexcat.fits: nested-aperture photometry catalog from SExtractor
          = sciimgdao.psf: spatially varying PSF estimate in DAOPhot's lookup table format

= sciimgdaopsfcent.fits: PSF estimate at science image center as a FITS image
= sciimlog.txt: log output from instrumental calibration pipeline, that created sciimg.fits
= scimrefdiffimg.fits.fz: difference image (science *minus* reference); *fpack*-compressed
= diffimgpsf.fits: PSF estimate for difference image as a FITS image
= alerts.tar.gz: tar-gzipped directory containing alert packet files in *Avro* format for all difference-image-detected events; see §9 for details
= diffimlog.txt: log output from image subtraction and extraction pipeline
= log.txt: overall system summary log from realtime pipeline

7.3   Calibration image product files

Files are CCD-quadrant (readout-amplifier) based.

Generically:
ztf_<YYYYMMDD>_<filterID>_c<ccdID>_q<quadID>_<ptype>

where:
  - first three quantities in angled brackets are the same as in the raw-file root name (see §7.1)
  - quadID = 1-digit quadrant ID in ccd: 1, 2, 3, or 4 (see Figure 1 for ordering)
  - ptype = bias.fits: bias calibration image product
        = biasunc.fits: 1-σ uncertainty image for bias image using trimmed stack StdDev/$\sqrt{N}$
        = biascmask.fits: bad pixel calibration mask tagging bad detector / noisy pixels
        = biaslog.txt: log output from bias calibration pipeline
        = hifreqflat.fits: high-frequency relative pixel-to-pixel responsivity image product
        = hifreqflatunc.fits: 1-σ unc. image for hifreqflat using trimmed stack StdDev/$\sqrt{N}$
        = hifreqflatlog.txt: log output from high-frequency calibration pipeline

7.4   Reference image (co-add) product files

Files are CCD-quadrant (readout-amplifier) based.

Generically:
ztf_<fieldID>_<filterID>_c<ccdID>_q<quadID>_<ptype>

where:
  - first three quantities in angled brackets are the same as in the raw-file root name (see §7.1)
  - quadID = 1-digit quadrant ID in ccd: 1, 2, 3, or 4 (see Figure 1 for ordering)
  - ptype = refimg.fits: primary reference image product
        = refunc.fits: 1-σ uncertainty image for reference image product
        = refcov.fits: depth-of-coverage map following stack-outlier rejection
        = refpsfcat.fits: PSF-fit photometry catalog from DAOPhot
        = refsexcat.fits: Nested-aperture photometry catalog from SExtractor
        = refimlog.txt: log output from reference image generation pipeline
        = log.txt: overall system summary log from reference-image pipeline

7.5   Source *Matchfiles* (Lightcurve files)

Files are CCD-quadrant (readout-amplifier) based.

Generically:
ztf_<fieldID>_<filterID>_c<ccdID>_q<quadID>_<ptype>

where:
  - first three quantities in angled brackets are the same as in the raw-file root name (see §7.1)
  - quadID = 1-digit quadrant ID in ccd: 1, 2, 3, or 4 (see Figure 1 for ordering)
  - ptype = match.pytable: primary *matchfile* product
          = match.pytable_log.txt: log output from source-matching/relative photometry pipeline

7.6   Archive Access Methods

- All epochal science image products, calibration products, raw image data files, and reference-image products can be accessed via links on the main IRSA-ZTF page:
  https://irsa.ipac.caltech.edu/Missions/ztf.html
  See below for links to specific services with further documentation.

- The archive file-system tree shown in Section 7 can be browsed directly at:
  https://irsa.ipac.caltech.edu/ibe/data/ztf/

- Queries using the Application Programming Interface (API), including image-cutouts are described in:
  https://irsa.ipac.caltech.edu/docs/program_interface/ztf_api.html

- A Graphical User Interface (GUI) for interactively querying/viewing image and catalog files with source overlays, and generating cutouts can be accessed at:
  https://irsa.ipac.caltech.edu/applications/ztf/

- All image metadata metrics against which the archive products can be queried (through either the API or GUI) are described in:
  https://irsa.ipac.caltech.edu/docs/program_interface/ztf_metadata.html

- The Moving Object Search Tool (MOST) GUI can be accessed at:
  http://irsa.ipac.caltech.edu/applications/MOST/

- The API equivalent of the MOST GUI is documented at:
  http://irsa.ipac.caltech.edu/applications/MOST/MOSTProgramInterface.html

- A query-able database of ZTF-detected Objects that can be used as input to the lightcurve retrieval GUI and Time-Series Analysis Tool (see next bullet) is at:
  https://irsa.ipac.caltech.edu/cgi-bin/Gator/nph-scan?mission=irsa&submit=Select&projshort=ZTF

- Following a query for ZTF Objects (see previous bullet), one can retrieve their lightcurves by clicking on the following icon at the bottom of the Objects results page:

  When presented with a list of *Object IDs*, you can click on any of them to spawn the interactive Time-Series Analysis Tool in your browser. This tool includes a visualization of the epoch-based science images centered on the object as well as a period finder.

- Entire *source-matchfile* products (in pytable format) containing all the lightcurves for a given field, CCD, quadrant and filter, and from which lightcurves are constructed for the lightcurve/time-series tool. These files are available to specific project-team members only. Access to *matchfiles* containing epochal data from **all *program IDs***: *https://ztfweb.ipac.caltech.edu/ztf/ops/srcmatch/* Access to *matchfiles* containing data from **only partnership** observations (*programID*=2): *https://ztfweb.ipac.caltech.edu/ztf/ops/srcmatch_programid2/*

8 SUMMARY OF PRODUCTS IN ZTF-DEPOT

ZTF-Depot is a project-internal file-based retrieval portal accessed through a web-server to facilitate two purposes:

1. Allow specific internal-project members to access QA information and survey performance metrics in near-realtime for monitoring and long-term trending. This includes cumulative and incremental (nightly) survey-coverage maps.

2. Serve products to support near-realtime Solar System discovery science, also restricted to specific individuals in the Solar System Working group (primarily through the ZTF Solar System Marshal). These products comprise:
   a. single-epoch image streaks with cutouts and metadata. See pipeline description in Section 6.6.
   b. moving-object tracks constructed from linking difference-image-detected *point-source* transients (not streaks). Cutouts, metadata, and orbit-fit information are also provided. See the pipeline description in Section 6.9.

8.1 ZTF-Depot Directory Structure

Figure 12 shows a layout of the ZTF-Depot products. Directories and filenames with angled brackets denote variable quantities.

adhoc/ – directory for storing additional non-realtime-critical products requested by users, i.e., to support external analyses. Such products may not be available in the archive.

readme_depot.html – definitions of all QA metrics listed in files under <YYYYMMDD>/.

<YYYYMMDD>/ – date of night being processed: based on UT. At most, two date directories will be retained (previous and current). Intent is to enable recovery from server and/or client-side hardware outages.

MOPSprods/ – directory for products from moving-object pipeline system (object tracks from linking point-source transient candidates identified by the ZMODE software).

CovMaps/ – all-sky exposure depth-of-coverage maps following *end-of-night* processing for (i) only that night that was processed; (ii) cumulative over the course of the survey up to and including that night; and (iii) separately for data from the public (MSIP) survey. These are also generated for separate filters in *equ* and *gal* coordinate systems in a Hammer-Aitoff projection. Format is both FITS and PNG.

CalQA/ – directory containing QA metadata files for image-based calibration files, i.e., for bias and dome-screen flat frames after splitting into quadrants and actual calibration (ensemble) products made from products acquired on <YYYYMMDD>/.

AggregateStats.txt – file summarizing number of images observed versus processed etc. as available at query time (See Section 6.1).

<ObsTimeRange>/ – directory name for realtime *science products* corresponding to query interval; one choice is the UT-based exposure-observation time range falling in <YYYYMMDD>, rounded to nearest second. E.g., 21h33m50s-21h43m50s/, 21h43m51s-21h53m53s/ …

ccd<ID>/ – subdirectory containing *discovery-based* products for a specific CCD ID [01..16]. A breakdown by CCD ID will keep the number of product files per directory at a manageable level.

<rootfname> – unique root filename of chip-quadrant (processing unit in pipeline), e.g., *ztf_20160320398531_000630_zr_c01_o_q2*, where fields are exposure date-time string; field ID [~ 000001 – 002000]; filter name; CCD ID [01..16]; quadrant ID therein [1..4].

<rootfname>_qa.txt – chip-quadrant-image based QA metrics: raw images, calibration products (under CalQA/ above), instrumentally calibrated science and subtraction images (under specific <ObsTimeRange>/).

<rootfname>_streaks.txt – IDs and metrics for streak-transient candidates.

strkid<ID>_pid<ID>_scimref.jpg – JPEG image-cutout for streak candidate with <ID> and other information stored in <rootname>_streaks.txt. All streak cutouts are 65×65 pix^2 in size.

strkid<ID>_pid<ID>_ades.xml – XML-formatted streak-metadata file in the ADES format to support deliveries to the MPC.
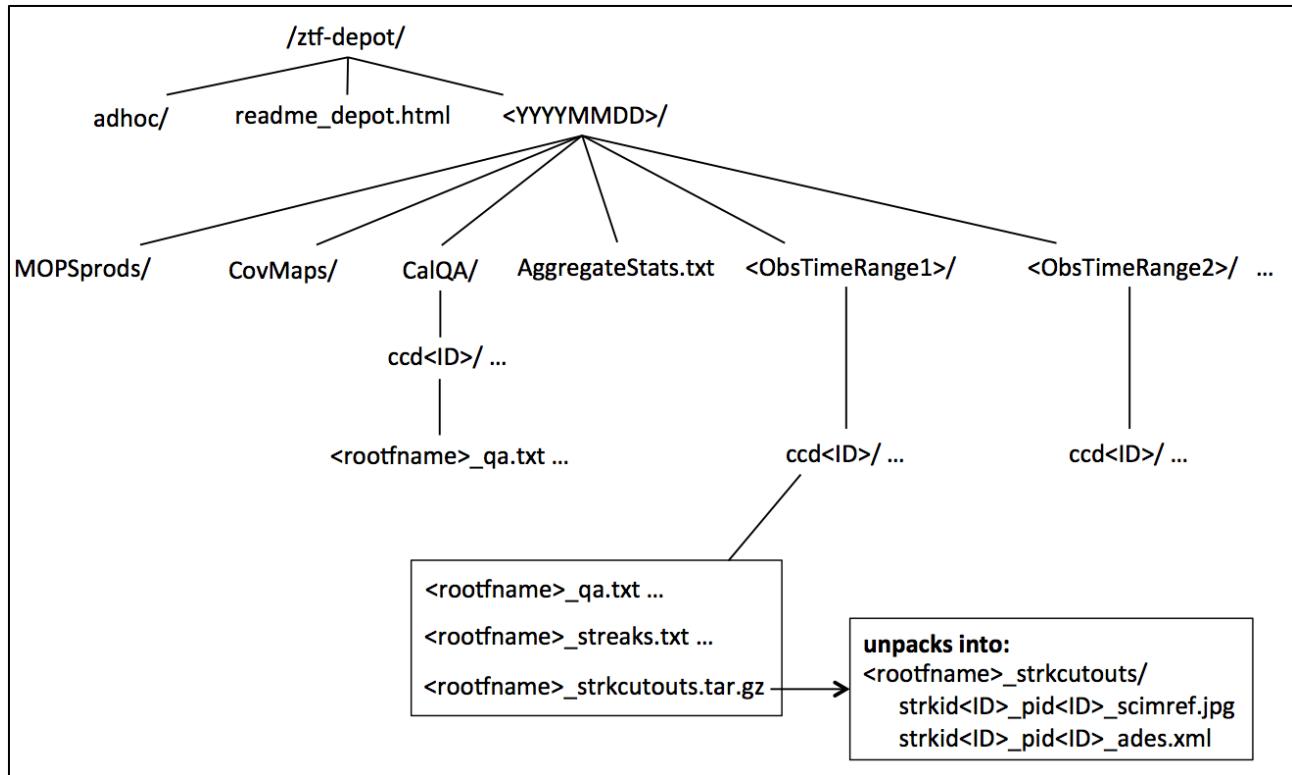
Figure 12: directory layout and files in ztf-depot (see Section 8.1 for definitions).

8.2 ZTF-Depot Access

The ztf-depot file-system can be accessed using the following URL:
https://ztfweb.ipac.caltech.edu/ztf/depot/
Access is restricted to internal project-team members.

9 ALERT (EVENT) STREAM

9.1 Triggering Criteria for Alerts

Alert packets are binary files in the *Avro*-format. Each packet corresponds to a single target event extracted from either a *positive* (science minus reference) or *negative* (reference minus science) difference image. Packets are generated from raw events that satisfy the following criteria. The intent is to mitigate obvious image artifacts and residuals from bad subtractions.

- PSF-fit flux S/N ≥ *thres_snr* (currently = 5.0 in all filters)

  Note: the S/N here is internally computed from "$1.0857/sigmapsf$", where the numerator is $2.5/log_e(10)$ and *sigmapsf* is the 1-σ uncertainty in the PSF-fit magnitude. *sigmapsf* is included in the alert packets.

- PSF-fit magnitude: *magpsf* ≤ *thres_magpsf* (currently = 23.5 mag in all filters)

- Number of bad pixels in a centered 5×5 pixel region: *nbad* ≤ *thres_nbad* (currently = 4 pixels)

- FWHM of source profile: *fwhm* ≤ *thres_fwhm* (currently = 7 pixels)

- Source elongation (ratio A/B from elliptical isophot): *elong* ≤ *thres_elong* (currently = 1.6)

- Fixed aperture – PSF-fit magnitude difference (*magap* – *magpsf*): *thres_magdiff_lo* ≤ *magdiff* ≤ *thres_magdiff_hi* (currently -0.4 ≤ *magdiff* ≤ 0.75)

- Random Forest Real-Bogus quality score: *rb* ≥ *thres_rb* (currently = 0)
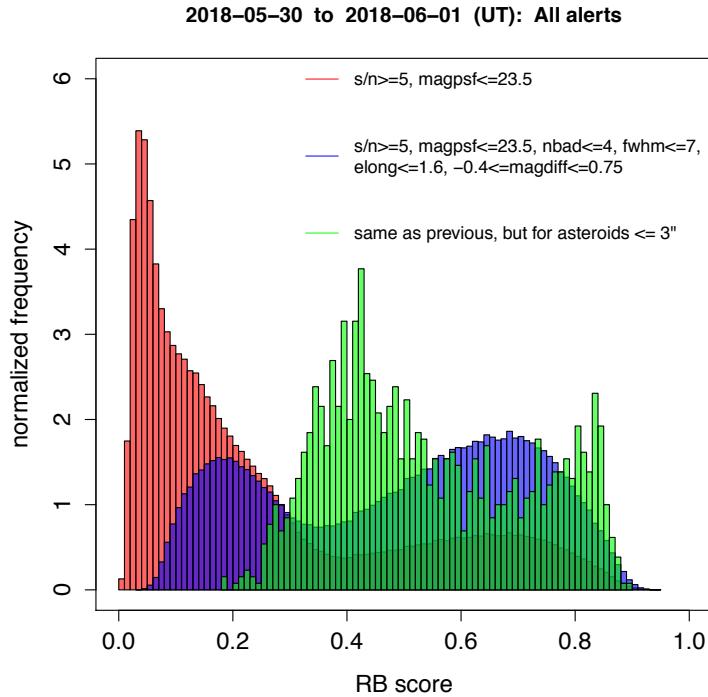
All alert packets leaving IPAC have already been filtered using the above criteria. Figure 9.1 compares the distribution of the Random Forest *rb* scores of events with and without these cuts, including those positionally associated with known asteroids. These events are from three nights of data that used the same *rb* machine-learned training model.
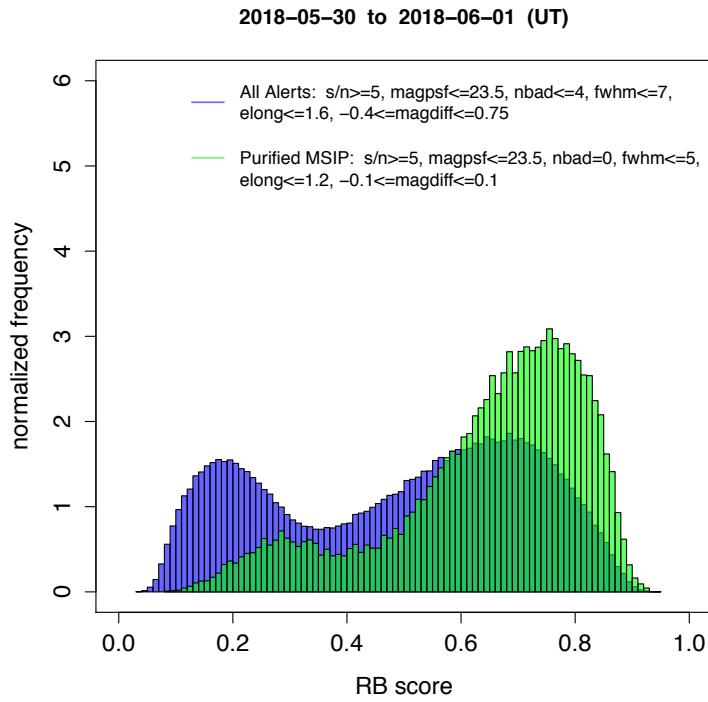
If one is interested in filtering the alert stream even further to improve reliability or purity (at the expense of completeness), we suggest the following filtering on the metadata in each packet:

- Random Forest Real-Bogus quality score: *rb* ≥ 0.65
- Number of bad pixels: *nbad* = 0
- FWHM of source profile: *fwhm* ≤ 5 pixels
- Source elongation: *elong* ≤ 1.2
- Aperture – PSF-fit magnitude difference (absolute value): | *magdiff* | ≤ 0.1 mag

Figure 9.2 compares the distribution of the Random Forest *rb* score for events extracted from the MSIP (public) alert stream (see §9.3) using these purity cuts (without the *rb* cut) to those from the full alert stream.

Figure 9.1: Distributions of RB score for alerts generated from raw events (red); with filtering applied to weed out obvious false positives (blue); and from this, those associated with known asteroids (green). The number of alerts in each are: 643860 (red); 228287 (blue); 1300 (green).

2018–05–30 to 2018–06–01 (UT)

All Alerts: s/n>=5, magpsf<=23.5, nbad<=4, fwhm<=7, elong<=1.6, −0.4<=magdiff<=0.75

Purified MSIP: s/n>=5, magpsf<=23.5, nbad=0, fwhm<=5, elong<=1.2, −0.1<=magdiff<=0.1

Figure 9.2: Distribution of RB for "purified" public alerts (green) versus all alerts (blue). The number of alerts in each are: 24581 (green) and 228287 (blue; same as Figure 9.1).

9.2 File-naming and Contents of Alert Packets

Alert packets have filenames formatted as *<candidateID>*.avro, where *candidateID* is a unique, internally-generated identifier from the *candidates* database table. This is the *candid* column therein (see schema in Section 11.3). The *candid* is also included in the alert packet.

Each alert packet is structured using a JSON-based schema, with metadata organized as follows:

- Target alert name (*objectId*): see Section 9.4 for naming mechanism. Note: the same *objectId* can correspond to multiple *candidateID*s (separate alert packets).

- Event source-specific metrics ("features") and difference-image metrics/metadata from which it was extracted (see listing below). Most of these are from internal database queries (specifically the *candidates* and *subtractions* tables with schemas given in Section 11.3).

- Metadata for the closest known Solar System object within a specific radius if found.

- Metadata for the closest, second closest, and third closest PS1 source within a specific radius if found, along with ML-predicted star-galaxy classification scores where available.

67

- Metadata for the closest source from the *Gaia* DR1 catalog, including the brightest source below some magnitude limit (~ ZTF saturation limit) within a specific radius. If found, distances and *Gaia G*-band magnitudes are reported.

- Previous (historical) events falling on/near target alert position (if any) with source- and image-based metrics for each, regardless of filter (*g*, *r*, *i*) and regardless if the event was detected in a positive (*science minus reference*) or negative (*reference minus science*) difference image. The matching radius is 1.5 arcsec and search span (look-back time) is currently 30 days. These historical events have relaxed extraction criteria than those used to trigger the initial alert (see above). Currently, the difference-image extraction threshold for any historically associated event is S/N = 5. See Section 9.3 for other restrictions.

- If there are no previously-detected events on/near the alert position (previous bullet), flux upper-limits are provided for all epochal images that covered (touched) that position, also going back 30 days. These upper limits are not local estimates (e.g., from forced photometry), but global estimates from the magnitude limit *per difference-image*.

- Three 63 × 63 pixel image cutouts (native pixel size) centered on the alert position from the science, reference, and difference images. These cutouts are in lossless *.fits.gz* format and the pixel values therein retain the full 32-bit floating point precision inherited from the input images.

The metadata in an alert packet are arranged in JSON-based schemas. For those with access, these are on GitHub:
https://github.com/Caltech-IPAC/ztf/tree/master/src/pl/avroalerts/schema
Access to this repository is restricted to ZTF project team members and partners.

For reference, below is the primary metadata associated with the event that triggered an alert, i.e., the *candidate* portion of a packet. This is followed by a definition of the metadata from any historically-associated events, including non-detections, i.e., the *prv_candidates* portion of a packet. The *prv_candidates* metadata comprises a subset of the primary *candidate* metadata.

candidate schema in packet:

```
metric                 description
jd              Observation Julian date at start of exposure [days]
fid             Filter ID (1=g; 2=R; 3=i)
pid             Processing ID for science image to facilitate archive retrieval
diffmaglim      Expected 5-sigma mag limit in difference image based on global noise
                estimate [mag]
pdiffimfilename filename of positive (sci minus ref) difference image
programpi        Principal investigator attached to program ID
programid        Program ID: encodes either public collab or caltech mode
candid          Candidate ID from operations DB
isdiffpos       t or 1 => candidate is from positive (sci minus ref) subtraction;
                f or 0 => candidate is from negative (ref minus sci) subtraction
tblid           Internal pipeline table extraction ID
nid             Night ID
rcid            Readout channel ID [00 .. 63]
field           ZTF field ID
xpos            x-image position of candidate [pixels]
ypos            y-image position of candidate [pixels]
```

68

```
ra              Right Ascension of candidate; J2000 [deg]
dec             Declination of candidate; J2000 [deg]
magpsf          Magnitude from PSF-fit photometry [mag]
sigmapsf        1-sigma uncertainty in magpsf [mag]
chipsf          Reduced chi-square for PSF-fit
magap           Aperture mag using 14 pixel diameter aperture [mag]
sigmagap        1-sigma uncertainty in magap [mag]
distnr          distance to nearest source in reference image PSF-catalog [pixels]
magnr           magnitude of nearest source in reference image PSF-catalog [mag]
sigmagnr        1-sigma uncertainty in magnr [mag]
chinr           DAOPhot chi parameter of nearest source in reference image PSF-catalog
sharpnr         DAOPhot sharp parameter of nearest source in reference image PSF-catalog
sky             Local sky background estimate [DN]
magdiff         Difference: magap - magpsf [mag]
fwhm            Full Width Half Max assuming a Gaussian core from SExtractor [pixels]
classtar        Star/Galaxy classification score from SExtractor
mindtoedge      Distance to nearest edge in image [pixels]
magfromlim      Difference: diffmaglim - magap [mag]
seeratio        Ratio: difffwhm / fwhm
aimage          Windowed profile RMS afloat major axis from SExtractor [pixels]
bimage          Windowed profile RMS afloat minor axis from SExtractor [pixels]
aimagerat       Ratio: aimage / fwhm
bimagerat       Ratio: bimage / fwhm
elong           Ratio: aimage / bimage
nneg            number of negative pixels in a 5 x 5 pixel stamp
nbad            number of prior-tagged bad pixels in a 5 x 5 pixel stamp
rb              RealBogus quality score from Random Forest classifier; range is 0 to 1 where
                closer to 1 is more reliable
drb             RealBogus quality score from Deep-Learning-based classifier; range is 0 to 1 where
                closer to 1 is more reliable
ssdistnr        distance to nearest known solar system object if exists within 30 arcsec [arcsec]
ssmagnr         magnitude of nearest known solar system object if exists within 30 arcsec
                (usually V-band from MPC archive) [mag]
ssnamenr        name of nearest known solar system object if exists within 30 arcsec
                (from MPC archive)
sumrat          Ratio: sum(pixels) / sum(|pixels|) in a 5 x 5 pixel stamp where stamp is
                first median-filtered to mitigate outliers
magapbig        Aperture mag using 18 pixel diameter aperture [mag]
sigmagapbig     1-sigma uncertainty in magapbig [mag]
ranr            Right Ascension of nearest source in reference image PSF-catalog; J2000 [deg]
decnr           Declination of nearest source in reference image PSF-catalog; J2000 [deg]
sgmag1          g-band PSF-fit magnitude of closest source from PS1 catalog; if exists
                within 30 arcsec [mag]
srmag1          r-band PSF-fit magnitude of closest source from PS1 catalog; if exists
                within 30 arcsec [mag]
simag1          i-band PSF-fit magnitude of closest source from PS1 catalog; if exists
                within 30 arcsec [mag]
szmag1          z-band PSF-fit magnitude of closest source from PS1 catalog; if exists
                within 30 arcsec [mag]
sgscore1        Star/Galaxy score of closest source from PS1 catalog; if exists within 30 arcsec:
                0 <= sgscore <= 1 where closer to 1 implies higher likelihood of being a star
distpsnr1       Distance to closest source from PS1 catalog; if exists within 30 arcsec [arcsec]
ndethist        Number of spatially-coincident detections falling within 1.5 arcsec going back to
                beginning of survey; only detections that fell on the same field and
                readout-channel ID where the input candidate was observed are counted.
                All raw detections down to a photometric S/N of ~ 3 are included
ncovhist        Number of times input candidate position fell on any field and readout-channel
                going back to beginning of survey
jdstarthist     Earliest Julian date of epoch corresponding to ndethist [days]
jdendhist       Latest Julian date of epoch corresponding to ndethist [days]
scorr           Peak-pixel signal-to-noise ratio in point source matched-filtered detection image
tooflag         1 => candidate is from a Target-of-Opportunity (ToO) exposure;
                0 => candidate is from a non-ToO exposure
objectidps1     Object ID of closest source from PS1 catalog; if exists within 30 arcsec
objectidps2     Object ID of second closest source from PS1 catalog; if exists within 30 arcsec
sgmag2          g-band PSF-fit magnitude of second closest source from PS1 catalog; if exists
                within 30 arcsec [mag]
srmag2          r-band PSF-fit magnitude of second closest source from PS1 catalog; if exists
                within 30 arcsec [mag]
simag2          i-band PSF-fit magnitude of second closest source from PS1 catalog; if exists
                within 30 arcsec [mag]
```

szmag2	z-band PSF-fit magnitude of second closest source from PS1 catalog; if exists within 30 arcsec [mag]
sgscore2	Star/Galaxy score of second closest source from PS1 catalog; if exists within 30 arcsec: 0 <= sgscore <= 1 where closer to 1 => higher likelihood of being a star
distpsnr2	Distance to second closest source from PS1 catalog; if exists within 30 arcsec [arcsec]
objectidps3	Object ID of third closest source from PS1 catalog; if exists within 30 arcsec
sgmag3	g-band PSF-fit magnitude of third closest source from PS1 catalog; if exists within 30 arcsec [mag]
srmag3	r-band PSF-fit magnitude of third closest source from PS1 catalog; if exists within 30 arcsec [mag]
simag3	i-band PSF-fit magnitude of third closest source from PS1 catalog; if exists within 30 arcsec [mag]
szmag3	z-band PSF-fit magnitude of third closest source from PS1 catalog; if exists within 30 arcsec [mag]
sgscore3	Star/Galaxy score of third closest source from PS1 catalog; if exists within 30 arcsec: 0 <= sgscore <= 1 where closer to 1 => higher likelihood of being a star
distpsnr3	Distance to third closest source from PS1 catalog; if exists within 30 arcsec [arcsec]
nmtchps	Number of source matches from PS1 catalog falling within 30 arcsec
rfid	Processing ID for reference image to facilitate archive retrieval
jdstartref	Observation Julian date of earliest exposure used to generate reference image [days]
jdendref	Observation Julian date of latest exposure used to generate reference image [days]
nframesref	Number of frames (epochal images) used to generate reference image
rbversion	Version of Random Forest classifier model used to assign RealBogus (rb) quality score
drbversion	Version of Deep-Learning-based classifier model used to assign RealBogus (drb) quality score
dsnrms	Ratio: D/stddev(D) on event position where D = difference image
ssnrms	Ratio: S/stddev(S) on event position where S = image of convolution: D (x) PSF(D)
dsdiff	Difference of statistics: dsnrms - ssnrms
magzpsci	Magnitude zero point for photometry estimates [mag]
magzpsciunc	Magnitude zero point uncertainty (in magzpsci) [mag]
magzpscirms	RMS (deviation from average) in all differences between instrumental photometry and matched photometric calibrators from science image processing [mag]
nmatches	Number of PS1 photometric calibrators used to calibrate science image from science image processing
clrcoeff	Color coefficient from linear fit from photometric calibration of science image
clrcounc	Color coefficient uncertainty from linear fit (corresponding to clrcoeff)
zpclrcov	Covariance in magzpsci and clrcoeff from science image processing [mag^2]
zpmed	Magnitude zero point from median of all differences between instrumental photometry and matched photometric calibrators from science image processing [mag]
clrmed	Median color of all PS1 photometric calibrators used from science image processing [mag]: for filter (fid) = 1 2 3 PS1 color used = g-r g-r r-i respectively
clrrms	RMS color (deviation from average) of all PS1 photometric calibrators used from science image processing [mag]
neargaia	Distance to closest source from Gaia DR1 catalog irrespective of magnitude; if exists within 90 arcsec [arcsec]
neargaiabright	Distance to closest source from Gaia DR1 catalog brighter than magnitude 14; if exists within 90 arcsec [arcsec]
maggaia	Gaia (G-band) magnitude of closest source from Gaia DR1 catalog irrespective of magnitude; if exists within 90 arcsec [mag]
maggaiabright	Gaia (G-band) magnitude of closest source from Gaia DR1 catalog brighter than magnitude 14; if exists within 90 arcsec [mag]
exptime	Integration time of camera exposure [sec]

prv_candidates schema in packet:

metric	description
jd	Observation Julian date at start of exposure [days]
fid	Filter ID (1=g; 2=R; 3=i)
pid	Processing ID for image
diffmaglim	Expected 5-sigma mag limit in difference image based on global noise estimate [mag]
pdiffimfilename	filename of positive (sci minus ref) difference image
programpi	Principal investigator attached to program ID
programid	Program ID: encodes either public collab or caltech mode

```
candid          Candidate ID from operations DB
isdiffpos       t or 1 => candidate is from positive (sci minus ref) subtraction;
                f or 0 => candidate is from negative (ref minus sci) subtraction
tblid           Internal pipeline table extraction ID
nid             Night ID
rcid            Readout channel ID [00 .. 63]
field           ZTF field ID
xpos            x-image position of candidate [pixels]
ypos            y-image position of candidate [pixels]
ra              Right Ascension of candidate; J2000 [deg]
dec             Declination of candidate; J2000 [deg]
magpsf          Magnitude from PSF-fit photometry [mag]
sigmapsf        1-sigma uncertainty in magpsf [mag]
chipsf          Reduced chi-square for PSF-fit
magap           Aperture mag using 14 pixel diameter aperture [mag]
sigmagap        1-sigma uncertainty in magap [mag]
distnr          distance to nearest source in reference image PSF-catalog [pixels]
magnr           magnitude of nearest source in reference image PSF-catalog [mag]
sigmagnr        1-sigma uncertainty in magnr [mag]
chinr           DAOPhot chi parameter of nearest source in reference image PSF-catalog
sharpnr         DAOPhot sharp parameter of nearest source in reference image PSF-catalog
sky             Local sky background estimate [DN]
magdiff         Difference: magap - magpsf [mag]
fwhm            Full Width Half Max assuming a Gaussian core from SExtractor [pixels]
classtar        Star/Galaxy classification score from SExtractor
mindtoedge      Distance to nearest edge in image [pixels]
magfromlim      Difference: diffmaglim - magap [mag]
seeratio        Ratio: difffwhm / fwhm
aimage          Windowed profile RMS afloat major axis from SExtractor [pixels]
bimage          Windowed profile RMS afloat minor axis from SExtractor [pixels]
aimagerat       Ratio: aimage / fwhm
bimagerat       Ratio: bimage / fwhm
elong           Ratio: aimage / bimage
nneg            number of negative pixels in a 5 x 5 pixel stamp
nbad            number of prior-tagged bad pixels in a 5 x 5 pixel stamp
rb              RealBogus quality score; range is 0 to 1 where closer to 1 is more reliable
rbversion       version of RealBogus model/classifier used to assign rb quality score
ssdistnr        distance to nearest known solar system object if exists within 30 arcsec [arcsec]
ssmagnr         magnitude of nearest known solar system object if exists within 30 arcsec
                (usually V-band from MPC archive) [mag]
ssnamenr        name of nearest known solar system object if exists within 30 arcsec
                (from MPC archive)
sumrat          Ratio: sum(pixels) / sum(|pixels|) in a 5 x 5 pixel stamp where stamp is
                first median-filtered to mitigate outliers
magapbig        Aperture mag using 18 pixel diameter aperture [mag]
sigmagapbig     1-sigma uncertainty in magapbig [mag]
ranr            Right Ascension of nearest source in reference image PSF-catalog; J2000 [deg]
decnr           Declination of nearest source in reference image PSF-catalog; J2000 [deg]
scorr           Peak-pixel signal-to-noise ratio in point source matched-filtered detection image
magzpsci        Magnitude zero point for photometry estimates [mag]
magzpsciunc     Magnitude zero point uncertainty (in magzpsci) [mag]
magzpscirms     RMS (deviation from average) in all differences between instrumental photometry
                and matched photometric calibrators from science image processing [mag]
clrcoeff        Color coefficient from linear fit from photometric calibration of science image
clrcounc        Color coefficient uncertainty from linear fit (corresponding to clrcoeff)
```

9.3   Program-ID based Restrictions

An important note regarding the association of historical events (if found), including the assignment of flux upper-limits from previous overlapping epochs (if any) is *programID*-based filtering. The *programID* is an identifier (= 0, 1, 2, or 3) attached to each image exposure to distinguish between exposures observed under **public (MSIP)** survey time (ID=1); **collaboration or "private" time** (ID=2); or through **Caltech time** (ID=3; granted by the TAC). *ProgramID*=0 signifies no restriction according to ZTF observing time. Here are the access and historical-association rules for alerts discovered under each of these categories:
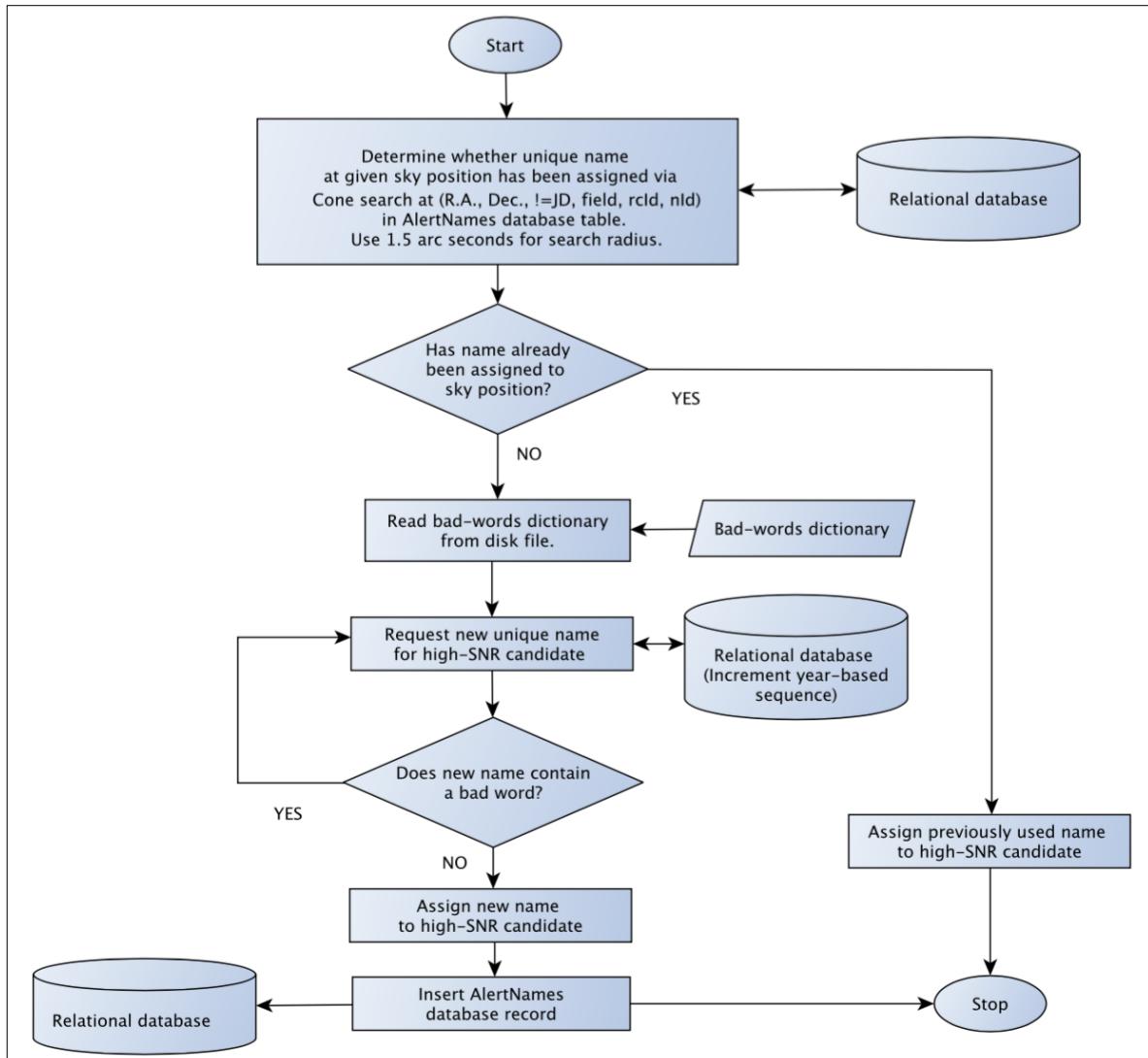
- Public-time alert packets (triggered from events detected in the public survey) will only contain public historical metadata.

- Private-time alert packets (triggered from events detected in the partnership surveys) will only contain historical metadata from the private surveys and the public survey.

- Caltech-time alert packets (triggered from events detected in Caltech TAC-allocated proposals) can contain historical metadata from all programs: Caltech-time, private surveys and the public survey.

9.4   Alert (object) Naming Process

Alert packets are assigned names (written in the *objectId* metadata field) using the algorithm outlined in Figure 13. Some details are as follows:

- The rationale is to "brand" each alert with a name to indicate its origin is from ZTF.

- In a nutshell, names (*objectId*s) are initially assigned to newly discovered alerts. Subsequent alerts that fall on/near a previously named alert (based on a positional search) going back to the beginning of commissioning (~ UT 2017-10-15) inherit that same name. The closest matching alert within the search radius (currently 1.5 arcsec) is used to handle ambiguous cases.

- *objectId* has the following format, for example: *ZTF18abcdefg*. This consists of the fixed ZTF prefix followed by the year and a seven (alphabet-based) character string, with range *a…z* for each character. Seven characters allows for $26^7 = 8,031,810,176$ possible unique names to be assigned over the course of a year. Sizing estimates for the number of possible alerts per year are well under this limit. This depends on the thresholds for alert generation (see Section 9.1).

- The assignment of *objectId* is sequential, beginning with ZTF18aaaaaaa. The naming process commenced during the early commissioning period (see above). Therefore, it's possible to obtain *objectId*s with prefix ZTF17, for example, for alerts that reoccur on a variable star.

- The *AlertNames* database table (schema given in Section 11.3) is used to bookkeep and re-assign *objectId*s from older alerts. This table will store one record for each *objectId* used, i.e., when it is first assigned. This includes associated metadata for the discovery epoch/position of the alert it was first assigned to.

Figure 13: Process used to assign or re-assign names (*objectId*) to alert packets

10 RECIPES / RECOMMENDATIONS ON DATA PRODUCT USAGE

10.1 Applying Photometric Calibration Solutions

10.1.1 Single exposure-(epochal) based catalogs

This section describes how photometric calibration solutions can be applied to the instrumental photometry in both the aperture and PSF-fit based single-epoch catalog products (file suffixes *sexcat.fits* and *psfcat.fits* respectively). This includes any new photometry you plan to measure on either the science or difference images. For the aperture-photometry based catalogs, see also Section 10.2 on aperture corrections.

The calibration of the readout-channel image filter-based zeropoints (ZP_f) and color-coefficients (c_f) was described in Section 6.5.2. Each image product is assigned its own unique solution for ZP_f, c_f – respectively the intercept and slope from a linear fit using matches of ZTF sources to a subset of calibrators selected from the PS1 catalog. The pertinent photometric calibration parameters reside in the epochal image and catalog FITS headers, for example:

```
MAGZP   =     27.4922460092007 / Magnitude zero point [mag]
NMATCHES=                 1802 / Number of photometric calibrators used
MAGZPUNC= 2.99175185805833E-07 / Magnitude zero point uncertainty [mag]
MAGZPRMS=  0.00552663035443809 / Magnitude zero point RMS deviation [mag]
CLRCOEFF=  0.00115165083397832 / Color coefficient from linear fit
CLRCOUNC= 5.61729479571772E-07 / Color coefficient uncertainty from linear fit
ZPCLRCOV= -3.9816364328644E-07 / Covariance in MAGZP and CLRCOEFF [mag^2]
PCOLOR  = 'g-R     '           / Calibrator color used for calibration
```

You can convert any instrumental g, r, i magnitude (m_{inst}) from a PSF-fit catalog (*mag* column therein) to calibrated magnitude *in the PS1 system* using:

$$m_{cal} = m_{inst} + ZP_f + c_f \left(m_1^{PS1} - m_2^{PS1} \right),$$

<div align="right">(Eq. 2)</div>

where ZP_f, c_f = MAGZP, CLRCOEFF respectively from the catalog or associated image header, and $m_1^{PS1} - m_2^{PS1}$ is the color of your source *in the PS1 filter system*. I.e., the filters m_1, m_2 must be reasonably close to those initially used to calibrate the specific filter in question (g, r, i): see Equation 1 in Section 6.5.2. The color used for the photometric calibration is indicated by the PCOLOR FITS keyword value. If no color information is available, you can set $m_1^{PS1} - m_2^{PS1} = 0$. ZP_f therefore represents the zeropoint for $m_1^{PS1} - m_2^{PS1} = 0$ and zero color implies a flat spectrum in flux density units: $f_\nu = constant$. This is because the PS1 calibrator magnitudes are already in the *AB* photometric system. It is also worth noting that the SDSS g' and r' magnitudes are very close to *AB* and therefore colors derived from these magnitudes can also be used in Equation 2.

Therefore, to account for source SED shape, you will require knowledge of the source's color in the *AB* system, more specifically, $m_1^{PS1} - m_2^{PS1}$ as defined by the PS1 effective filter wavelengths (λ_1, λ_2 : g, r, i, z). For a power-law approximation to the SED of the form $f_\nu \sim \nu^{-\alpha}$, the *AB*-based color can be estimated from the spectral index α using:

$$m_1 - m_2 \approx \alpha 2.5 \log_{10}\left(\frac{\lambda_2}{\lambda_1}\right). \qquad\qquad\qquad \text{(Eq. 3)}$$

Furthermore, in the *AB* (PS1) photometric system, calibrated magnitudes (m_{cal}) from Equation 2 can be converted to a monochromatic flux density using:

$$f_\nu = 3631 \times 10^{-0.4 m_{cal}} \, Jy. \qquad\qquad\qquad \text{(Eq. 4)}$$

10.1.2 Reference-image based catalogs

For reference-image catalog products, both aperture and PSF-fit based (file suffixes *sexcat.fits* and *psfcat.fits* respectively), the following information resides in their FITS headers, for example:

```
MAGZP   =               26.325 / Magnitude zero point [mag]
NMATCHES=                 1135 / Number of photometric calibrators used (median)
MAGZPUNC=          1.9728E-06 / Magnitude zero point uncertainty [mag] (median)
MAGZPRMS=         0.0012305597 / Magnitude zero point RMS deviation [mag]
CLRCOEFF=           -0.0005335 / Color coefficient from linear fit (median)
CLRCOUNC=           2.9719E-06 / Color coefficient uncertainty from linear fit (median)
PCOLOR  = 'g-R      '         / Calibrator color used for calibration
```

It's important to note that reference images are co-adds of multiple single-epoch images, which were individually photometrically calibrated upstream (Section 6.5.2). The single-epoch photometric solutions are propagated to a reference image product by rescaling the input single-epoch images (pixel values) to a common photometric zeropoint before co-addition (see Section 6.7). Currently, the fixed reference image magnitude zeropoints for *g, R, i* filters are MAGZP = 26.325, 26.275, 25.660 respectively. I.e., the reference image products are not recalibrated against any external catalog.

Furthermore, the following four metrics in the reference-image product headers are derived from a median of the corresponding input single-epoch image values: MAGZPUNC, NMATCHES, CLRCOEFF, and CLRCOUNC. The MAGZPRMS metric for reference-image products represents a robust RMS (based on the MAD) of the input single-epoch MAGZP values.

Photometric calibration solutions can be applied to reference-image products using the same methodology as for single-epoch image products outlined in Section 10.1.1. For the reference-image aperture-based catalogs, see also Section 10.2 on aperture corrections.

10.2 PSF-to-Aperture Photometry Offset Corrections

In all single-epoch and reference image *aperture-photometry* based catalogs (file suffix *sexcat.fits*), the following information is present in their FITS headers, for example:

```
FIXAPERS= '2,3,4,6,10,14'       / fixed aperture diameters used [pixels]
APCOR1  =   -9.321069700000E-01 / PSF_mag - AP_mag for AP.diam= 2.0 pix.
APCORUN1=    4.370083000000E-04 / Uncertainty in APCOR1 [mag]
APCOR2  =   -2.767840000000E-01 / PSF_mag - AP_mag for AP.diam= 3.0 pix.
APCORUN2=    2.819453000000E-04 / Uncertainty in APCOR2 [mag]
APCOR3  =   -9.507000000000E-02 / PSF_mag - AP_mag for AP.diam= 4.0 pix.
APCORUN3=    1.734361000000E-04 / Uncertainty in APCOR3 [mag]
APCOR4  =   -4.820000000000E-03 / PSF_mag - AP_mag for AP.diam= 6.0 pix.
APCORUN4=    1.309872000000E-04 / Uncertainty in APCOR4 [mag]
APCOR5  =   -3.620000000000E-04 / PSF_mag - AP_mag for AP.diam=10.0 pix.
APCORUN5=    1.910256000000E-04 / Uncertainty in APCOR5 [mag]
APCOR6  =    1.445000000000E-04 / PSF_mag - AP_mag for AP.diam=14.0 pix.
APCORUN6=    3.204096000000E-04 / Uncertainty in APCOR6 [mag]
```

The aperture-photometry catalogs contain **fixed**-concentric aperture photometry measurements for several apertures. The fixed aperture diameters used are listed in the `FIXAPERS` keyword. The corresponding fixed aperture measurements reside under the column name vector `MAG_APER(`i`)` in the FITS binary tables, where i is the aperture index (1…6) in order of increasing diameter as listed in `FIXAPERS`.

The aperture photometry catalogs were generated independently of the PSF-fit catalogs and contain a mixture of extended (non-PSF-like) and PSF-like sources. Due to variations in seeing across observation epochs, the *instrumental* photometry from *fixed* aperture photometry will not match that from PSF-fit photometry. The photometric calibration described in Section 6.5.2 was derived exclusively using filtered point-sources from the PSF-fit catalogs. Therefore, the per-image (single-exposure) calibration solutions are strictly only applicable to point-sources extracted therefrom, using the same PSF-fitting template.

The example header metadata above can be used to correct the aperture photometry measurements to achieve consistency (within random error) with those from PSF-fit photometry. This will allow photometric calibration solutions to be applied to the **fixed**-aperture measurements as well. For a raw aperture (instrumental) measurement on a *point-source* corresponding to a specific aperture diameter listed in `FIXAPERS`, the relevant PSF-fit to aperture correction (in magnitudes) is given by the relevant `APCOR`i keyword value (for the ith aperture with 1-σ uncertainty `APCORUN`i). The correction is applied in the sense:

$$m_{aper}(\text{corrected}) = m_{aper}(\text{raw}) + \texttt{APCOR}i, \qquad\qquad (\text{Eq. 5})$$

where m_{aper}(raw) is the corresponding fixed-aperture instrumental magnitude (`MAG_APER(`i`)` described above). It's important to note that each individual epochal-based image and reference-image product will have its own set of aperture corrections in the header.

Following this aperture correction, the appropriate photometric calibration solution from the FITS header can then be applied (Section 10.1.1 or 10.1.2). Putting it all together, the calibrated magnitude for a fixed instrumental aperture measurement on a point source is computed using:

$$\texttt{MAG\_APER}_{cal}(i) = \texttt{MAG\_APER}(i) + \texttt{APCORi} + \texttt{MAGZP} + \texttt{CLRCOEFF}(m_1^{PS1} - m_2^{PS2}),\qquad \text{(Eq. 6)}$$

where the calibration terms are described in Section 10.1.1.

For cases when the seeing is "reasonably good" (FWHM <~ 2.5 arcsec), the corrections for the largest apertures (D >~ 5FWHM or D >~ 12 pixels) are expected to be zero (within uncertainties). In this case, a correction need not be applied. If however the seeing is "reasonably good" but you still see significant corrections (i.e., not converging to zero for the largest apertures), it means a systematic related to initial PSF-fitting is present, i.e., PSF-fit instrumental fluxes are not catching the *total instrumental flux for isolated point sources*. For example, this will occur if the initial PSF-template estimate was incorrect. In the end, any PSF-fit catalog with a constant bias in the PSF-fit *instrumental* magnitudes *should* be implicitly corrected following application of its calibration solution. We recommend that any significantly non-zero aperture correction be always applied to the instrumental fixed aperture magnitudes to bring them in line with the PSF-fit magnitudes. This ensures the point-source calibration solutions can be applied in a consistent manner, as described by Equation 6.

As a rule of thumb, the optimal aperture diameter to use in *relatively unconfused fields*, where by "optimal" we mean in the maximal S/N sense, is ≈ 1.4*FWHM, where FWHM is the image quality. Therefore, for FWHM = 2.5 pixels, the optimal aperture diameter is 3.5 pixels. We recommend using the closest aperture diameter available, in this case either 3 or 4 pixels. It's worth examining measurements for the closest apertures above and below the optimal value to gauge overall accuracy. In confused high source-density fields however, the aperture photometry and initial aperture corrections derived relative to the PSF-fit catalog are likely to be suspect. In this case, you may want to use measurements from a small fixed-aperture, along with its aperture correction. This will help minimize confusion noise. In ultra-dense fields, you are probably better off using only the PSF-fit catalogs since these have the added benefit of de-blending.

Other instrumental measurements in the *sexcat.fits* catalogs (e.g., MAG_AUTO, MAG_BEST, MAG_ISO, MAG_PETRO, and related measures) cannot be immediately or "blindly" calibrated using the framework described above. These require offline analysis to place them on the *AB* (or PS1) photometric system. These measurements are provided for completeness, for example, to support extended-source photometry. However, there is still hope at calibrating these magnitudes. This is possible when the quoted aperture corrections (APCORi) converge to zero with increasing aperture size, which is often the case. This means that PSF-fit instrumental fluxes are catching the *total instrumental flux for point sources*. Furthermore, for such cases, the photometric calibration solution for a specific image is then generic in the sense that MAGZP represents the magnitude for 1 DN (1 ADU) of signal. Any large-enough aperture that catches the *total* instrumental flux from a source, regardless of its shape or size, can be calibrated directly using the calibration solution.

10.3 Science (epochal) pixel-level Bit-Mask Definitions

The archive product files with file suffix _mskimg.fits_ (see Section 7.2) represent bit-masks that encode at the pixel level: the state of the camera pixels (e.g., those tagged during characterization as permanently "bad" and unusable); encountered to be transiently anomalous during realtime processing (e.g., saturated, associated with a hot pixel, cosmic ray, and/or artifact); or is associated with signal from an extracted source. The bit definitions implemented at the time of writing are below.

For example, to find all science-image pixels which are uncontaminated, but which may contain clean extracted-source signal (bits 1 or 11), one would "_logically_ **AND**" the corresponding mask-image with the template value 6141 (= $2^0 + 2^2 + 2^3 + 2^4 + 2^5 + 2^6 + 2^7 + 2^8 + 2^9 + 2^{10} + 2^{12}$) and retain pixels where this operation yields _zero_. To then find which of these pixels contain source signal, one would "**AND**" the resulting image with 2050 (= $2^1 + 2^{11}$) and retain pixels where this operation is _non-zero_.

Both the aperture-based and PSF-fit photometry catalog products encode any masked pixels that fall within the effective measurement areas of individual sources through their "`FLAGS`" column. If more than one condition below was applicable, the bits were combined using a "_logical_ **OR**".

```
BIT00   =                    0 / AIRCRAFT/SATELLITE TRACK
BIT01   =                    1 / CONTAINS SEXTRACTOR DETECTION
BIT02   =                    2 / LOW RESPONSIVITY
BIT03   =                    3 / HIGH RESPONSIVITY
BIT04   =                    4 / NOISY
BIT05   =                    5 / GHOST FROM BRIGHT SOURCE
BIT06   =                    6 / GHOST FROM CHARGE SPILLAGE (ONLY FOR OBSMJD <= 58779)
BIT07   =                    7 / PIXEL SPIKE (POSSIBLE RAD HIT)
BIT08   =                    8 / SATURATED
BIT09   =                    9 / DEAD (UNRESPONSIVE)
BIT10   =                   10 / NAN (not a number)
BIT11   =                   11 / CONTAINS PSF-EXTRACTED SOURCE POSITION
BIT12   =                   12 / HALO FROM BRIGHT SOURCE
BIT13   =                   13 / RESERVED FOR FUTURE USE
BIT14   =                   14 / RESERVED FOR FUTURE USE
BIT15   =                   15 / RESERVED FOR FUTURE USE
```

10.4 Science (epochal) image-level Quality Flags

These are located in the FITS headers of the science image product files; suffix: _sciimg.fits_ (Section 7.2).

First, there's an overall "INFOBITS" string, for example:

```
INFOBITS=                        1234 / infoBits string; bit(s) = 1,4,6,7,10
```

I.e., $1234 = 2^1 + 2^4 + 2^6 + 2^7 + 2^{10}$. The INFOBITS value is based on any of the following being set. The thresholds were determined from analyzing the distributions of image/catalog quality metrics.

```
BIT00   =                         0 / MEDIAN BACKGROUND LEVEL is >  4000.0 DN
BIT01   =                         1 / ROBUST PIXEL NOISE (RMS) is > 100.0 DN
BIT02   =                         2 / PERCENT OF SATURATED PIXELS is > 50.0%
BIT03   =                         3 / IMAGE SEEING (FWHM) is > 5.0"
BIT04   =                         4 / MAGLIM FROM PSF CATALOG UNCERTS is < 17.5 mag
BIT05   =                         5 / MAGLIM FROM EMPIRICAL FORMULA is < 17.5 mag
BIT06   =                         6 / NUMBER OF PSF CATALOG SOURCES is < 10
BIT07   =                         7 / NUMBER OF SEXTRACTOR CAT SOURCES is < 20
BIT08   =                         8 / PHOTOMETRIC CALIBRATION NOT POSSIBLE
BIT09   =                         9 / NUMBER OF PHOTOMETRIC CALIBRATORS is < 30
BIT10   =                        10 / NUMBER OF MATCHES FOR PHOTOMETRIC CAL < 5
BIT11   =                        11 / RMS in PHOTCAL RESIDUALS (ZPs) is > 0.2 mag
BIT12   =                        12 / ASTROMETRIC CALIBRATION NOT POSSIBLE
BIT13   =                        13 / PCT OF REF-CAT MATCHES FOR ASTROM is < 0%
BIT14   =                        14 / NUMBER OF MATCHES FOR ASTROM CAL is < 100
BIT15   =                        15 / MIN NUM OF IN-GRID MATCHES FOR ASTROM is < 5
BIT16   =                        16 / REL-PIXEL SCALE OUT OF RANGE: -0.20% - 2.00%
BIT17   =                        17 / PCT CHANGE IN SCAMP CHI-SQUARE is > 0.0%
BIT18   =                        18 / CHI-SQUARE FROM SECOND PASS SCAMP is > 30 (MJD <= 58813)
BIT19   =                        19 / MAX IN-GRID MEDIAN ASTROM RESID is > 0.7"
BIT20   =                        20 / PSF CREATION (HENCE PSF-FIT PHOT) NOT POSSIBLE
BIT21   =                        21 / PCT OF DETECTED MATCHES FOR ASTROM is < 50%
```

An overall science-image processing status flag is also written to the header (keyword STATUS) and is set to 0 to indicate the image is unusable. At the time of writing, this occurs if any of the following bits from above were set in the INFOBITS string: 2,6,7,8,10,12,13,14,15,16,17,18,19,20. These bits are considered "fatal". Currently, only images with "status = 1" (i.e., not assigned any of these fatal bits) are copied to the archive. Therefore, all images in the archive will have the following in their headers:

```
STATUS  =                         1 / overall processing flag: 0 = bad; 1 = usable
```

10.5 Reference image-level Quality Flags

These are located in the FITS headers of the reference image product files; suffix: _refimg.fits_ (Section 7.4).

First, there's an overall "INFOBITS" string, for example:

```
INFOBITS=                      3 / infoBits string; bit(s) = 0,1
```

I.e., $3 = 2^0 + 2^1$. The INFOBITS value is based on any of the following being set. The thresholds were determined from analyzing the distributions of image/catalog quality metrics. Most of the metrics here are medians or maxima over the input science-images used to construct the reference image. These are still purely suggestive and subject to change.

```
BIT00   =                      0 / MEDIAN BACKGROUND LEVEL is >  4000.0 DN
BIT01   =                      1 / ROBUST PIXEL NOISE (RMS) is > 100.0 DN
BIT02   =                      2 / PERCENT OF BAD/SATURATED PIXELS is > 65.0%
BIT03   =                      3 / MEDIAN OF INPUT SEEING (FWHM) is > 5.0"
BIT04   =                      4 / MAXIMUM OF INPUT FWHM is > 4.0"
BIT05   =                      5 / MAGLIM FROM PSF CATALOG UNCERTS is < 18.0 mag
BIT06   =                      6 / MAGLIM FROM EMPIRICAL FORMULA is < 18.0 mag
BIT07   =                      7 / NUMBER OF PSF CATALOG SOURCES is < 10
BIT08   =                      8 / NUMBER OF SEXTRACTOR CAT SOURCES is < 10
BIT09   =                      9 / MEDIAN NUM. OF INP. PHOTCAL MATCHES is < 15
BIT10   =                     10 / ROBUST RMS OF INPUT IMAGE ZPs is >  3.5 mag
BIT16   =                     16 / PSF CREATION (HENCE PSF-FIT PHOT) NOT POSSIBLE
```

An overall reference-image processing status flag is also written to the header and is set to 0 (to indicate a probably unusable image) if any of the following bits from above were set in the INFOBITS string: 2,5,7,8,16. These bits are considered "fatal".

```
STATUS  =                      1 / overall processing flag: 0 = bad; 1 = usable
```

10.6 Column Definitions in Aperture & PSF-fit Catalog Files

For the aperture-based single-epoch and reference image photometry catalogs (filename suffixes *_sexcat.fits and *_refsexcat.fits respectively), all metrics are defined by header keywords in their FITS binary tables. More information on these columns can be obtained from the *SExtractor* documentation: *https://www.astromatic.net/pubsvn/software/sextractor/trunk/doc/sextractor.pdf*

For the PSF-fit photometry catalogs however (filename suffixes *_psfcat.fits and *_refpsfcat.fits), the columns are not defined in their FITS binary tables. These are defined as follows.

sourceid – source counter (sequential).
xpos [pixels] – source *x* position in native CCD-quadrant image.
ypos [pixels] – source *y* position in native CCD-quadrant image.
ra [degrees] – Right Ascension (J2000)
dec [degrees] – Declination (J2000)
flux [DN] – integrated counts for source in detector digital units following ADC.
sigflux [DN] – one-sigma uncertainty in flux measurement.
mag [mag] – instrumental magnitude ($= -2.5\log_{10}[\text{flux}]$).
sigmag [mag] – one-sigma uncertainty in instrumental magnitude ($\sim 1.086\,\text{sigflux} / \text{flux}$).
snr – signal-to-noise ratio (= flux / sigflux)

chi – robust estimate of ratio:
   RMS in PSF-fit residuals / expected RMS from uncertainties.

sharp – equal to the difference: $\text{FWHM}^2_{\text{obs}} - \text{FWHM}^2_{\text{psfmodel}}$;
   $\sim 0 \Rightarrow$ source is PSF like
   $\gg 0 \Rightarrow$ source is extended relative to PSF model
   $\ll 0 \Rightarrow$ cosmic ray glitch (pixel spike or hard-edged artifact).

flags – source quality flags:
   $-1 \Rightarrow$ source falls on/near an image edge where source size is defined by size of PSF-stamp. All other values represent the *bitwise-OR* of all pixel values in the corresponding bit-mask image falling in a 5×5 pixel region centered on the source *x, y* position. The bits were defined in Section 10.3.

11 PIPELINE OPERATIONS DATABASE SCHEMA

The ZSDS is heavily reliant on a database to store and bookkeep all raw, processed, and archived products files, image metadata, observing-system telemetry, QA metrics, timing information, extracted transient events, and alert packets. The operations database stores data going back to the beginning of the commissioning period (~ mid October 2017). Schema diagrams are shown below. Data dictionaries (with definitions of all columns) are available on GitHub, specifically:
https://github.com/Caltech-IPAC/ztf/blob/master/database/schema/ztfOpsTables.sql
https://github.com/Caltech-IPAC/ztf/blob/master/database/schema/ztfideTables.sql
Access to this repository is restricted to ZTF project team members and partners.

11.1 Tables related to raw exposure image ingestion
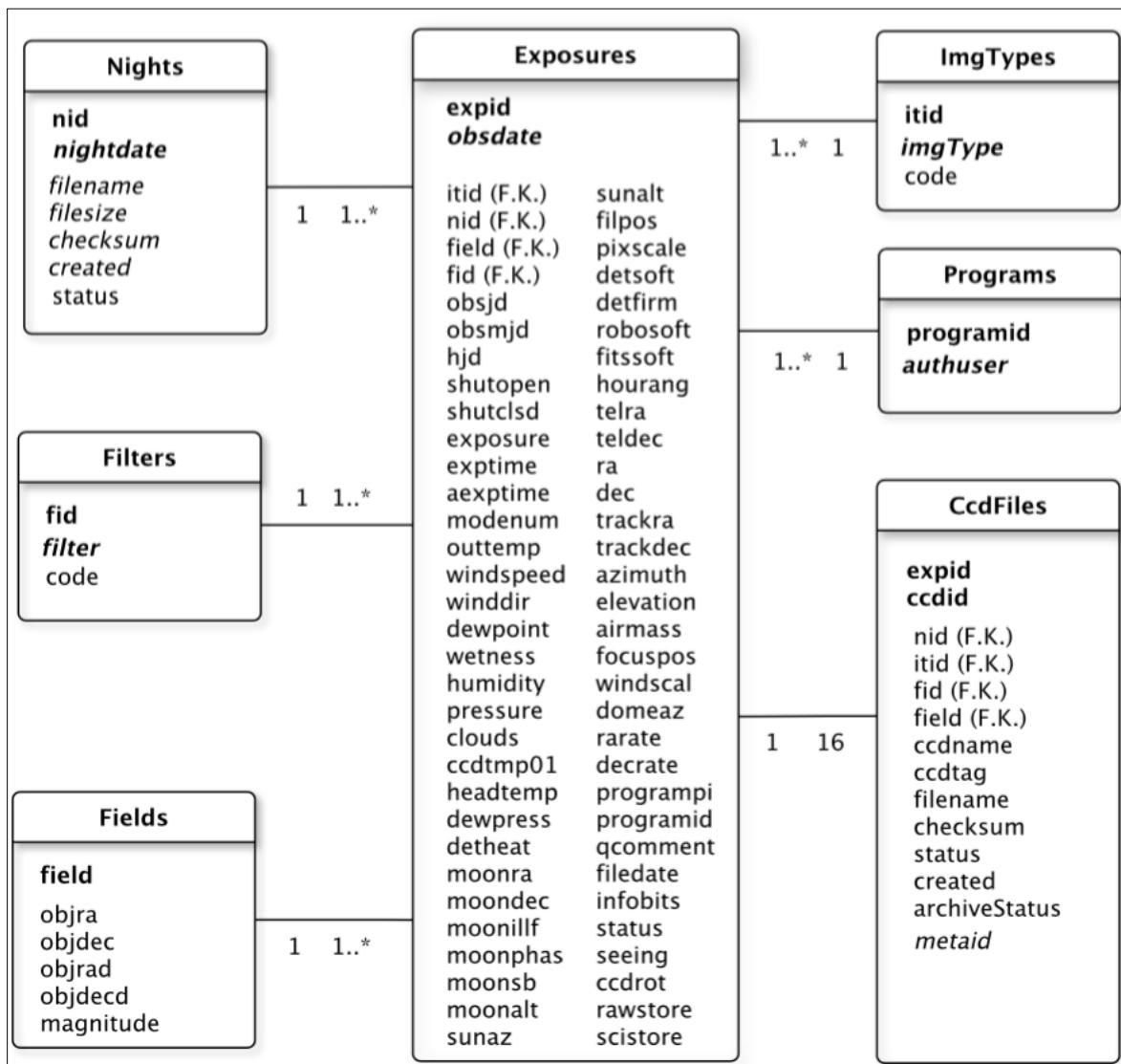
Figure 14: Schema for tables related to raw-image ingestion

11.2 Tables related to readout-channel image-based products and QA
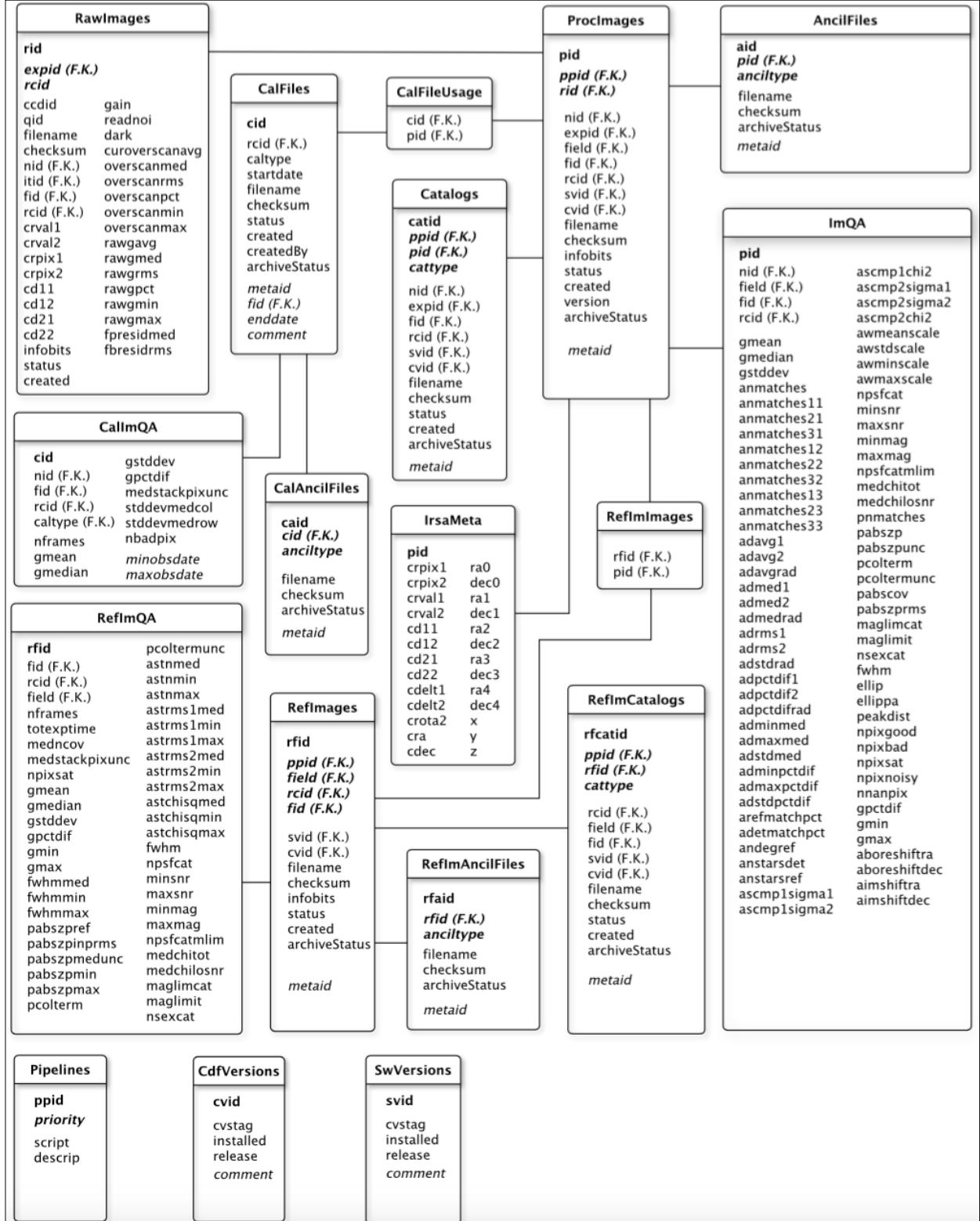
Figure 15: Schema for tables related to image data products and QA

11.3 Tables related to Transient Source Candidates from Image Subtraction
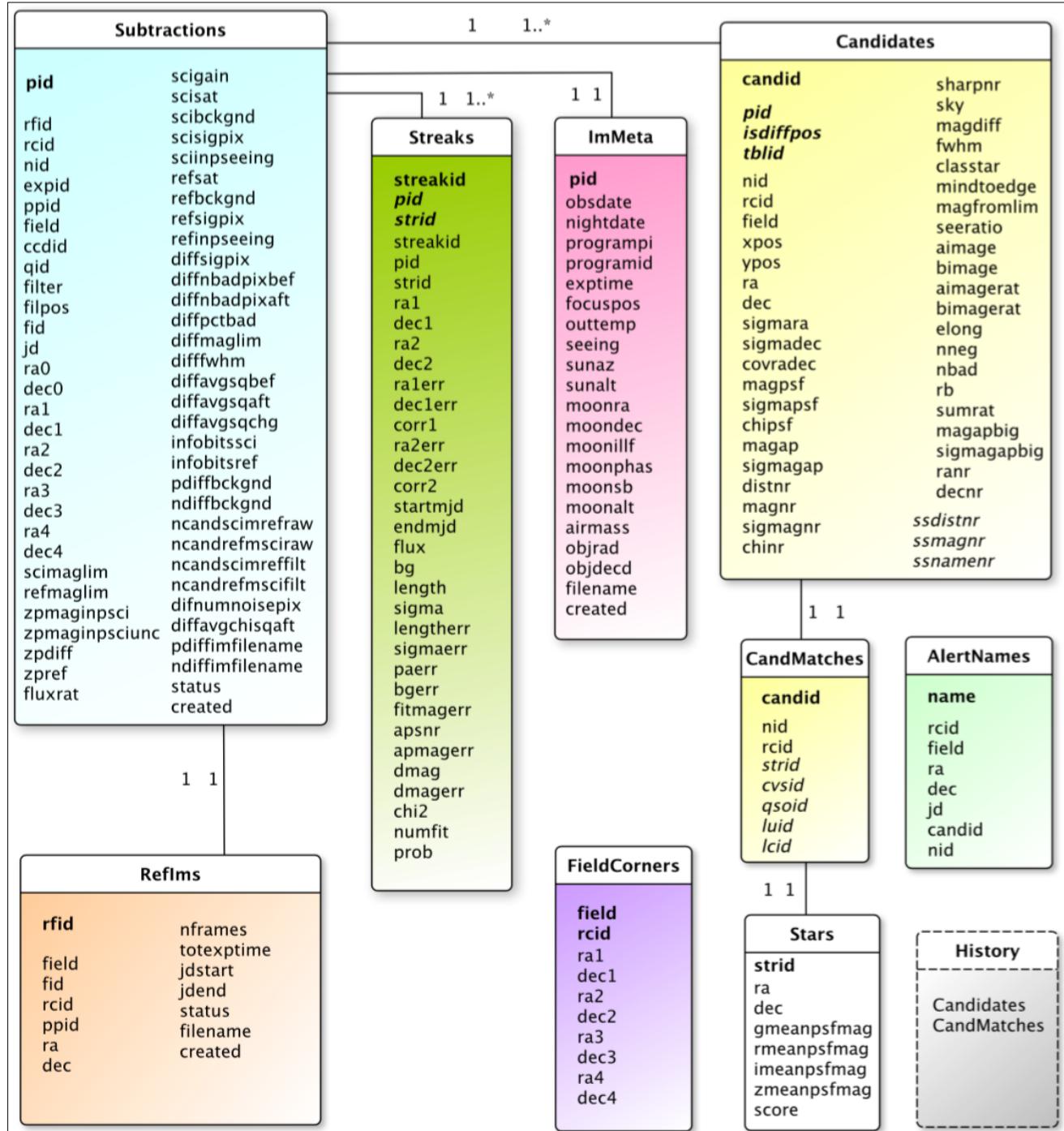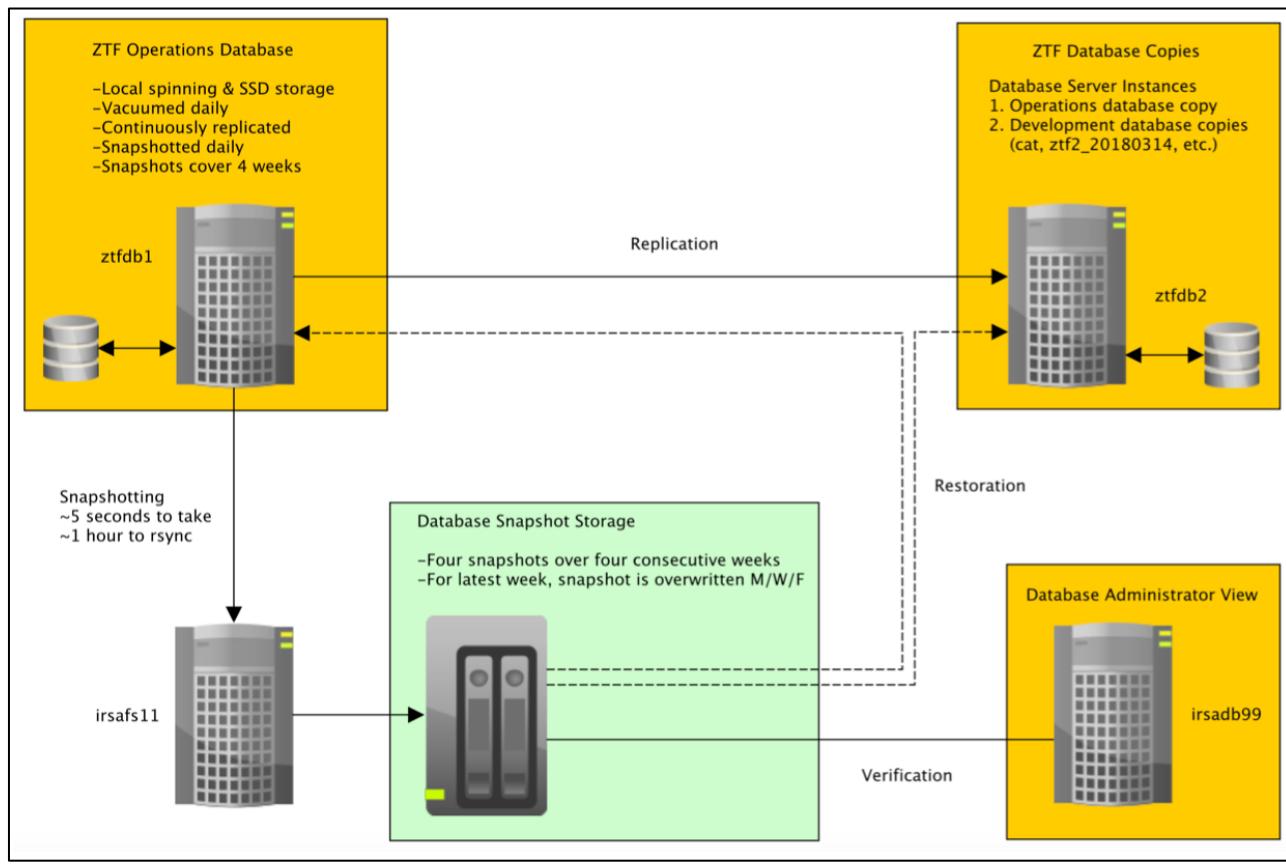
Figure 16: Schema for tables related to transient events (candidates) and diff-image QA

The figure below gives an overview of the database management process during operations. The primary operations database is snapshotted (backed-up) regularly to protect its data, as well as continuously replicated into a separate copy. The reason for this replication is so that any analyses can be performed outside the operations environment, with minimal impact to the operations database. Currently, access is restricted to a subset of project team members.
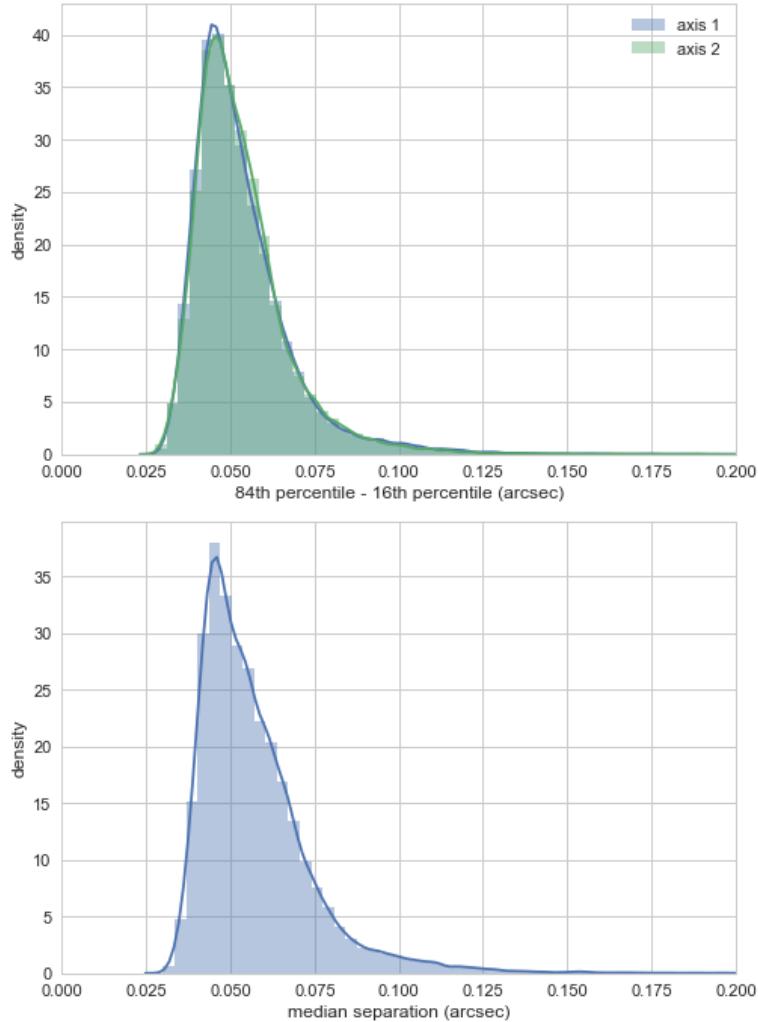
Database snapshots are copied to a network-attached filesystem. Snapshots cover four weeks of the latest content and are available at any given time. Tape backups of the snapshotted data are also made on a regular schedule, and moved to an off-site repository. Both the primary and replicated database servers run PostgreSQL 9.6 server software under the Solaris 11 operating system.

Database management plan: snapshotting and replication

12 PRELIMINARY ON-SKY PERFORMANCE
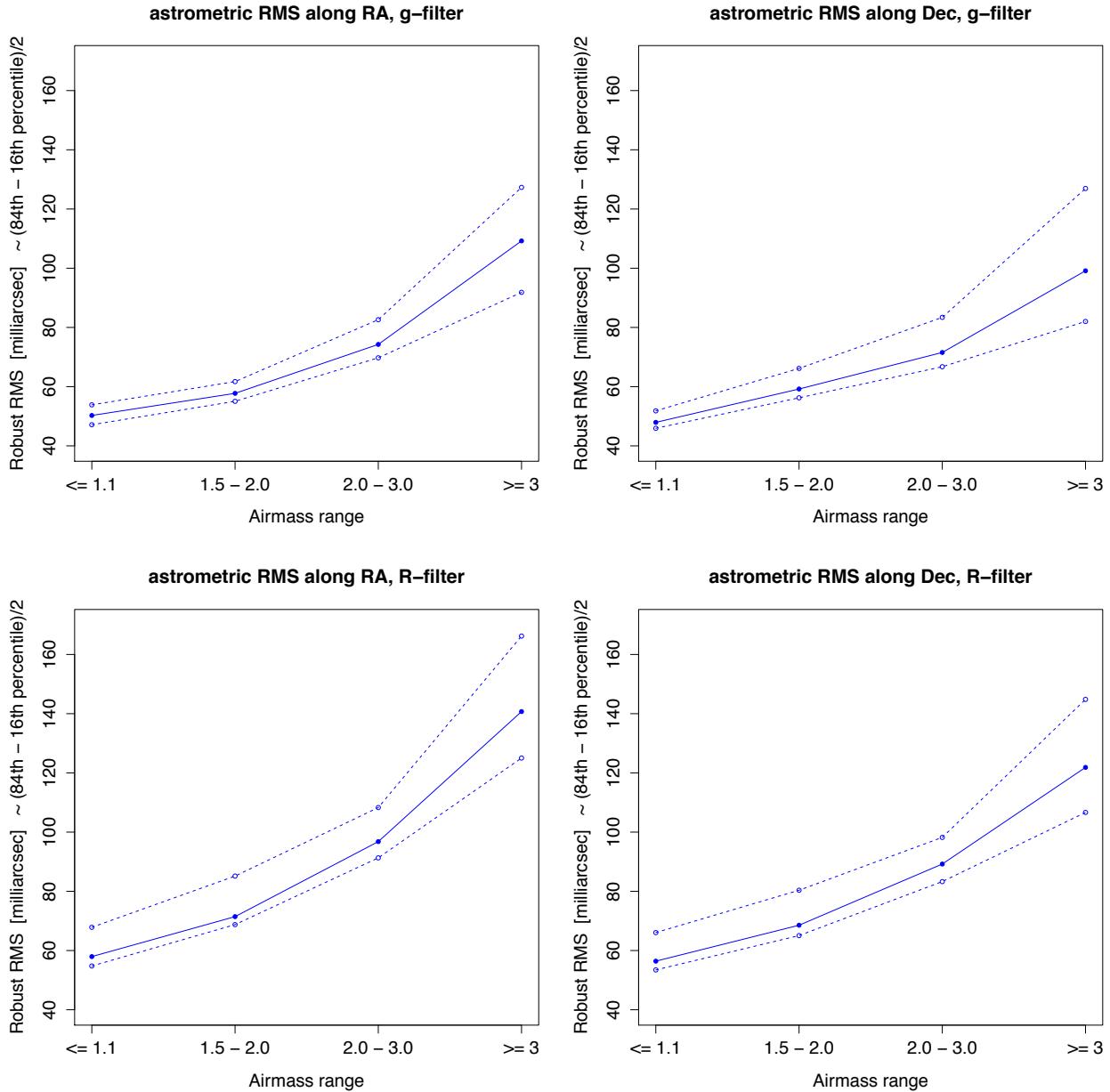
12.1 Astrometric Accuracy

The accuracy in the pipeline-reconstructed astrometry with respect to *Gaia* DR1 per quadrant image is typically 45 to 50 milli-arcsec (peak RMS deviation per R.A. or Dec. axis on the sky) at low airmass. For a 2D Gaussian distribution, this is equivalent to the median radial separation. The distributions in Figure 17 were generated from fields observed on the night of 2017-11-21 for *airmass* < 3. This is a representative night with several different science-validation experiments being executed at a range of galactic latitudes.

Figure 17: overall astrometric accuracy per processed quadrant image with respect to *Gaia* DR1. _Top_: distribution in robust estimate for the RMS deviation based on half the 84th - 16th percentile range where axes 1,2 = R.A., Dec. _Bottom_: distribution in the median radial separation.

Figure 18 shows the robust RMS measure (equivalent to that shown in the top histogram in Figure 17) as a function of airmass range, for the *g* and *R* filters using fields observed from 2017-12-01 onwards. Above *airmass* = 3, there is significant degradation in the astrometry from differential atmospheric refraction on the scale of a readout channel. Corrections for this effect will be considered in future. Currently, the primary survey is limited to airmass ≤ 2 where the astrometric residuals are relatively small for most science programs. This corresponds to altitudes of >∼ 30° above the horizon and gives the requisite 3π sky coverage for ZTF over six months.

Figure 18: astrometric accuracy as a function of airmass range. _Top-panels_ are for _g_-filter; _bottom-panels_ are for _R_-filter; _left_ and _right_ panels are for R.A. and Dec. respectively. The middle filled circles (with solid lines) are median RMS values over all 64 readout channel
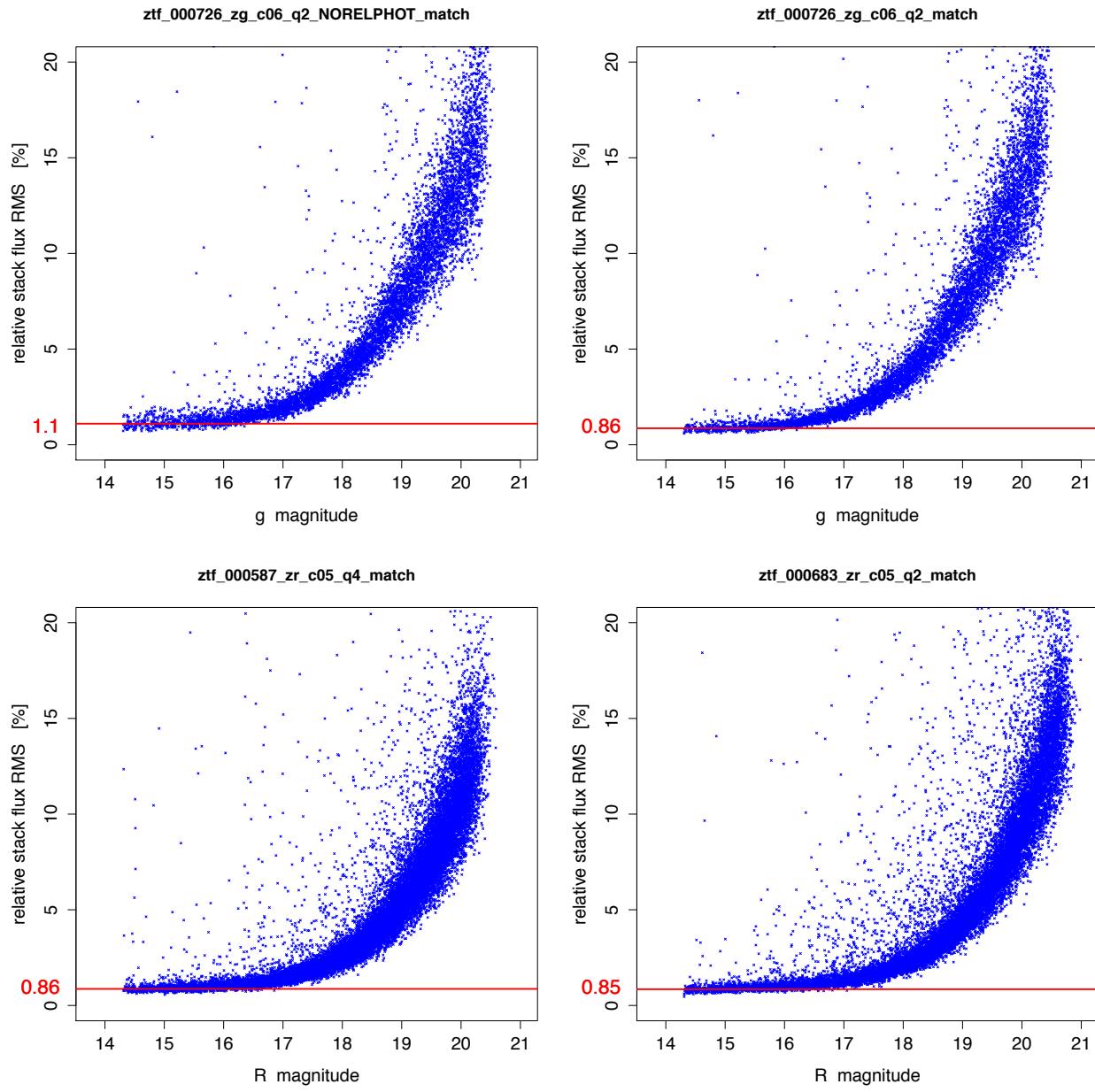
images; top open circles (with dashed lines) are maxima over all readout-channels; bottom open circles (with dashed lines) are minima over all readout-channels.

12.2 Photometric Precision (Repeatability)
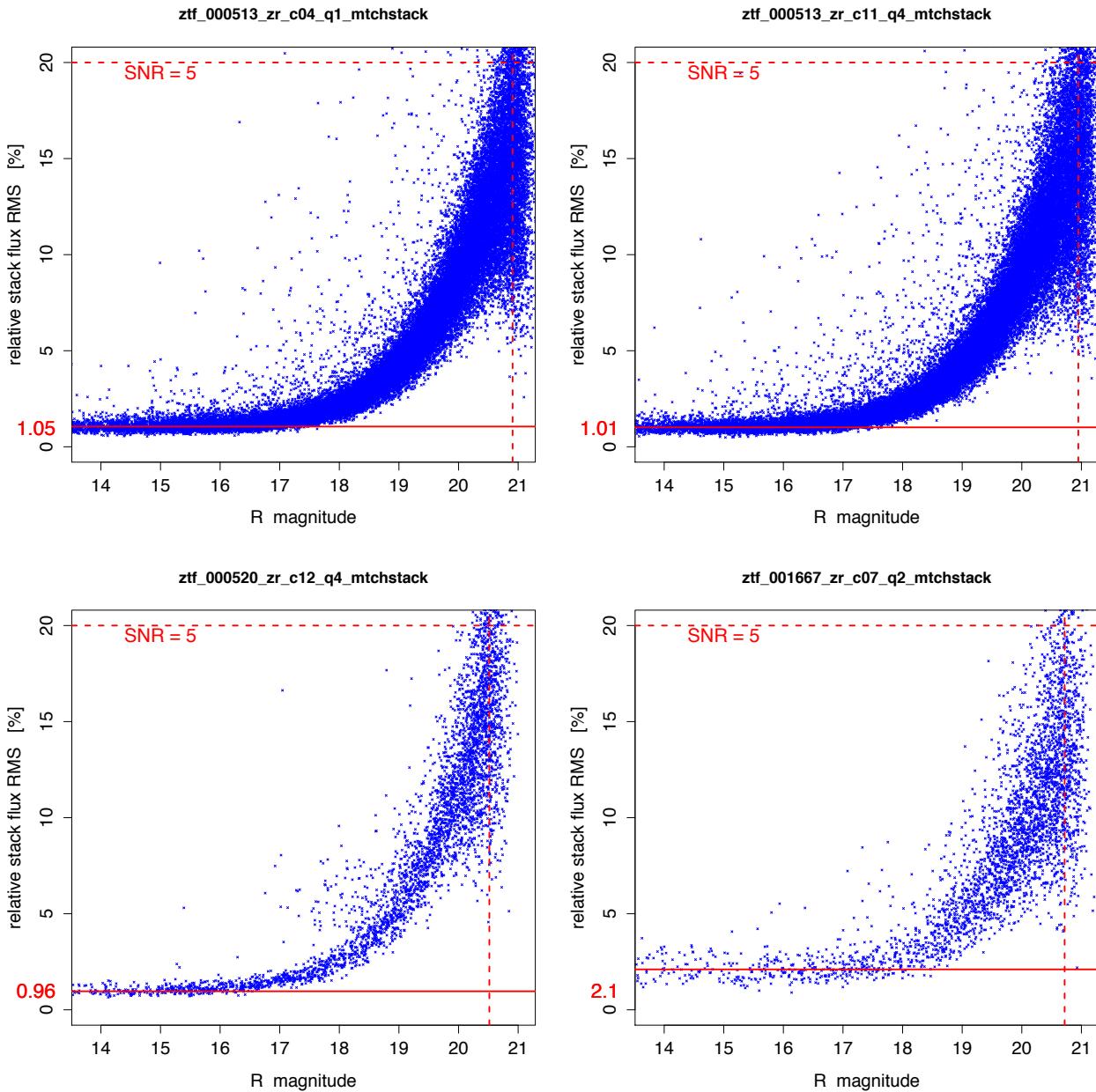
Figure 19 quantifies the photometric repeatability as a function of magnitude using the *magrms* lightcurve statistics extracted from the *matchfile* products (see Section 6.8 for a description). The inputs are PSF-fit catalogs (**psfcat.fits*). The *matchfile* creation includes a step to refine the photometric zero-points across epochs (relative to each other) prior to source matching. The top panels in Figure 19 compare the flux RMS without (left) and with relative photometry turned on. The improvement for this image is ~ 15% at bright fluxes, resulting in a relative precision of 0.79% or ~ 8.5 millmag.

In general, the internal photometric precision typically lies between 5 and 10 millimag. Improvements due to relative photometric refinement are modest, although one should expect better improvement. The primary limiting factor that determines the resulting photometric precision for a field/readout-quadrant is flat-fielding accuracy. Therefore, the plots shown below can be used to assess flat-fielding accuracy.

For comparison, Figure 20 shows the results of an independent analysis that does not use the *matchfile* products. Here, the PSF-fit catalogs were positionally-matched against each other and the source fluxes therein are globally rescaled to match those in the reference image (co-add) PSF-fit catalog. I.e., each input epochal catalog is rescaled to a common zeropoint equal to that of the reference image. There is no mutual refinement of the individual epochal zeropoints as done in *matchfile* creation. Only epochal catalogs extracted from overlapping images observed at airmass ≤ 1.2 and between 2018-02-04 to 2018-03-04 were used. These plots provide an independent check of the *matchfile*-derived stack-RMS estimates. In general, the results are consistent with those from the *matchfile* products (Figure 19) with no relative photometric refinement. Furthermore, the 5-σ magnitude limits (corresponding to a 20% flux-RMS) are consistent with those from uncertainty model estimates (Section 12.3).

Figure 19: Relative flux RMS in stacked multi-epochal photometry (repeatability) extracted from *matchfile* products. The inputs are PSF-fit catalogs. The *top left and right panels* correspond to the same field/readout-quadrant image and are for runs where the relative-photometric refinement step was turned off and on respectively.
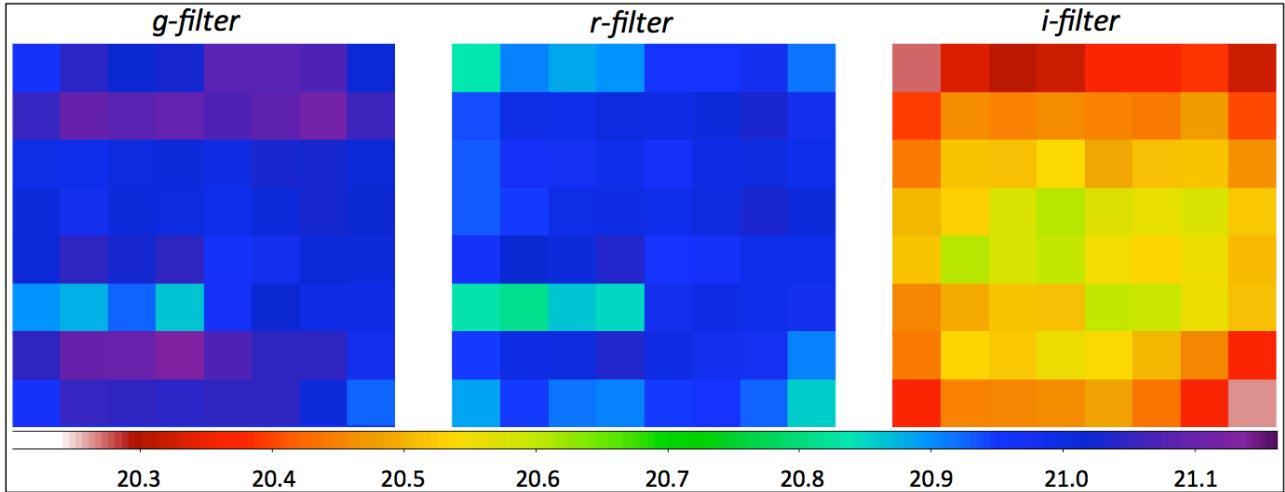
Figure 20: Independent estimates of the relative flux-RMS in stacked multi-epochal photometry (repeatability) using the PSF-fit catalogs. Unlike Figure 19, these do not use the *matchfile* products, but are based on matching the catalogs directly. Vertical dashed lines correspond to the approximate 5-σ magnitude limits (20% flux-RMS). See Section 12.2 for details.
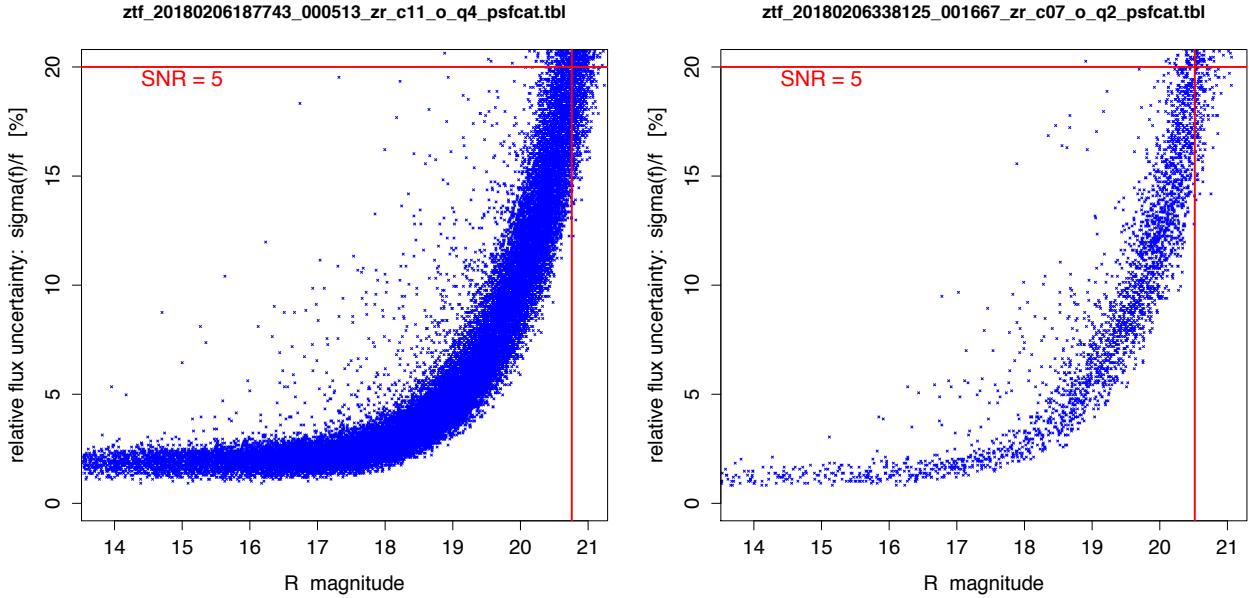
12.3 Sensitivity Limits

We provide an estimate of the 5-σ magnitude limit in each science image FITS header (*_sciimg.fits* product). These are represented by the header keyword MAGLIM. This is estimated from the accompanying PSF-fit catalog product (*_psfcat.fits*) by computing the median magnitude of sources falling within a narrow range of S/N centered on S/N=5. This S/N is based on the photometric uncertainties in the PSF-fit catalog as propagated from an uncertainty model involving the electronic gain, pixel RMS, and fitting error. Therefore, the accuracy of MAGLIM is contingent on the plausibility of the photometric uncertainty model. These estimates have been validated in a global sense using matches to sources in the PS1 catalog and photometric errors reported therein. Furthermore, these are consistent with the 5-σ (= 20% flux-RMS) temporal repeatability estimates from Figure 20.

Figure 21 shows the spatial variation in MAGLIM for each filter by medianing the values from several thousand quadrants in g and r, and a few hundred quadrants in i across 75 photometric nights spanning 2018-04-01UT to 2018-06-30UT, or ~ 3 lunations. All are at airmass ≤ 1.05 (i.e., near zenith) and all are from 30 sec exposures, the nominal integration time for ZTF. The overall 5σ median magnitude limits across all quadrants and epochs used in these maps are $g = 21.067 \pm 0.003$, $r = 21.012 \pm 0.002$, and $i = 20.51 \pm 0.01$ mag. The 1σ uncertainties are based on $\sqrt{N}$ statistics. The drop in sensitivity in the corners is due to vignetting. Figure 22 shows the relative flux-uncertainty as a function of magnitude from two PSF-catalogs. The 5-σ magnitude limits (20% flux-uncertainty) are shown.

Figure 21: focal plane map of median MAGLIM for *g, r, i* (*left* to *right*). Color bar is shown on the bottom. The CCD quadrants follow the same layout as in Figure 1. The input science images span Apr 1 to Jun 30 2018, were observed at airmass ≤ 1.05, and all have EXPTIME=30 sec.
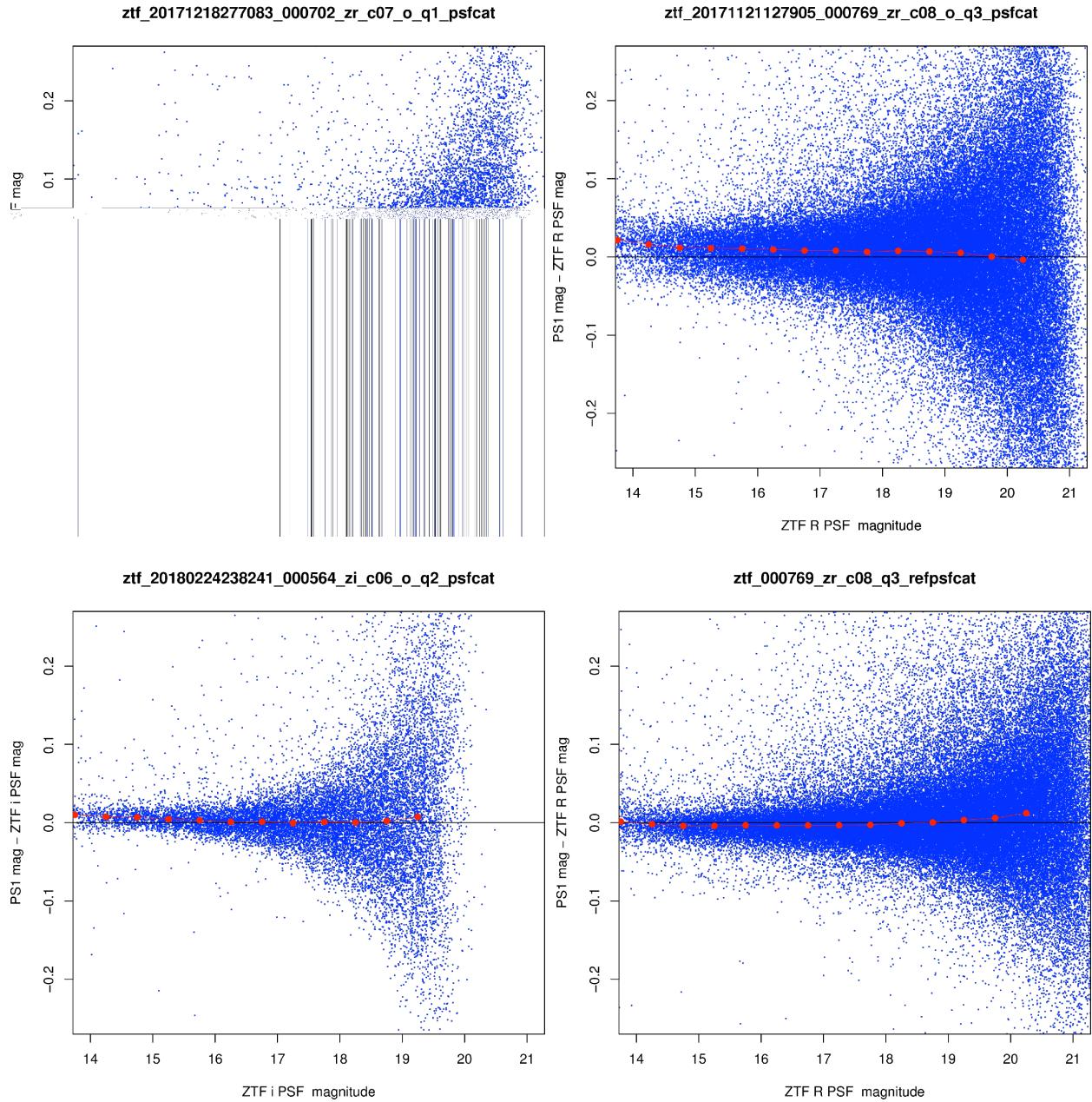
Figure 22: relative (model-based) flux uncertainty as reported in two PSF-fit catalogs, for a dense galactic plane field (*left*) and a high galactic latitude field (*right*). Vertical lines correspond to the approximate 5-σ magnitude limits (20% flux-uncertainty cut).

12.4 Photometric Calibration Check
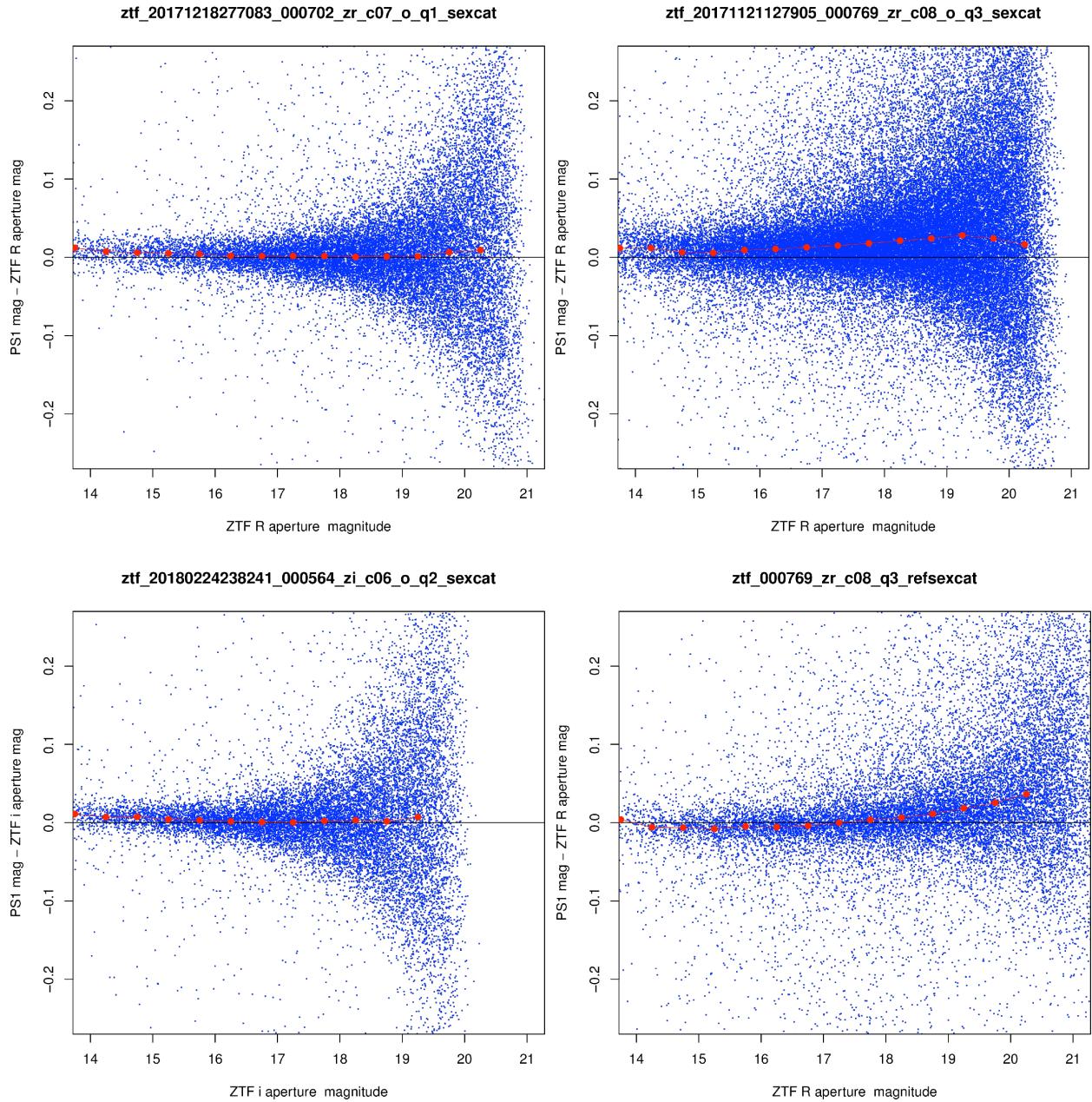
Figures 23 and 24 show the difference between PS1 and calibrated ZTF magnitudes for a number of quadrant-image PSF-fit (**_psfcat.fits*) and aperture-based (**_sexcat.fits*) catalogs respectively. Equivalent plots for one reference image are also shown. The purpose here is to (i) determine the level of any photometric biases, and (ii) quantify the overall calibration accuracy with respect to PS1.

Global analyses are still in progress, but early results show there are photometric biases from PSF-fitting of up to 0.02 mag for predominately bright sources with g, r, $i <\sim$ 15.5 mag. These biases are not yet understood. The results are also field-dependent. As expected, biases will be larger for the aperture-based catalogs in high source-density fields across the full magnitude range (Figure 24). PSF-fitting is more immune to contamination from crowding.
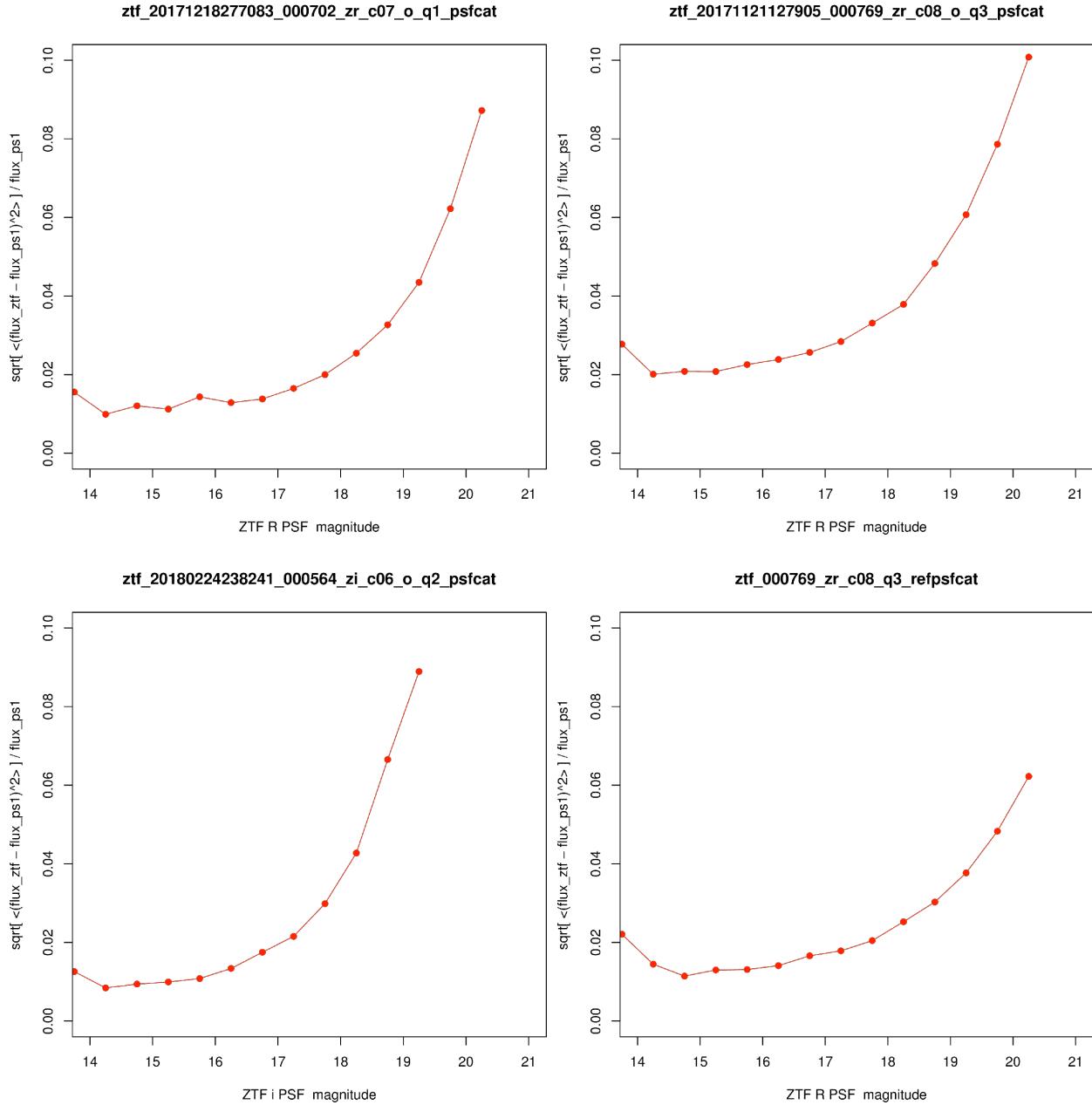
Figure 25 shows relative flux RMS-deviation with respect to PS1 in 0.5-magnitude bins corresponding to the PSF-fit residuals in Figure 23. The dispersion here is dominated by the measured instrumental photometry, e.g., local background estimation, PSF estimation, flat-fielding, source confusion, and errors in global astrometry and centroiding. There is also some scatter in the PS1 measurements themselves. The RMS scatter with respect to PS1 is field-dependent and varies from ~ 1% to ~ 2% at bright fluxes. A more global analysis is pending.

Figure 23: difference between PS1 and ZTF PSF-fit magnitudes from PSF-fit photometry catalogs for three epochal images (*top left, top right, bottom left*) and a reference image (*bottom right*). Comparisons using the corresponding aperture photometry catalogs are shown in Figure 24. Case 000769_zr_c08_q3 corresponds to a galactic-plane observation.

Figure 24: difference between PS1 and ZTF aperture-derived magnitudes from aperture photometry catalogs for three epochal science images (*top left, top right, bottom left*) and a reference image (*bottom right*). Comparisons using the corresponding PSF-fit catalogs are shown in Figure 23. Case 000769_zr_c08_q3 corresponds to a galactic-plane observation.

94

Figure 25: relative flux RMS-deviation with respect to PS1 corresponding to the PSF-fit comparisons in Figure 23: *top left, top right, and bottom left* **are for three epochal science images;** *bottom right* **is for a reference image. Case 000769_zr_c08_q3 corresponds to a galactic-plane observation.**

13 CAUTIONARY NOTES

Below we summarize the caveats and warnings when using science products from the ZSDS. Advisories on how to identify and omit bad-quality data from archive queries are also given. These notes are a work in progress and will be revised as issues become known or are resolved. If you come across any new issues, "features", or tips not listed below, please let us know.

13.1 General

1. The ZSDS archive contains products with basic *automated* data quality assurance applied during pipeline processing. These are products where astrometric and photometric calibration was possible according to thresholds defined by metrics in the INFOBITS string (§10.4). This yields potentially usable products, however, "bad" data is still possible. This may be data affected by clouds, low atmospheric transparency, moon contamination, bright source scattering artifacts, unmasked aircraft and artificial satellites, transient detector behavior, and/or inaccurate telescope tracking. The archive metadata contains flags for identifying and omitting bad quality data. For images and accompanying source catalog files derived therefrom, the appropriate metric is INFOBITS. For lightcurve measurements (on a per-epoch basis), the metric is *catflags*. These are described further in §13.4 and §13.6 respectively.

2. The date/timestamps attached to all file-based products (i.e., values for the FITS header keywords: OBSJD, OBSMJD, SHUTOPEN, DATE–OBS) pertain to the *start* of the exposure, not the middle. The timestamps attached to the lightcurve photometry measurements however (*hjd, and mjd*) pertain to the *middle* of each exposure, i.e., with 0.5*EXPTIME added to the start observation times.

3. The start of an exposure, as defined by the start of photon integration (keywords in note 2 above) pertains to the approximate N-S centerline of the focal plane, i.e., parallel to the shutter blades. Due to the finite time it takes for the shutter to open and close, the actual total integration time depends on a source's position in the focal plane. The effective exposure time is smaller by at most 60 millisec at the E-W edges of the focal plane relative to its center. This decrease towards the edges translates to a relative reduction in photometric throughput of $<\sim 0.2\%$. This spatial variation is implicitly caught by the photometric zeropoint calibration at the CCD-quadrant level.

4. The EXPTIME (exposure or integration time) keyword value is always 30 seconds for public (MSIP) exposures. This is also the nominal integration time for private and Caltech-time observations, however, for some science programs, it can be 60, 90, 300, or 600 sec.

5. The root filenames and identifiers assigned to the single epoch-based file products are described in §7.1. The fractional time of day represented in a filename (i.e., the *dddddd* in ztf_<YYYYMMDD*dddddd*>…) is UT-based. E.g., 176377 => 0.176377 day => 04h:13m:58.97s UT. It's important to note that these times are typically one second earlier than the start-observation timestamps in the FITS headers (e.g., OBSJD). This difference is

due to the asynchronous FITS-keyword/telemetry computation and image write time at the P48.

13.2 Raw Image Files

1. The raw CCD data are packaged as Rice-(*fpack*-)compressed multi-extension FITS files (see §6.2). There is one file per CCD, with 16 CCD files per exposure of the focal plane. There are nine FITS extensions: the first extension is a global header; the next four contain the CCD photon-sensitive quadrant pixel data (or readout-channels); the last four contain the over-scan (bias) strips for each respective CCD-quadrant.

2. The photon-sensitive CCD-quadrant pixel data in the raw CCD files *have not* been over-scan corrected using their corresponding over-scan data. There are no other raw intermediate products in the ZSDS where the raw CCD data have been over-scan (bias) corrected.

13.3 Single-Exposure (Epochal) Image products

1. Beware of unmasked artifacts in all images, i.e., not flagged in the accompanying bit-mask images (§10.3). The detection and masking of optical and detector artifacts was performed on a best effort basis. There is no guarantee that all flavors of ghosts, cross-talk, bright-star CCD-bleed artifacts and bad-pixels were tagged. Their presence will have also contaminated the photometry catalogs (§13.4), but not necessarily alert packets generated from difference-images (§13.8) since for these, other *source*-based metrics were used to mitigate artifacts following detection. The machine-learned reliability score also provides a metric for filtering alerts triggered from artifacts.

2. Photometric calibration accuracy (estimates of zeropoints and color coefficients with respect to PS1; §6.5.2) depends on many variables, e.g., seeing, airmass, non-uniform transparency due to intermittent/partial cloud cover, background level, CCD/detector-noise, scattering from the moon and other bright objects, and level of source confusion. There is one estimate of MAGZP and CLRCOEFF *per CCD-quadrant*. Spatial variations in transparency at the intra-quadrant level will lead to errant and often unusable science products. This includes difference images and alerts derived/triggered therefrom. See note 2 in §13.4 for instructions on how to identify and omit bad quality data from your queries.

3. In the Public Data Releases, a difference image (file suffix *scimrefdiffimg.fits.fz*) will only accompany a single-exposure science image in the archive if a reference image corresponding to the same field/CCD-quadrant/filter *existed* on/before a single-exposure image was processed. Reference image generation continued throughout 2018 and 2019 and image differencing (with alert generation) could only be triggered for fields/CCD-quadrants/filters where a survey-quality reference image was present in the archive. There was no attempt to generate difference images for historical epochs. We expect future data releases to be more complete in terms of the difference images across epochs.

13.4 Single-Exposure (Epochal) Photometry Catalogs

1. Only non-saturated sources are detected and measured, both during PSF-fitting and aperture photometry. The saturation limits lie within ~ 12.5 to 13.2 mag for all filters and CCD-quadrants. The limits are fuzzy due to (i) electronic detector-gain variations and (ii) a given star will fluctuate between being saturated in its core when the seeing is good and not saturated when the seeing is bad. The limit of 12.5 mag is typically for quadrants near the center of the focal plane and with a derived image quality of ~ 2.0 arcsec.

2. The ZTF Observing System is a fully robotic system. One of the drawbacks is that it cannot determine in advance when conditions are non-photometric or the atmospheric transparency is very low, for example, when there are intermittent clouds partially covering the FOV during an exposure. This severely affected the accuracy of the derived photometric calibration solutions. There is one estimate of MAGZP and CLRCOEFF *per CCD-quadrant.* Therefore, spatial variations in transparency at the intra-quadrant level will lead to errant and often unusable science products.

   "Bad-quality" or generally unusable CCD-quadrant-based images from individual epochs (including accompanying source catalog files) can be omitted when querying the archive by thresholding the INFOBITS value in the archive metadata. If INFOBITS for an image has value < 33554432 (i.e., does not contain bit 25), the image and catalog data are probably usable.

   Data flagged as bad/unusable were determined beforehand by examining metrics from processed CCD-quadrants for 42 survey fields covering different parts of the sky and observing conditions, and then deriving thresholds. The criteria used to set the bad-data quality flag (specifically bit 25 in the INFOBITS string) for all image/catalog file-based products in the Public Data Releases are:

   ```
   if fid=1 (g filter):
    magzp > 26.7 – 0.2*airmass OR
    magzprms > 0.06 OR
    nps1matches < 80 OR
    magzp < zp_thres[rcid] – 0.2*airmass

   if fid=2 (r filter):
    magzp > 26.65 – 0.15*airmass OR
    magzprms > 0.05 OR
    nps1matches < 120 OR
    magzp < zp_thres[rcid] – 0.15*airmass

   if fid=3 (i filter):
    magzp > 26.0 – 0.07*airmass OR
    magzprms > 0.06 OR
    nps1matches < 100 OR
    magzp < zp_thres[rcid] – 0.07*airmass
   ```
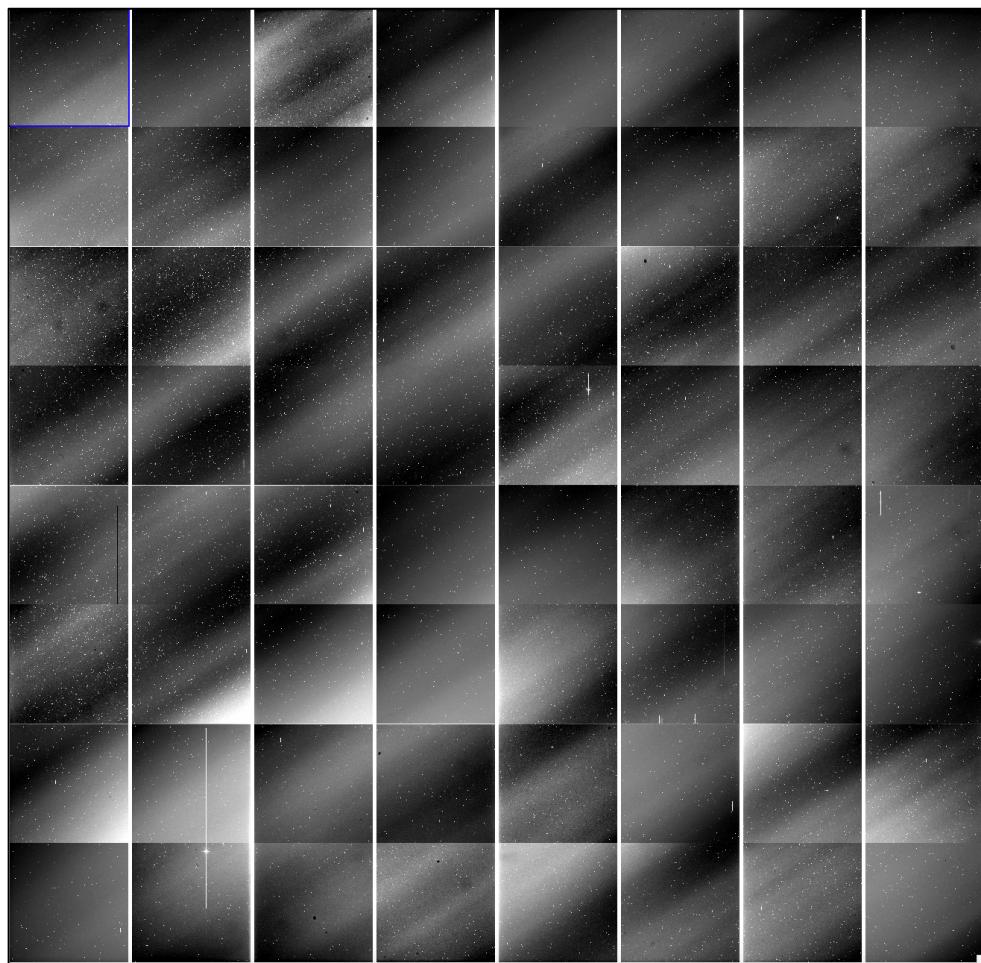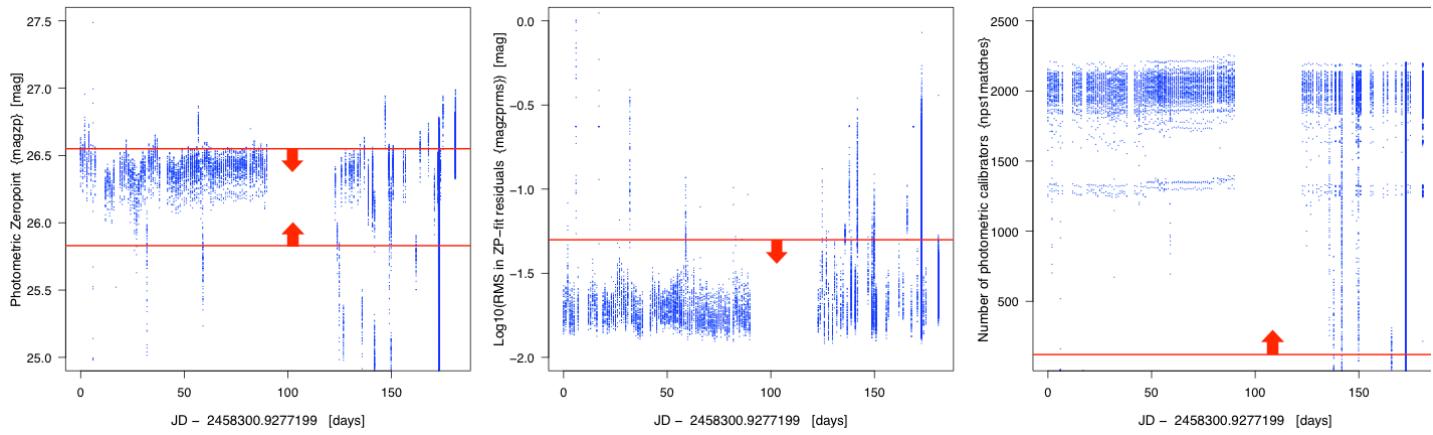
   zp_thres[*rcid*] means the threshold is CCD-quadrant dependent, where *rcid* = 0…63.

The `zp_thres[`*rcid*`]` values are listed in
http://web.ipac.caltech.edu/staff/fmasci/ztf/zp_thresholds_quadID.txt

To demonstrate these criteria, Figure 26 shows an example exposure with all 64 CCD-quadrants impacted by clouds that also happen to be reflecting moonlight onto the focal plane. This exposure is on galactic plane field 806 and was observed on 2018-12-21 UT. The photometry catalogs from these quadrants are not usable. Figure 27 shows trends in *magzp* (photometric zeropoint), *magzprms* (RMS in *magzp*-fit residuals), and *nps1matches* (number of PS1 calibrators) for the same survey field. The scatter at an individual observation JD represents the dispersion across all 64 CCD-quadrants of an exposure. There is also intra-night scatter at this resolution. The *magzp* values shown here were corrected for airmass variations (to a fixed airmass of 1) and for possible variable integration times (to the nominal 30 sec) using $magzp + 2.5\log_{10}(\text{AIRMASS}) - 2.5\log_{10}(\text{EXPTIME}/30)$. The bad exposure in Figure 26 is at relative *obsJD* ~ 174 in Figure 27 and is indicated by a huge scatter in the metrics. The red horizontal lines are thresholds derived from the above flagging criteria. The usable *magzp* range is conservative in the sense that the tightest "good-data" *magzp* range across all CCD-quadrants is shown.

Figure 26: example exposure (one epoch on survey field 806) covered by intermittent clouds that are also reflecting moonlight onto the focal plane. This exposure resides at relative *obsJD* ~ 174 in Figure 27. Sky north is up and east is to the right.

Figure 27: trends in *magzp*, *magzprms*, and *nps1matches* per CCD quadrant in *r*-filter exposures for survey field 806 from 2018-07-01 to 2018-12-29. Red lines with arrows indicate approximate thresholds for "good" (usable) image data. Data outside these limits are flagged in the archive metadata. See cautionary note (2) in Section 3.4.

3. Varying atmospheric transparency on the scale of an individual CCD-quadrant can have a significant impact on the quality of the *i-filter* fringe corrections. Fringe corrections were calibrated from high throughput images acquired under photometric conditions. The presence of intermittent cloud cover, depending on its spatial variation, can lead to erroneous corrections and unusable images.

4. The INFOBITS value stored in the archive metadata database (used to flag "bad-quality" data; see note in red text in note 2 above) is not always consistent with that present in the FITS header of a single-epoch image. INFOBITS in the FITS headers do not record bit 25, – the bad-data quality flag. This bit was assigned following archival analysis. All other bits (see §10.4 for definitions) are consistent between FITS headers and records in the archive metadata database.

5. The aperture and PSF-fit based single-epoch catalog files (file suffixes *sexcat.fits* and *psfcat.fits* respectively) were constructed independent of each other. I.e., the catalogs are not merged according to position. The PSF-fit catalogs usually contain more sources due to de-blending during PSF-fitting, particularly in crowded regions.

6. Comparisons between calibrated ZTF photometry and PS1 on a CCD-quadrant-basis show biases in the PSF-fit fluxes of up to 0.02 mag for predominately bright sources with *g, r, i,* $<\sim$ 15.5 mag (§12.4). These biases are not yet understood. They are also field-dependent and whether there was significant contamination by the moon.

7. Similar photometric biases are seen in the aperture-based catalogs with respect to PS1 (§12.4). As expected, biases in the aperture photometry are typically larger in high source-density fields across the full flux range due to contamination from source crowding. PSF-fitting is more immune to such contamination.

8. Photometric uncertainties in the PSF-fit based single-epoch catalog files (file suffix *psfcat.fits*) *can be* overestimated by factors of up to 3 relative to estimates of the RMS in photometric repeatability using non-variable (noisy) source populations. This overestimation is a function of source flux and is generally larger for brighter sources. These uncertainties have been corrected in the final lightcurve photometry following source-matching across the single epoch catalog files (see note 3 in §13.6).

9. Bad photometric quality can also arise from inaccurate over-scan (bias) subtractions during processing of CCD-quadrant images. The presence of bright/saturating stars near or on an over-scan boundary can lead to anomalous bias-level estimates as a function of pixel position and therefore impact the entire image and photometry derived therefrom. See note 10 in §13.6 for a method to avoid selecting lightcurve data that may have been impacted by this effect. A fix was implemented to mitigate this effect on February 21, 2020, following the DR3 data cutoff of Dec 31, 2019.

13.5 Reference Images (co-adds) and their Source Catalogs

1. Reference images are co-adds of individual epochal images on a CCD-quadrant/filter basis that have been selected to satisfy the quality criteria outlined in §6.7 (step 1 therein). Their depth varies from 15 (minimum) to 40 single-epoch images (maximum).

2. Epochal images from **all** survey programs (public, private, and Caltech-time) that satisfy the quality criteria (§6.7) are used to construct reference images. This means there could be a mixture of input integration times (30, 60, 90, 300 or 600 sec). The total effective integration time is indicated by the TOTEXPT keyword in the reference image FITS headers.

3. Only non-saturated sources are detected and measured for inclusion in the reference image photometry catalogs, both during PSF-fitting and aperture photometry. The saturation limits lie within ~ 12.5 to 13.2 mag for all filters and CCD-quadrants, similar to those in the single epoch images (§13.4).

4. Photometric calibration solutions are not explicitly derived for reference image products. The single-epoch photometric solutions are propagated to a reference image product by rescaling the input single-epoch images (pixel values) to a common photometric zeropoint before co-addition (see §6.7). Currently, the fixed reference image magnitude zeropoints (MAGZP) for *g, R, i* filters are MAGZP = 26.325, 26.275, 25.660 respectively. I.e., the reference image products are not recalibrated against any external catalog.

5. The following four metrics residing in the reference-image FITS headers are derived from a median of the corresponding input single-epoch image values: MAGZPUNC, NMATCHES, CLRCOEFF, and CLRCOUNC. The MAGZPRMS metric for reference-image products represents a robust RMS (based on the MAD) of the input single-epoch MAGZP values.

6. The aperture and PSF-fit based reference image catalog products (file suffixes *refsexcat.fits* and *refpsfcat.fits* respectively) were constructed independent of each other. I.e., the catalogs are not merged according to position. The PSF-fit catalogs usually contain more sources due to de-blending during PSF-fitting, particularly in crowded regions.

7. Comparisons between calibrated ZTF photometry and PS1 on a CCD-quadrant-basis show biases in the PSF-fit fluxes of up to 0.02 mag for predominately bright sources with *g, r, i*, $<\sim$ 15.5 mag. These biases are not yet understood. They are also field-dependent.

8. Similar photometric biases are seen in the aperture-based catalogs with respect to PS1. As expected, biases in the aperture photometry are typically larger in high source-density fields across the full flux range due to contamination from source crowding. PSF-fitting is more immune to such contamination.

13.6 Lightcurves ingested from Source-Matchfile products

Access to the lightcurve query service was described in §7.6. Table columns and metrics returned by this service are defined in §18. *Matchfile* (lightcurve) generation was described in §6.8.

1. Lightcurves are generated per field, CCD-quadrant, and filter. Due to overlapping fields/quadrants on the sky-tiling grid, it is likely the query service will return multiple lightcurves for the same object. There is no merging of same-object lightcurves corresponding to the same filter. You will need to splice them after retrieving them. Furthermore, the different fields/quadrants will likely have different photometric calibration solutions with respect to PS1. The final lightcurve photometry however is calibrated on the same system.

2. Lightcurves are constructed using seed detections from the deeper (co-add) Reference Image PSF-fit catalogs. These "seed sources" are loaded into the searchable Objects Table (§13.7) where their photometry is $\sim$ 2 to 2.5 magnitudes deeper than the single-exposure catalogs. Therefore, a majority of faint sources in the Objects Table (fainter than the single-exposure sensitivity limits) will not have lightcurves. Futhermore, objects with photometry hovering near the single-exposure detection limits will have incomplete temporal-coverage and therefore more sparsely sampled lightcurves.

3. The primary lightcurve photometry is represented by column *mag* with 1σ uncertainty: ± *magerr*. The *mag* values already have their epoch-dependent photometric zeropoints applied (column: *magzp*). By default, all lightcurve photometry is in the native ZTF photometric system. No color corrections are applied (or rather, zero color in the *AB* system is assumed). To transform your photometry onto the *AB* (PS1) system, you will require knowledge of the source color in that system. The exact color used during initial photometric calibration against PS1 is defined by column *pcolor*, the actual color coefficient to enable transformation to the PS1 system is given by column *clrcoeff*, and the procedure for applying color corrections was described in §10.1.1.

4. The magnitude uncertainties *magerr* include empirically-derived corrections so they are overall consistent with the temporal RMS inferred from the photometric repeatability in stationary (noisy) lightcurves at the given *mag*. For details, see §6.8.1.

5. Analogous to the flagging of "bad-quality" images and associated catalogs (§3.4), bad or generally unusable observation epochs in lightcurves can be omitted by thresholding the *catflags* column. If *catflags* for an image has value < 32768 (i.e., does not contain bit 15), the photometry at that epoch is probably usable. This flagging removes epochs based on their overall image/calibration quality. The *catflags* value also encodes possible source-specific issues, for example, bad-pixel information propagated from the input catalog *flags* defined in §10.6. If you demand perfectly clean extractions, we advise specifying *catflags* = 0 instead when querying lightcurve epochs.

6. Related to the previous note, the number "good" observation epochs in a lightcurve, i.e., with *catflags* = 0 is given by the *ngoodobs* metric. The total number of epochs is *nobs*. Analogous to these counters is the actual number of epochs in a public release (usually a subset of all epochs). These numbers are given by *nobsrel* (all epochs, regardless of quality) or *ngoodobsrel* (with *catflags* = 0).

7. As of April 8, 2019, relative photometric refinement is turned off during lightcurve generation. Analyses have shown that this post-processing step can give erratic ZP corrections, usually caused by occasional bad (unflagged) data going into the solver, which was difficult to track.

8. A weak but significant periodic signal with period ~ 28 to 29 days is seen in the lightcurves of faint sources with *g* or *r* >~ 20.5 mag (close to the 5σ detection single exposure limit). This no doubt is due to background contamination by the moon and is a function of its distance. The local background estimation during PSF-fit photometry is relatively robust against linearly-varying background gradients, but can be biased if the background has a large variation or is significantly non-linear at high spatial frequencies. Furthermore, if you are applying color corrections to obtain photometry in the *AB* system (note 3 above), contamination by the moon may have biased the color of your source due to residuals from inconsistent background subtraction across filters.

9. Lightcurve photometry and all associated products are not replaced if there is any reprocessing or replacement of the reference image and/or single-epoch catalog products in the archive. Publicly-released lightcurve products are frozen at release time and never updated. Any updates to the inputs will propagate into future public releases of the lightcurve database.

10. To additionally avoid possibly corrupt lightcurve measurements (e.g., due to unmasked ghosts or inaccurate over-scan subtractions as described in note 9 of §13.4), one can threshold on the *chi* metric, which is available on a per-epoch basis, but only from the lightcurve GUI service (not lightcurve tarballs). A recommended threshold for eliminating

bad photometric measurements is *chi* > 4. This would be in addition to any *catflags* filtering (note 5 above).

13.7 Objects Database Table to support Lightcurve Queries

1. As described in §7.6, access to the lightcurve database first entails performing a spatial-search (combined with other optional filters) on the ZTF Objects Table. These are all the objects detected and extracted (with PSF-fit photometry) from the reference images, i.e., those that reside in the CCD-quadrant-based reference image catalog files (§13.5). The cone search for a single query is currently limited to a radius of 10 arcmin.

2. Currently, the ZTF Objects Table does not contain flux transients, that is, single epoch detections with no counterpart in a reference image. Only sources detected in the reference images are represented. The reason is that a large fraction of the transients appear to be associated with image artifacts, detections falling off the overlapping reference image footprint, and blended detections with ambiguous (non-unique) counterparts in the deeper, better quality reference images. The alert packets (§9 and §13.8) are the primary products from which transients should be mined.

3. The Objects Table also contains a selection of collapsed-lightcurve metrics. Source data spanning all observation epochs from **all** survey programs (public, private, and Caltech-time) are used to compute these metrics. The total number of epochs used to compute these metrics is given by *nobs*. Publicly-released lighturves however contain a subset of the measurements are their number is indicated by *nobsrel* which is typically < *nobs*. Any lightcurve-collapsed metric can be included and filtered as part of your spatial query.

4. Photometry and metrics in the Object Table are not replaced if there is reprocessing or replacement of the reference image and/or single-epoch catalog products in the archive following initial source matching to generate lightcurves. Publicly-released lightcurve products are frozen at release time and never updated. Updates to the inputs will propagate into future public releases of the lightcurve database.

13.8 Alert Packets

The contents and structure of the alert packets were described in Section 9. Below are the known issues and behaviors to be aware of when using these products.

1. All packets have been (lightly) filtered to remove obvious false-positives using the metrics therein (see §9.1). Therefore, some of the metrics will have limits set by the thresholds used in this filtering. Be aware that not all packets correspond to reliable events. Low-level instrumental calibration residuals and anomalies in the image-subtraction process still lurk.

2. The differential photometry (*magpsf* or *magap*) can sometimes exhibit significant intra-night scatter (larger than the quoted errors) when constructing lightcurves from either the photometric history in each packet, or by merging packets corresponding to the same

objectId. We believe this is due to systematics that need to be corrected upstream. We appreciate such cases be reported to the Data System team.

3. If the number of bad pixels in a 5x5 pixel region (*nbad*) is non-zero, the *magpsf* is more likely to be erroneous. In such cases, we force *magpsf = magap*. The *magap* is likely to yield a more accurate measurement due to internal bad-pixel interpolation used during the aperture photometry step. This will be revised when the issue in (2) is addressed.

4. When merging packets to generate lightcurves for a given *objectId* (from both detections and associated photometric histories, including upper limits), it is possible for some packets to have upper-limits in their histories at a specific observation epoch and other packets to have bona fide detections at the *same* epoch. This is due to the finite matching radius (1.5 arcsec) used to (i) assign the initial *objectId* and (ii) construct the photometric history in each individual packet. That is, each alert packet's history is constructed using a positional match to the *detected* event that triggered it, independent of previous alerts. Astrometric and centroiding error can cause positional discrepancies across packets. We advise only reporting photometric upper limits at a given epoch in a lightcurve if detections from other packets don't already exist.

5. Due to astrometric and centroiding error, alerts from the same astrophysical object may not receive the same *objectId* if their measured positions hover around the 1.5 arcsec matching radius used for *objectId* assignment or reuse across epochs. The important point is that the *nearest* previously assigned *objectId* (if found) is reused. It is therefore possible for two very close alerts (detected at different epochs) being assigned different *objectIds* according to this rule (see §9.4 for details).

6. Multiple alerts detected within 1.5 arcsec of each other on the *same image* (observation epoch) may lead to all/some being assigned the same *objectId*. This would be the *objectId* of a previously named (historic) alert if found within 1.5 arcsec (see §9.4 for details). If no previously named alert exist, each alert would acquire a unique *objectId* for the first time. Matching to future alerts (if they reoccur close to these earlier alerts) will then lead to ambiguities in the assignment of *objectId,* again due to astrometric and centroiding error. These ambiguities are more likely to occur in high source density regions, e.g., the galactic plane.

7. The PSF-fit differential photometry estimates (metrics *magpsf* and *sigmapsf* in the alert packets) cannot easily be reconstructed using the available image products (difference image and accompanying PSF from the archive). These estimates are computed from other internally generated products from ZOGY (see §6.6). The intent here was to reduce photometric scatter early in the project (see note 2 above) without disrupting existing image products from which all other metrics are computed and used to train and update the *RealBogus* classifiers.

8. Beware when interpreting the lightcurves constructed from merging the differential photometry from multiple alert packets for an *objectId*. There are instances where an object

is covered by different overlapping survey fields and detector quadrants. As a reminder, each field/quadrant is treated independently in the ZSDS, where each has its own reference image, possibly constructed from different input-image epochs. There is no guarantee that the reference images are not contaminated (by different amounts) by the transient or variable source flux sought. This means that the differential photometry can show significant scatter if packets originating from different fields/quadrants (with different "reference baselines") are combined.

9. It can be a challenge to ascertain the reliability of a single alert using its packet metadata alone, including the supplied *rb* or *drb* scores. Unless the alert is associated with a moving object (as indicated in the metadata), one recommendation is to perform a spatial cross-match with earlier or later alerts (if available) and construct a lightcurve. Temporal correlation is a powerful way to determine if a collection of alerts are real (astrophysical) or not.

10. The *sgscore* (star/galaxy separation score) metrics attached to the nearest PS1-catalog associations have at times found to be unreliable, most notably for events from variable stars. This is due to issues with the original PS1 photometry catalogs since these were used to train the machine-learned star/galaxy classifiers outside the ZSDS.

11. All alert packets generated per CCD-quadrant subtraction image are archived at the ZSDS and stored in gzipped tar files (generic suffix: *_alerts.tar.gz*). These can be accessed from IRSA using the same GUI and API tools for querying the single epoch CCD-quadrant image products (§7.6).

12. The reliability of an alert or its photometric accuracy strongly depends on the calibration quality of the CCD-quadrant used during image subtraction. Photometric calibration accuracy (estimates of zeropoints and color coefficients with respect to PS1; §6.5.2) depends on many variables, e.g., seeing, airmass, non-uniform transparency due to intermittent/partial cloud cover, background level, CCD/detector-noise, scattering from the moon, and level of source confusion. There is one estimate of MAGZP and CLRCOEFF *per CCD-quadrant*. Spatial variations in transparency at the intra-quadrant level will lead to unreliable and often unusable alerts extracted from the quadrant subtraction image. See note 2 in §13.4 for a method on how to identify and omit all alerts triggered from a suspect badly-calibrated CCD-quadrant image.

13. The alert packets can also be combined to construct absolute photometric (DC) lightcurves, for example, for (a)periodic variable sources. Such lightcurves can be computed by combining the nearest reference image catalog magnitude (*magnr*), differential magnitude (*magpsf*), and *isdiffpos* (positive or negative difference image detection) as follows:
    $$mag_{abs} = -2.5\log_{10}(10^{[-0.4magnr]} + sign\,10^{[-0.4magpsf]}),$$
    where
    $sign = 1$ if *isdiffpos* = 't' or $sign = -1$ if *isdiffpos* = 'f'.
    Before using the nearest reference image source magnitude (*magnr*), you will need to ensure the source is close enough to be considered an association (e.g., *distnr* <~ 1.5 arcsec). It is

also advised you check the other associated metrics (*chinr* and/or *sharpnr*) to ensure it is a point source. These metrics are defined in §10.6. We recommend $0.5 <\sim chinr <\sim 1.5$ and/or $-0.5 <\sim sharpnr <\sim 0.5$.

14. NaN'd pixel values in the alert packet image cutouts are an indicator of bad (unusable) pixels. These NaNs were propagated from upstream processing and are also present in the archived difference image products. Bad (fatal) pixels identified in upstream processing are intentionally forced to NaN prior to differencing. Specifically, the bad pixels set to NaN are those tagged with bits 0,2,3,5,8,9,10,12 in the science images and are defined in §10.3. Various convolutions and interpolations during the PSF-matching process will have spatially-expanded these bad-pixel regions. NaNs were the simplest way to propagate and track this "badness".

15. Alert packets are generated close to realtime typically within 20 minutes of a camera exposure, using filtered events extracted from difference images. Image differencing requires the presence of a reference image. Therefore, alerts can only be generated if a reference image corresponding to the same field/CCD-quadrant/filter *existed* on/before a single-exposure image was processed. Reference image generation continued throughout the survey and therefore there are regions of sky where no alerts could be generated during the course of the survey.

16. A note of caution on the *ndethist* metric in alert packets. This metric is not related to the contents of the *prv_candidates* block in a packet. This metric enumerates every associated detection (within 1.5 arcsec of the target alert's position) from an *internal* candidates database down to S/N ~ 3. Only historical detections falling on the same survey field and CCD-quadrant from which the target alert was extracted are counted. The S/N here is shown as an approximation (~ 3) since it is based on a point source match-filtered estimate (peak S/N estimate, a proxy for the photometric S/N). This should not be confused with associations in the *prv_candidates* block which contain detections down to a *photometric* S/N cut of exactly 5 (precisely 1.0857/*sigmapsf*).

14 REFERENCES

Bertin, E. et al., 2002, ASP Conf. Series, 281, 228

Denneau L. et al. 2013, PASP, 125, 926, *http://iopscience.iop.org/article/10.1086/670337/meta*

Flewelling, H. A. et al., 2017, *https://arxiv.org/abs/1612.05243*

Høg, E. et al., 2000, A&A, 355, L27

Laher, R. R., et al., 2014, *https://arxiv.org/abs/1404.1953*

Levitan, D. B., et al., 2011, *http://thesis.library.caltech.edu/7857/1/Levitan_PhD_Thesis.pdf*

Masci, F. J. et al., 2017, *http://iopscience.iop.org/article/10.1088/1538-3873/129/971/014002/meta*

Masci, F. J. et al., 2018, *http://iopscience.iop.org/article/10.1088/1538-3873/aae8ac/meta*

Ofek, E., O., et al., 2011, *http://iopscience.iop.org/article/10.1088/0004-637X/740/2/65/pdf*

Rest, A., 2017, *https://confluence.stsci.edu/display/PANSTARRS/How+to+separate+stars+and+galaxies*

Waszczak, A., et al., 2017, *https://arxiv.org/abs/1609.08018*

Zackay, B., et al., 2016, *https://arxiv.org/abs/1601.02655*

15 ACRONYMS

ADES	Astrometry Data Exchange Standard
ADC	Analog to Digital Conversion
ADU	Analog Digital Units
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
CCD	Charge Coupled Device
CCDID	CCD Identifier: 01…16.
CPU	Central Processing Unit
DAOPhot	Dominion Astrophysical Observatory Photometry tool
DB	Database
DN	Data Number (analogous to ADU)
DR	Data Release
FITS	Flexible Image Transport System
FOV	Field-Of-View
FWHM	Full Width at Half Maximum
GUI	Graphical User Interface
HDF5	Hierarchical Data Formatted file
HDU	Header Data Unit (or FITS image extension)
I/O	Input / Output
IAU	International Astronomical Union
IPAC	Formerly referred to as the Infrared Processing and Analysis Center
iPTF	*intermediate* Palomar Transient Factory
IRSA	InfraRed Science Archive
JPEG	Joint Photographic Experts Group image format
LC	Lightcurve
LCDB	Lightcurve Database
LDAC	Leiden Data Analysis Center
LED	Light Emitting Diode
MAD	Median Absolute Deviation: $\sim 1.4826\,med\{\,|x_i - med\{x_i\}|\,\}$
MBA	Main Belt Object
MEF	Multi-Extension FITS file format
ML	Machine Learning (or Machine-Learned)
MW	Microwave data link
MODE	Moving Object Discovery Engine
MOPS	Moving Object Pipeline System
MOST	Moving Object Search Tool
MPC	Minor Planet Center
MSIP	Mid-Scale Innovations Program
NEA	Near-Earth Asteroid
NEO	Near-Earth Object
P48	Palomar 48 inch
PDT	Pacific Daylight-savings Time (UT − 7 hours)
PNG	Portable Network Graphics image format
PS1	Pan-STARRS 1 Data Release products

PSF	Point Spread Function
PST	Pacific Standard Time (UT − 8 hours)
PT	Pacific Time, regardless of time of year (i.e., PDT or PST)
PTF	Palomar Transient Factory
PTFIDE	PTF Image Differencing & Extraction pipeline
QA	Quality Assurance
QID	Quadrant Identifier: 1…4.
RB	RealBogus (moniker for machine-learned vetting)
RCID	Readout-Channel Identifier: 00…63.
RMS	Root-Mean-Square deviation
S/N	Signal-to-Noise ratio
SDSC	San Diego Supercomputing Center
SDSS	Sloan Digital Sky Survey
SED	Spectral Energy Distribution
SExtractor	Source Extraction tool
StdDev	Standard Deviation
SV	Science Validation
SVB	Slowly Varying Background (image)
TBC	To Be Confirmed
TBD	To Be Determined
TBR	To Be Resolved
TBW	To Be Written
ToO	Target of Opportunity (e.g., in context of a LIGO event or Fermi burst)
UDB	User's Database (replication of Operations Database)
URL	Uniform Resource Locator
UT	Universal Time
UW	University of Washington
VO	Virtual Observatory
VPO	Virtual Pipeline Operator
WCS	World Coordinate System
XML	Extensible Markup Language
ZADS	ZTF Alert Distribution System
ZMODE	ZTF Moving Object Discovery Engine
ZOGY	Zackay, Ofek, & Gal-Yam image-subtraction algorithm
ZP	Zero-Point (for photometric calibration in terms of magnitudes)
ZPVM	Zero-Point Variations Map generated from absolute calibration
ZSDC	ZTF Science Data Center
ZSDS	ZTF Science Data System
ZTF	Zwicky Transient Facility

16 APPENDIX A: ZMODE – ZTF's Moving Object Discovery Engine

ZMODE was initially developed at IPAC for the Palomar Transient Factory. It was developed for use on extractions from difference images, with provisions for scaling up to meet ZTF's computational demands by leveraging new hardware architectures. It uses a new algorithm, related to ideas presented in Waszczak et al. (2013; MNRAS, 433, 3115). From studies of the recovery fraction of known objects in PTF data (i.e., with well known orbits), ZMODE can detect moving objects at a completeness (efficiency) of > 90% and a reliability (accuracy) of > 98% to moderately faint flux levels ($R_{PTF} \sim 20$ mag).

The most significant difference between ZMODE and the classic MOPS implementation (Denneau et al. 2013) is the first step in linking candidate moving object detections into building blocks of candidate tracks. ZMODE requires a minimum of three detections to form a moving-object stringlet, whereas MOPS starts out by forming two-detection "tuples". ZMODE forms stringlets by matching the relative velocities of two adjacent pairs of detections that share a common middle detection. A schematic is shown in Figure A1. Their velocities are matched within some tolerance that depends on the time separation of the detections. The stringlet-construction step imposes two constraints on the time-separation of detections, i.e., their observation epochs: (i) one of the detection pairs in the stringlet must span ≤ 10 hours (i.e., the same night); and (ii) the other pair must span ≤ 2.5 days. Criterion (ii) is optimal for input data spanning four consecutive nights. Four consecutive nights is driven by ZTF's typical observing cadence and a minimum track length (number of detections) of four to declare a candidate track for vetting (see below).

After all possible stringlets in the detection stream have been identified, they are linked using velocity matching on a coarser grid to create moving-object candidate tracks. Two types of velocities are used in the final stringlet-merging step: (i) the mean velocity of the (intra-stringlet) detection-pairs; and (ii) the relative (inter-stringlet) velocities between the average position of detections in each of the stringlets. The candidate tracks from ZMODE are then vetted using a number of quality metrics, including an orbit fit to test if the candidate track is dynamically plausible (see below).
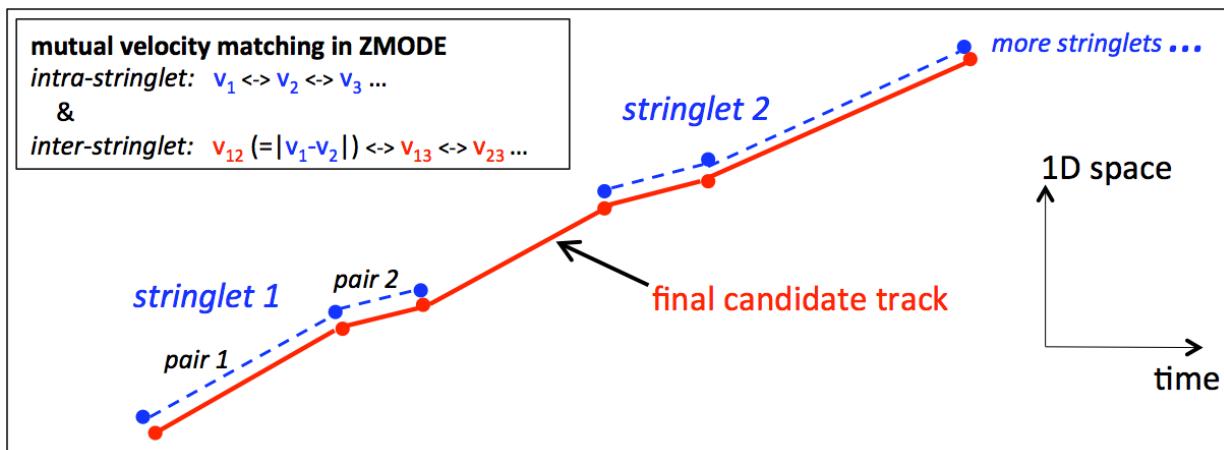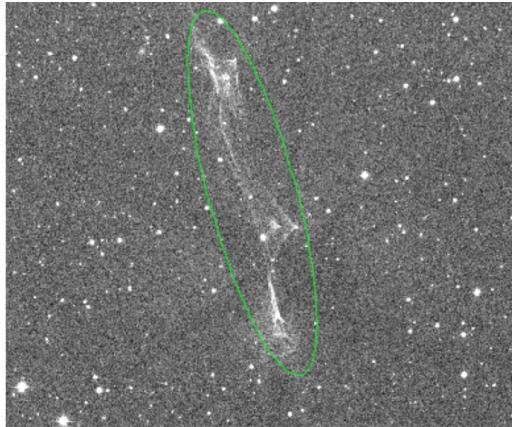
Figure A1: Schematic showing the construction of moving-object tracks in ZMODE.

By building three-detection stringlets, ZMODE moves part of the combinatorial challenge earlier in the processing. It helps eliminate spurious two-detection "tuples" that would otherwise add to the load at later stages of candidate track construction. Moreover, it significantly reduces the number of possible combinations that need to be processed and merged downstream, as opposed to MOPS which carries along all possible two-detection "tuples". The ZMODE design is therefore more efficient, for example, in regions of high source density for a given input timespan.

To reduce the incidence of false linkages, we automatically identify reliable tracks using orbit fitting. We find that tracks consisting of four or more linked detections accurately fitting a dynamically valid orbit constitute worthy candidates for further inspection. The orbit fitting is based on the *Find_Orb* orbit-fitting software, which has been adapted and tuned to process candidate tracks from ZMODE. *Find_Orb* generates an orbit quality metric (or score) along with an RMS of the fit residuals. These two metrics are thresholded to retain good quality tracks for both MBAs and NEOs.

17 APPENDIX B: PREDICTING AND MASKING COUNTER-MOVING GHOSTS

This section expands on the prediction and masking of *counter*-moving ghosts in the instrumental calibration pipeline (step 16 in Section 6.5). A counter-moving ghost is a "tendril"-shaped artifact caused by a bright source (the parent) falling elsewhere on the focal plane, usually on the opposite side of the telescope optical axis. All these ghosts have approximately the same shape. Figure B1 shows an example of a typical ghost and its elliptical masking region.

Figure B1: A ghost from a V=3 mag parent star.

In all cases, the CCD quadrant containing a purported ghost can be found from just knowing the CCD quadrant containing the parent star. The ghosts always reflect across the approximate center (optical axis) of the focal plane. For example, referring to Figure 1 in Section 3, a bright parent in CCD1,q4 will produce a ghost in CCD16,q2 and vice versa.

In the description below, let C_g, Q_g denote respectively the CCD and quadrant ID therein containing the purported ghost(s). This is also the quadrant that is being processed in the pipeline, in realtime. The corresponding CCD and quadrant therein that is searched for the parent (potentially ghost-inducing) stars is denoted C_p, Q_p. The steps used to predict and mask ghosts in C_g, Q_g are as follows:

1. Find the corresponding "parent image", i.e., to search for bright parent stars. According to the CCD and quadrant ordering in the focal plane (Figure 1 in Section 3), the parent CCD is given by $C_p = 17 - C_g$ and quadrant therein is $Q_p = 5 - Q_g + (-1)^{Q_g}$.
2. Determine the WCS parameters of the parent image (C_p, Q_p) by transforming the astrometric solution of the image being processed (C_g, Q_g). The astrometric solution for the latter is already available, but not necessarily for C_p, Q_p from the *same* exposure given the asynchronous nature of the quadrant-based parallel processing. The WCS mapping also uses a prior model for the FOV distortion, as well as prior calibrated pixel scales and relative offsets between all quadrants in the focal plane.
3. Use a subset of the Tycho-2 Catalogue of Bright Stars (Høg et al. 2000) to find which parents fall in the *x,y* coordinate frame of C_p, Q_p and store their positions x_p, y_p. This subset catalogue consists of stars with a *V*-band magnitude of ≤ 5.8 – the approximate limit where counter-moving ghosts start to appear above the background. Furthermore, these stars will
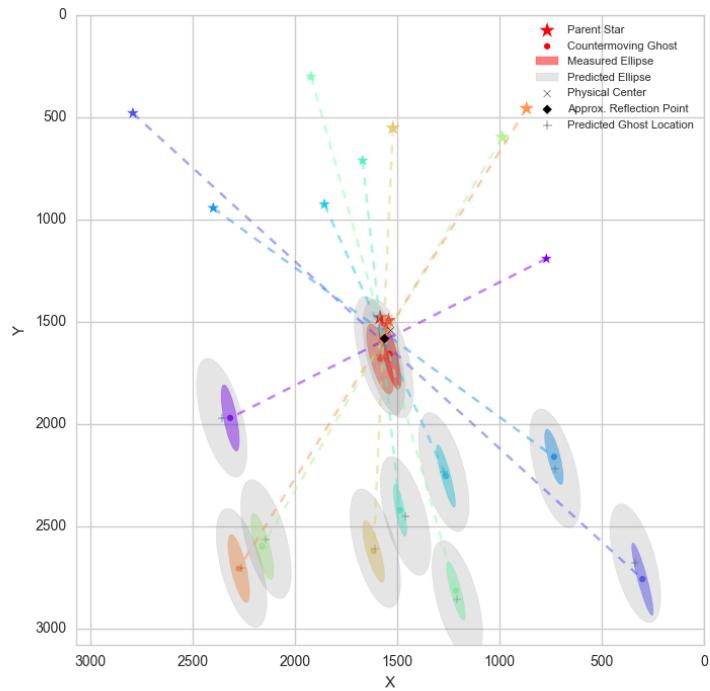
be heavily saturated in the nominal 30sec exposures and therefore not detected by the pipeline.

4. If ghost-inducing parents are found in C_p, Q_p, "pretend" that the ghost will appear in the same image but reflected across the fiducial point: $(x_r, y_r) = (1565, 1580)$. This point is represented by the diamond in Figure B2. The purported ghost position is given by: $(x_g, y_g) = (2x_r - x_p, 2y_r - y_p)$.

5. The ghost will appear at the position x_g, y_g derived in step 4 but in CCD quadrant C_g, Q_g.

6. Given the ghost position from 5, mask an elliptical region in the bit-mask image with center at this position; see Figure B2 for geometry. Each elliptical mask has a fixed conservative size: $a = 600$, $b = 200$ and rotated at a fixed angle of $+15°$ from the y-axis. The ellipse size is intentionally made large to accommodate uncertainties in the WCS/position predictions and parent-dependent variations in ghost morphology. In general however, these variations are very small. The bit-mask is updated with a special bit for counter-moving ghosts (see Section 10.3).

Currently, the ghost prediction and masking process has the following limitations:

- For parent stars falling in the corner quadrants of the focal plane, secondary ghosts are sometimes seen near the primary ghosts. Depending on where the parent falls, secondary ghosts are more difficult to predict and therefore are not masked.

- Ghosts from "parent" stars whose Tycho-2 catalog positions fall outside a focal-plane edge or within a CCD gap are not masked. Therefore, expect to see unmasked ghosts close to CCD edges.

- Due to inaccuracies in the predicted positions of ghosts in exposures observed at high *airmass* (primarily due to atmospheric refraction effects affecting the overall astrometric solution), there is no ghost-masking at *airmass* > 2.2. This is the approximate maximum *airmass* above which assumptions in the focal plane WCS used to map parents to ghosts start to show significant systematic residuals. As a reminder, *airmass* = 2 is the approximate survey limit.

- Related to the previous point, the predicted elliptically-masked centers can sometimes be off by several pixels. This is also due to approximations in the focal plane WCS when predicting ghost locations (see step 2 above). These offsets are difficult to characterize. The masked regions are made large enough to catch this uncertainty.

As mentioned in step 3 above, parent stars with magnitude $V \leq 5.8$ produce noticeable ghosts, and there are ~ 4000 such stars across the sky. If we assume they are uniformly distributed, we expect to see ~ 5 ghost-producing stars per exposure (full focal plane). Assuming our conservative mask size (see step 6), this translates to ~ 0.3% of the focal plane being masked on average. However, these stars are not uniformly distributed, but more likely concentrated near the galactic plane. If we ignore stars at $|b| \leq 10°$, the number of ghost-producing stars outside this latitude range is ~ 2700, which translates to ~ 0.2% of the focal plane being masked on average.

Figure B2: Schematic showing locations of parent stars and their counter-moving ghosts. The ghosts shown here *will not* actually appear in the same CCD quadrant as their parents. They would be masked in *another* quadrant reflected through the FOV center (optical axis).

18 APPENDIX C: *MATCHFILE* CONTENTS AND STRUCTURE

Matchfile (lightcurve) generation was described in Section 6.8. Here we summarize the contents of a *matchfile*. This file is an internal data-store for all lightcurves in each field/CCD-quadrant. A subset of the contents of these files can be accessed through the lightcurve query service (see §7.6).

FILE FORMAT

The ZTF matchfiles are stored in HDF5 format, and are readable and writable by the pytables Python package.

GROUPS WITHIN THE FILE

The HDF5 groups within each file is named "matches". For example, the hierarchy in a file is store.root.matches where store is the pytables file handle.

FILE LOCATIONS

The matchfiles are stored on the ZTF systems in subdirectories of /ztf/ops/srcmatch/, where the subdirectories take the form rcNN/frNNNNNN-NNNNNN similar to the reference images.

FILENAMING CONVENTION

The matchfiles follow the naming convention of reference images, ztf_NNNNNN_zC_cNN_qN_match.pytable where the capital items are field number, filter letter, chip number and quadrant number.

ATTRIBUTES OF THE MATCH GROUP

At the match group level, (e.g. store.root.matches), several attributes are recorded:

Attribute	Example Value
CLASS	'GROUP'
TITLE	'Matched Sources'
VERSION	'1.0'
calibratedSinceLastMatch	0
ccdID	7
datetimeCalibrated	b'2017-12-10 01:35:01.278087'
datetimeMatched	b'2017-12-09 01:15:01.278087'
fieldID	1674
filterID	2
goodSystematicErrorLimit	0.0099
nCalibrationSources	363
quadrantID	1
rcID	24

Attribute	Example Value
referenceImageID	16740224
referenceZeroPoint	27.0
referenceZeroPointErr	0.0
saturatedMagnitude	14.3

Three tables are attached to the match group:
  Exposures
  Sources
  Sourcedata

EXPOSURES TABLE

The exposures table contains one row per exposure.
The *expid* column is indexed with default pytables values.

Column Name	Data Type	Description
airmass	Float32	Airmass at approximately the center of the focal plane at time of exposure
clrcoeff	Float32	Color coefficient from linear fit
clrcounc	Float32	Color coefficient uncertainty from linear fit
clrmed	Float32	Median calibrator color [mag]
clrrms	Float32	RMS in calibrator color [mag]
diq	Float32	Derived image quality (FWHM) in arcseconds. Values of 0.0 indicate the PSF file could not be read.
expid	Int32	Exposure ID (DB counter)
exptime	Float32	Exposure time from scheduler [sec]
fieldid	UInt32	Field ID
filterid	UInt16	Filter ID
fracday	Uint64	YYYYMMDDdddddd Year, month, fractional day
gmedian	Float32	Median pixel (background) value of associated image [DN]
infobits	Int32	Quality indicators: see §10.4. Additionally, if *infobits* contains bit 25 (value $\geq$ 33554432), image at this epoch is likely unusable (e.g., affected by clouds and/or moonlight)
ipac_pub_date	Float32	IPAC publication date for this exposure, in modified Julian Date format
limitmag	Float32	Approximate 5σ limiting magnitude corresponding to epoch-based PSF-fit catalog [mag]

118

Column Name	Data Type	Description
magzp	Float32	Magnitude zeropoint from photometric calibration [mag]
magzprms	Float32	RMS deviation in magnitude zeropoint [mag]
magzpunc	Float32	Uncertainty in magnitude zeropoint [mag]
nps1matches	Int32	Number of PS1 photometric calibrators used
nsources	Int32	Number of *sourcedata* entries for this expid
nstars	Int32	nsources + ntransients
ntransients	Int32	Number of transientdata entries for this expid
obsmjd	Float64	Modified Julian date *at start* of observation [days]
obshjd	Float64	Heliocentric Julian date *at start* of observation and at commanded (scheduled) ra, dec of survey field center [days]
pcolor	StringCol(8)	Calibrator color used in photometric calibration, in PS1 system
programid	UInt16	Program Identifier
programpi	Char40	Program Principle Investigator
rcid	UInt8	Readout-channel ID (0-63)
zpclrcov	Float32	Covariance in magzp and clrcoeff [mag^2]
zpmed	Float32	Mag zero point from median of mag diffs [mag]

SOURCES TABLE

Each row in the sources table corresponds to an Object that appeared in the reference catalog that seeded the matchfile. Most of the columns are lightcurve statistics computed on the time series data in the sourcedata table. Note that several columns contain multidimensional arrays. The *matchid* column is indexed with the pytables defaults.

Column Name	Data Type	Description
astrometricrms	Float64Col	[degrees] Root Mean Squared deviation in epochal positions relative to the object's ra,dec; effectively the root-summed mean squared deviations along the ra and dec axes
chisq	Float32Col	Proxy for robust measure of reduced χ^2 metric using all nobs detections in lightcurve. Computed from ratio: $1.4826 medianabsdev / medmagerr$
con	Float32Col	fraction of nobs where three consecutive observations are more than 2x RMS from the median magnitude, plus 1
dec	Float64Col	Declination (deg)
lineartrend	Float32Col	[mag/day] Slope from linear fit to nobs

Column Name	Data Type	Description
		detections
magrms	Float32Col	[mag] Root Mean Squared deviation in nobs magnitudes
matchid	Int32Col	Index into the sourcedata table (for pulling out light curves)
maxmag	Float32Col	[mag] Maximum magnitude over nobs detections
maxslope	Float32Col	[mag/day] Maximum pairwise slope over all consecutive nobs detections
meanmag	Float32Col	[mag] Mean magnitude over nobs detections
medianabsdev	Float32Col	[mag] Median absolute deviation in magnitudes over nobs detections
medianmag	Float32Col	[mag] Median *mag* over nobs detections
medmagerr	Float32Col	[mag] Median of *magerr* values over nobs detections
minmag	Float32Col	[mag] Minimum magnitude over nobs detections
nabovemeanbystd	UInt16Col, shape=(3,)	Number of lightcurve points that are [1, 3, 5] standard deviations above the mean magnitude over nobs
nbelowmeanbystd	UInt16Col, shape=(3,)	Number of lightcurve points that are [1, 3, 5] standard deviations below the mean magnitude over nobs
nconsecabovemeanbystd	UInt16Col, shape=(3,)	Number of consecutive lightcurve points that are [1, 3, 5] standard deviations above the mean magnitude
nconsecbelowmeanbystd	UInt16Col, shape=(3,)	Number of consecutive lightcurve points that are [1, 3, 5] standard deviations below the mean magnitude
nconsecfrommeanbystd	UInt16Col, shape=(3,)	Total number of consecutive lightcurve points that are [1, 3, 5] standard deviations either above or below the mean magnitude
ngoodobs	UInt16Col	The number of epochal detections that were not flagged, masked, nor associated with bad pixels during the source extraction process; i.e., where all measurements have *catflags* = 0.
ngoodobsrel	UInt16Col	Similar to *ngoodobs* but only counting epochs that are in a public release.
nmedianbufferrange	UInt16Col	Number of points more than 20% of the lightcurve amplitude from the

Column Name	Data Type	Description
		weighted mean magnitude
nobs	UInt16Col	Total number of epochal detections (observation epochs) for this object
nobsrel	UInt16Col	Similar to *nobs* but only counting epochs that are in a public release.
npairposslope	UInt16Col	Number of positive slopes between all consecutive pairwise lightcurve points (mag[i+1] brighter than mag[i])
percentiles	Float32Col, shape=(12,)	[mag] [5th, 10th, 17.5^{th}, 25th, 32.5^{th}, 40th, 60th, 67.5^{th}, 75th, 82.5^{th}, 90th, 95th] percentiles in magnitude over nobs
ra	Float64Col	Right Ascension (deg)
refchi	Float32Col	Chi metric from reference catalog
refmag	Float32Col	Magnitude from reference catalog
refmagerr	Float32Col	Uncertainty in refmag
refsharp	Float32Col	*Sharp* metric from reference catalog
refsnr	Float32Col	SNR from reference catalog
skewness	Float32Col	Skew (third moment) in lightcurve magnitudes over nobs
smallkurtosis	Float32Col	Kurtosis (fourth moment) in lightcurve magnitudes over nobs for small samples (http://www.xycoon.com/peakedness_small_sample_test_1.htm)
stetsonj	Float32Col	Stetson-J statistic for this filter only; measures degree of autocorrelation in magnitudes over nobs
stetsonk	Float32Col	Stetson-K statistic; proportional to the kurtosis of the magnitude distribution over nobs
uncalibmeanmag	Float32Col	Mean of uncalibrated mags
vonneumannratio	Float32Col	The von Neumann Ratio statistic over nobs; measure of epoch-to-epoch variability
weightedmagrms	Float32Col	[mag] Root Mean Square magnitude over nobs using inverse-variance weighting.
weightedmeanmag	Float32Col	[mag] Inverse-variance weighted mean magnitude over nobs
x	Float64Col	cos(RA)*cos(Dec)
y	Float64Col	sin(RA)*cos(Dec)
z	Float64Col	cos(RA)

The sourcedata table contains one row for each observation of an Object in the sources table. The time series photometry is stored here.

The *matchid* column is indexed fully so that the table may be sorted by this column. The *programid* column is indexed with default values.

Column Name	Data Type	Description
catflags	Int16Col	Catalog flags from PSF-fitting catalog: see *flags* definition in §10.6. Additionally, if *catflags* contains bit 15 (value ≥ 32768), measurement at this epoch is likely unusable (e.g., affected by clouds and/or moonlight)
chi	Float32Col	Chi metric from PSF-fit catalog (see §10.6)
dec	Float64Col,	Declination [deg]
expid	Int32Col	Exposure ID (corresponds to exposures table)
hjd	Float64Col	Heliocentric Julian Date corresponding to midpoint of exposure and source ra, dec position. For speed, this is predicted from a 2D surface fit to the ra,dec centers of 81 partitions (a 9 x 9 grid) covering each matchfile footprint (= CCD quadrant).
mag	Float32Col	Magnitude that includes corrections for three photometric biases that depend on: **(i)** magnitude relative to PS1; **(ii)** position (flat-fielding); and **(iii)** survey field (sky location). For details, see: *http://nesssi.cacr.caltech.edu/ZTF/Web/Calib.html*
magerr	Float32Col	Uncertainty in *mag* measurement [mag]. Includes correction to conform to photometric repeatability RMS derived from "non-variable" population.
matchid	Int32Col	Match ID corresponding to entries in sources table
mjd	Float64Col	Modified Julian Date corresponding to midpoint of exposure (*obsmjd* + 0.5*exptime*).
programid	UInt16Col	Program ID
psfflux	Float32Col	Raw flux from PSF-fitting, with no post-corrections to account for possible photometric biases as described in link to mag column above [DN]
psffluxerr	Float32Col	Uncertainty in *psfflux*, with no post-corrections to account for possible biases or inconsistencies with predictions from photometric repeatability [DN]
ra	Float64Col	Right Ascension [deg]
sharp	Float32Col	Sharp parameter from PSF-fit catalog (see §10.6)
xpos	Float32Col	*x*-center in pixels
ypos	Float32Col	*y*-center in pixels

19 APPENDIX D: A DAY IN THE LIFE OF THE ZSDS

Figure D1 shows a timeline of the various operational steps in a typical day/night. The steps shown are all automated and executed by the Virtual Pipeline Operator (VPO). Further details and longer term operational tasks are summarized in:
http://web.ipac.caltech.edu/staff/fmasci/ztf/masci_pasadena_ops_03.19.18.pdf
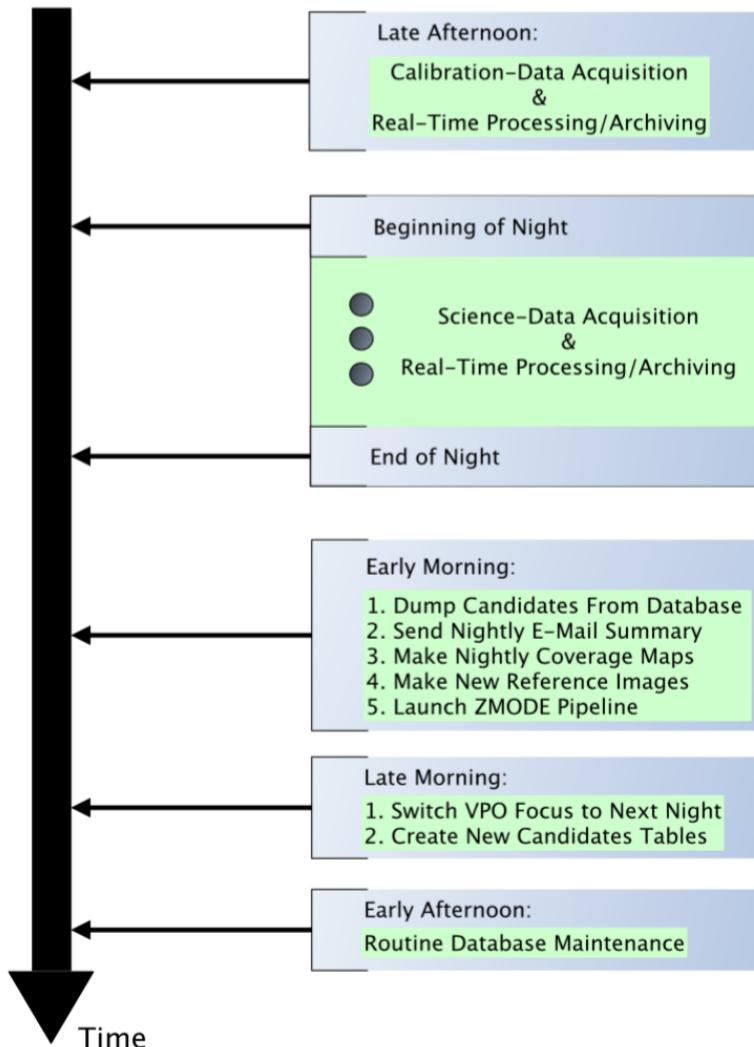
Figure D1: Schematic showing typical day in the life at the ZTF Science Data Center.