Imtrandetect: a new tool/methodology for transient detection

Frank Masci, IPAC / Caltech

Digging Faster and Deeper, Caltech, Dec 12-13, 2011

Goals / Overview

- How can we detect transient candidates to low S/N levels **optimally, reliably and quickly** from large data streams for (possible) follow-up, classification, and to support knowledge discovery?

- Mining deeper => more candidates => increases our chances of discovering rare and new events
  - Must be prepared for a higher rate of false-positives, e.g., instrumental glitches and contamination from a "fog" of uninteresting astrophysical transients. Hence need to work harder at finding those diamonds

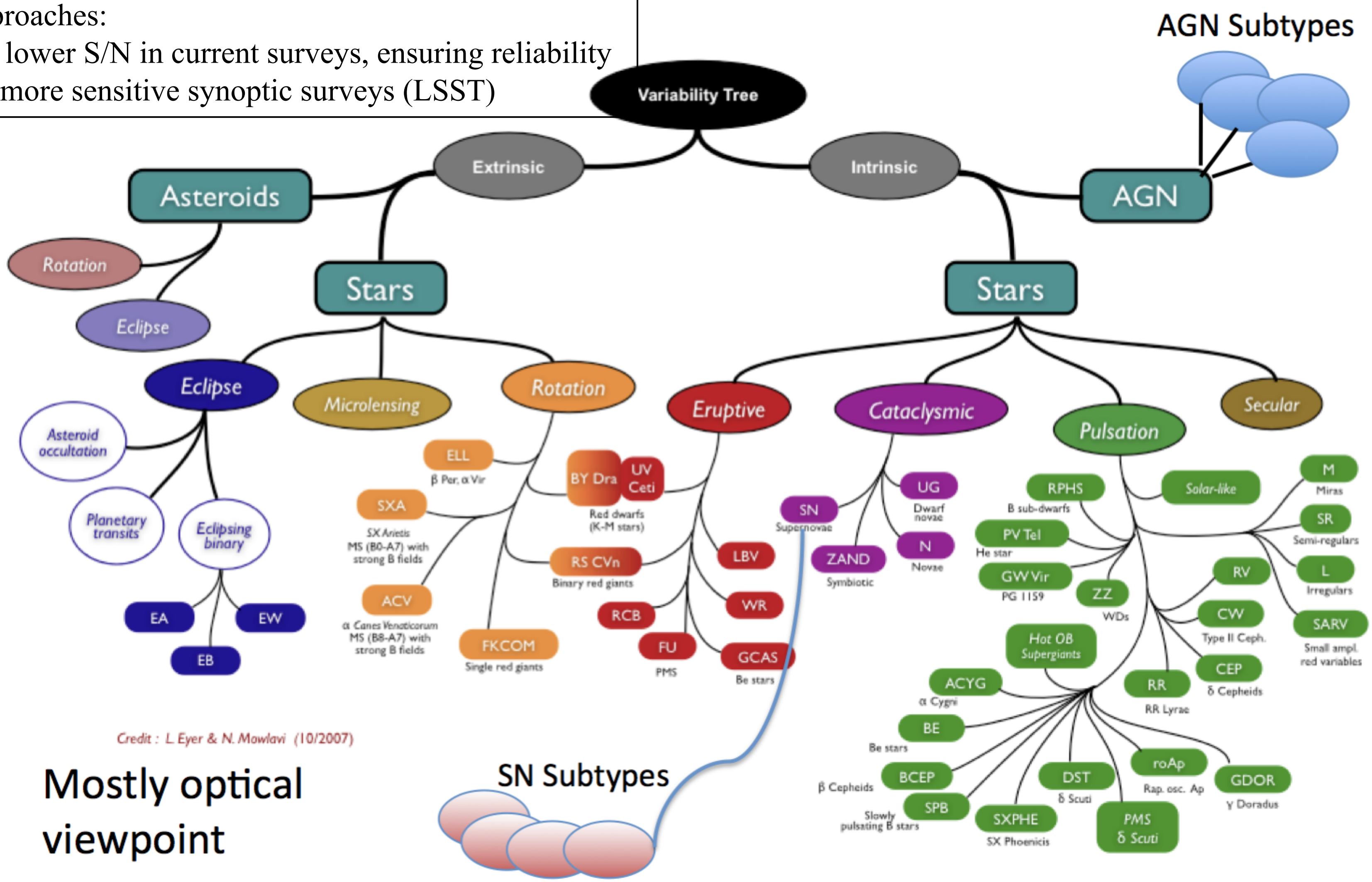
- An optimal image-based transient-detection method has been designed and implemented in an automated software tool (*imtrandetect*) that can be run in real-time for a synoptic sky survey
  - optimal method => one that maximizes the S/N of a quantity from a realization of measurements
  - emphasis is on reliability not completeness
  - the methods explored are still very experimental and software is a work in progress
  - will show examples from testing on CRTS and WISE data

- Method is mostly applicable to optical/IR data, but can be extended to harder/softer energies

- Only focus on initial transient detection process, not classification, although latter can drive former

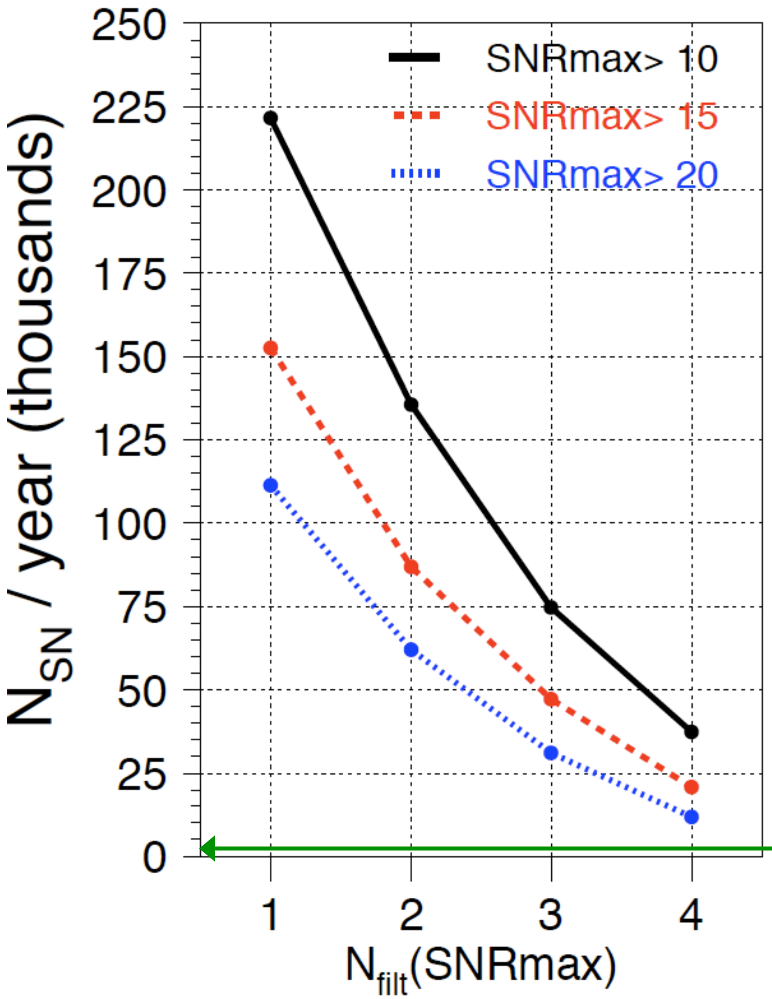
Tree of Knowledge

Which are the rare types? Where can we discover new types?
⇒ Need higher detection statistics for discovery;
Two approaches:
(1) probe to lower S/N in current surveys, ensuring reliability
(2) wait for more sensitive synoptic surveys (LSST)



Mostly optical viewpoint

E.G: Supernovae Type Ia discovery rates (now versus future)



number of LSST filters

- Will no longer be a rarity with LSST, even to modest S/N levels.
- Multiband photometry will be crucial to ensure reliable detections and classifications.
  => can afford to detect to lower S/N.

Now: <~ 900 SNe/year, CRTS + PTF + other

Bernstein et al, 2009, LSST Science Book, ch.11

Popular transient-detection methods

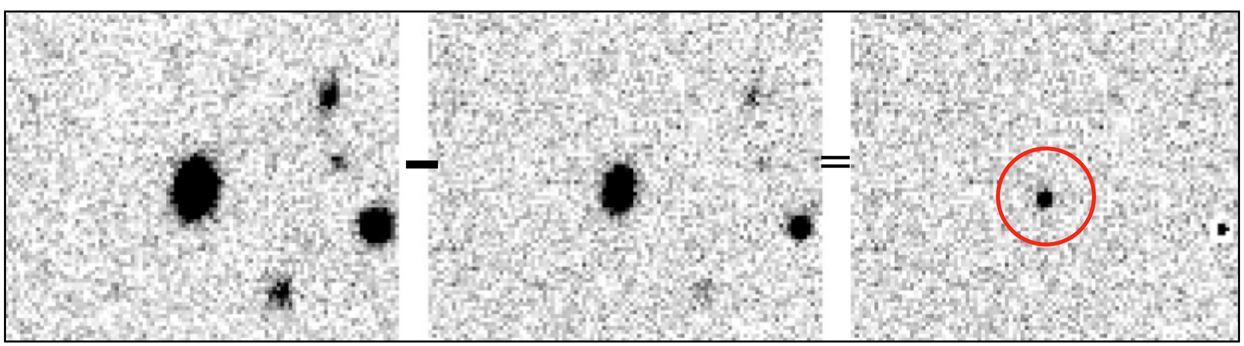
Two methods:

(1) source-catalog matching (e.g., *CRTS, Sloan, INTEGRAL …, Drake et al. 2009; Telezhinsky et al. 2010*)

- generate calibrated, clean source catalogs from deep co-adds of past images (reference catalog)
- match extractions from new images with reference catalog and search for significant flux changes

(2) image differencing (e.g., *PTF, ESSENCE, +many others …, Alard 2000; Bramich et al. 2008; Law et al. 2009*)

- create a deep template image of a field by co-adding a stack of images
- given a new observed image, register with template, perform PSF-matching and subtract them
- search for significant + or – flux changes relative to template
- powerful, but not suitable for all instruments (e.g., artifacts, residuals from atmospheric refraction)

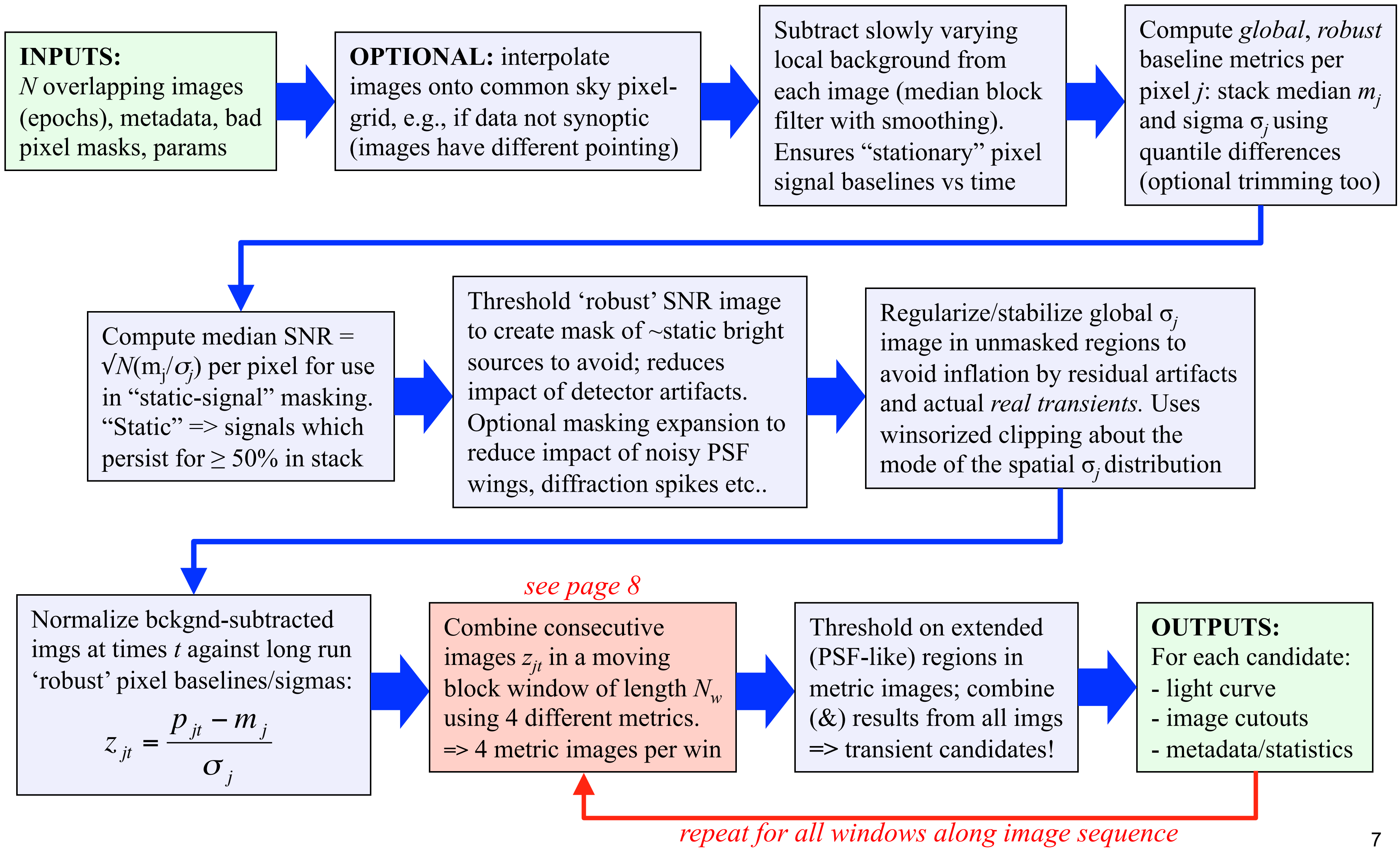


SN 2009av from PTF
Law et al. 2009

Ingredients for a robust transient-detection tool

- Mask instrumental artifacts: glints, ghosts, glitches, diffraction spikes, saturation bleeds..
- Not affected by PSF variations (temporal + spatially)
- Ability to combine images from different filters, or from contemporaneous observations from different instruments to improve S/N
- Handle data with irregularly spaced observation times, large gaps, varying throughput
- Handle images with non-uniform overlap (hence spatially-varying depth) across epochs
- Optimal use of available data given knowledge of all noise sources
- Relative photometry across image epochs is sufficient, absolute calibration not necessary
- Tunable to detect transients/variability to different S/N thresholds and characteristic timescales
- Optional use of priors to assist in isolating specific candidates (e.g., microlensing light-curves are symmetric and achromatic, while those of Supernovae and many other transients are not)
- Optional constraints to maximize chances that transient is real: e.g., must have at least n consecutive events above some S/N separated by $< \Delta t$, and must appear PSF-like (at least)
- Generate light-curves, image cutouts, transient metadata for decision making and QA
- Fully automated; able to process large data streams in real-time reasonably fast!

Processing flow in *imtrandetect* (v1.0)



Optimal image-combination metrics for transient detection

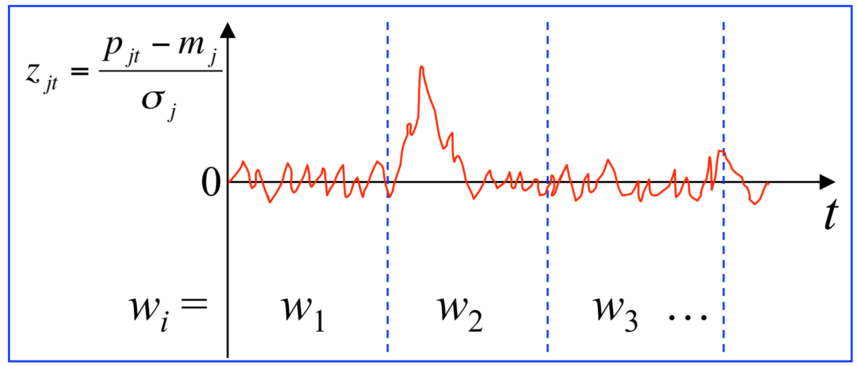
- Considered four image-combination metrics per window w_i (see below)
- An extension of *single*-image differencing method but more optimal since S/N is increased by combining multiple epochs where transient may be “active”.
- **Reason for windowing:** reduce dilution to metric S/N from baseline noise (see pg. 9)
- In practice, set window length N_{wi} to *smallest* possible such that metric S/N is high enough to detect *shortest* transient timescale of interest. For long-term variability, set length to entire data span

$$z_{j,\max} = \max\left\{ z_{jt} \ \forall \ t = 1,2,3 \ldots N_{wi} \right\} \Rightarrow \text{image of maxima}$$

$$R_j = \frac{Frac(z_{jt} \geq z_{thres})}{FracGaussian(\geq z_{thres})} \Rightarrow \text{image of frac. excesses above some threshold}$$

$$\chi_j^2 = \frac{1}{N_{wi} - 1} \sum_{t=1}^{N_{wi}} z_{jt}^2 \Rightarrow \text{image of reduced chi-squares}$$

$$S_j = \frac{\sqrt{N_{wi}(N_{wi} - 1)}}{N_{wi}(N_{wi} - 2)} \sum_{t=1}^{N_{wi}} z_{jt}^3 \Rightarrow \text{image of skewnesses}$$

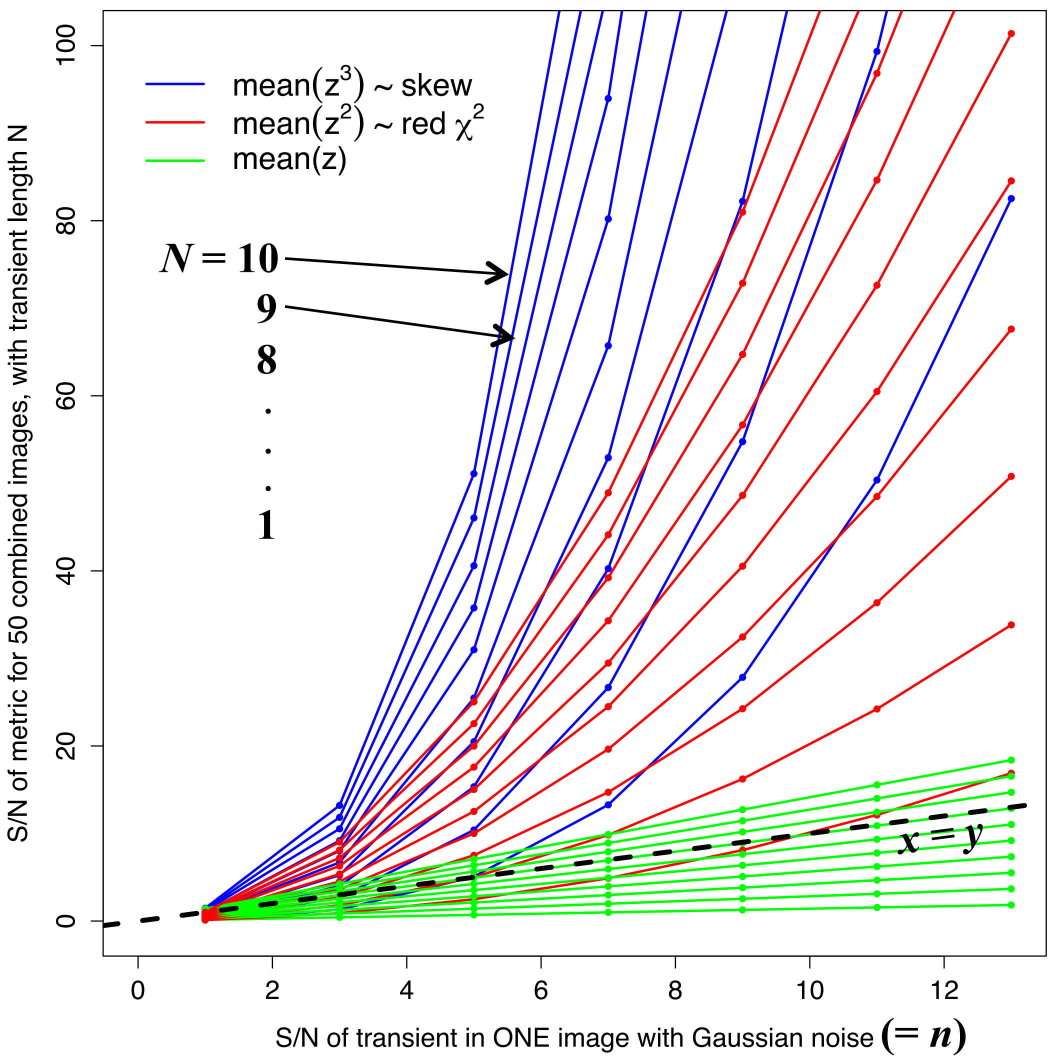


cf. to *Babu et al. 2006*, astro-ph/0612707:

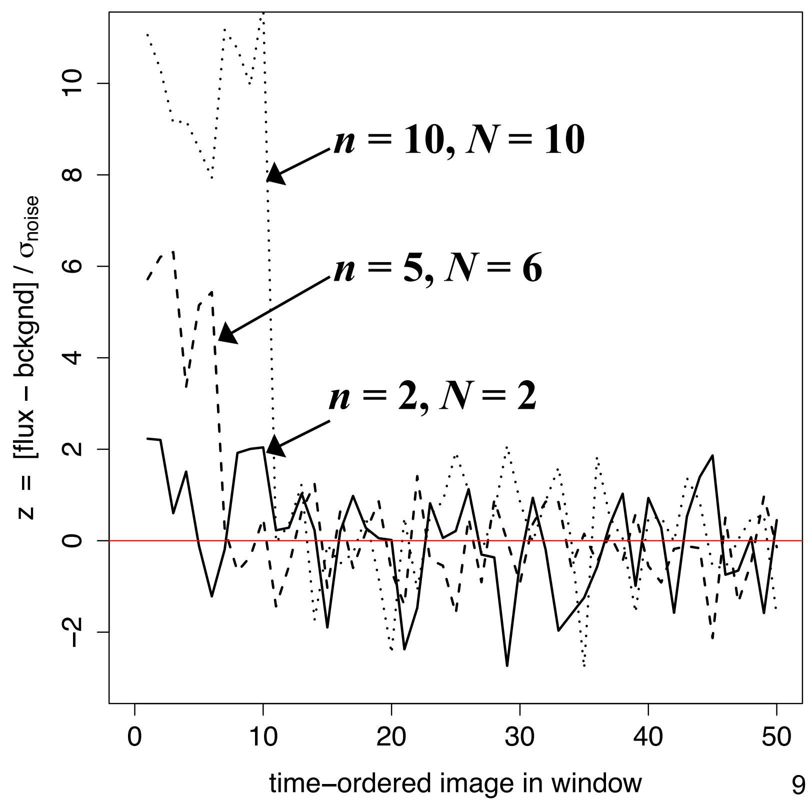
$$\chi_j^2 \equiv \left(p_j - m_j\right)^T \Omega_{Nw}^{-1} \left(p_j - m_j\right)$$

we assume cov. matrix Ω is diagonal

Sensitivity of image-combination metrics to transients

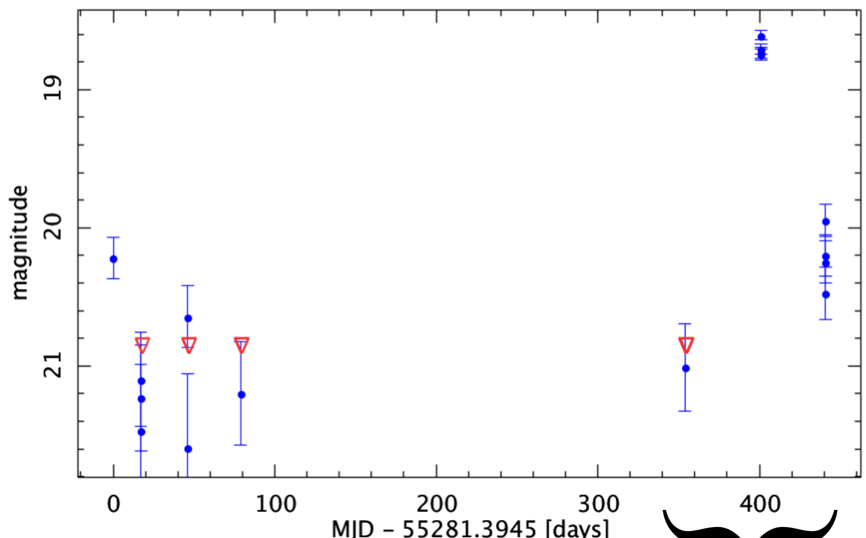


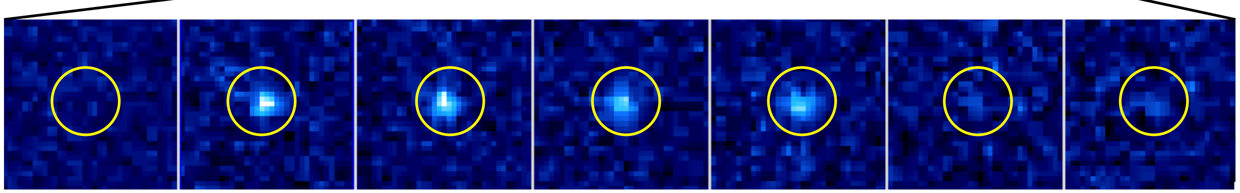
- Monte-Carlo simulation to test metric strength and dilution by uncorrelated Gaussian noise in a window of length = 50 images
- Assumed simple transient with flux ~ constant and single image S/N = n lasting for N images
- $<z^3>$ is most sensitive overall, e.g., has S/N > 10 for single-image $n \geq 3$ occurring $N \geq 8$ times

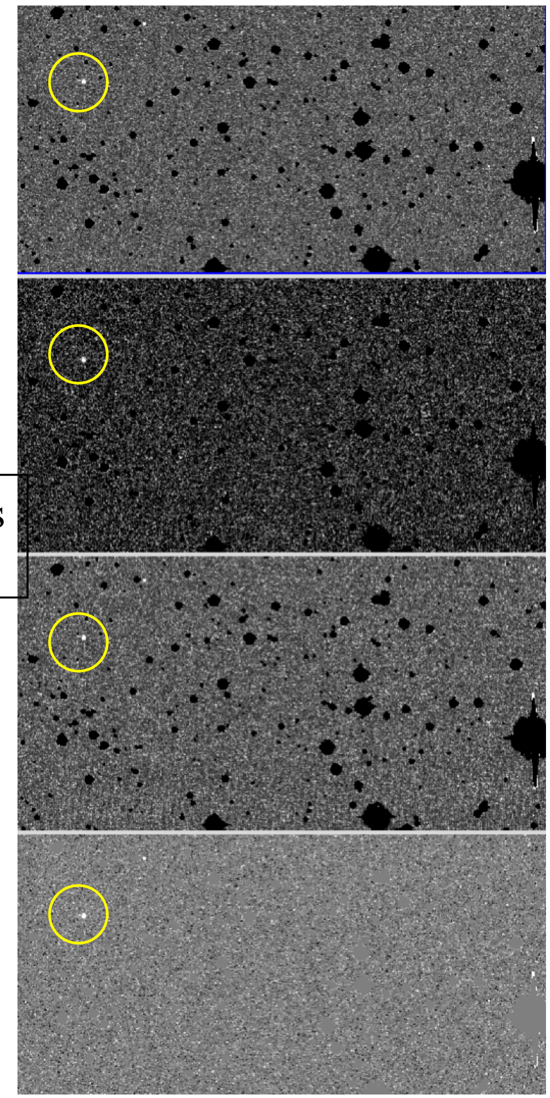


Testing on image data from the Catalina Real-Time Transient Survey (CRTS)

- Ran *imtrandetect* on ~ 0.5° × 0.5° fields containing CRTS-discovered transients (mostly Supernovae, no personal bias!) and searched for new transients.
- Below: products from a field containing SN 2011cw (Type IIn), discovered 5/1/2011. Used a running window of 15 images.



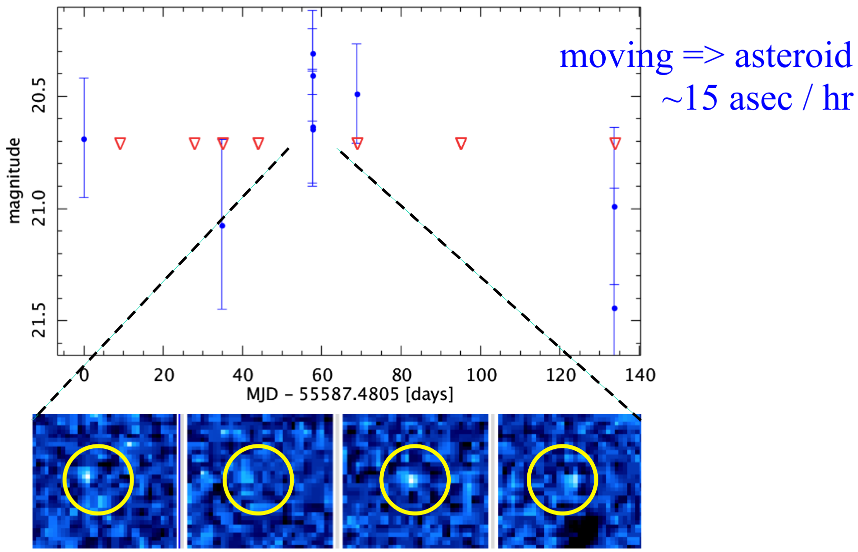


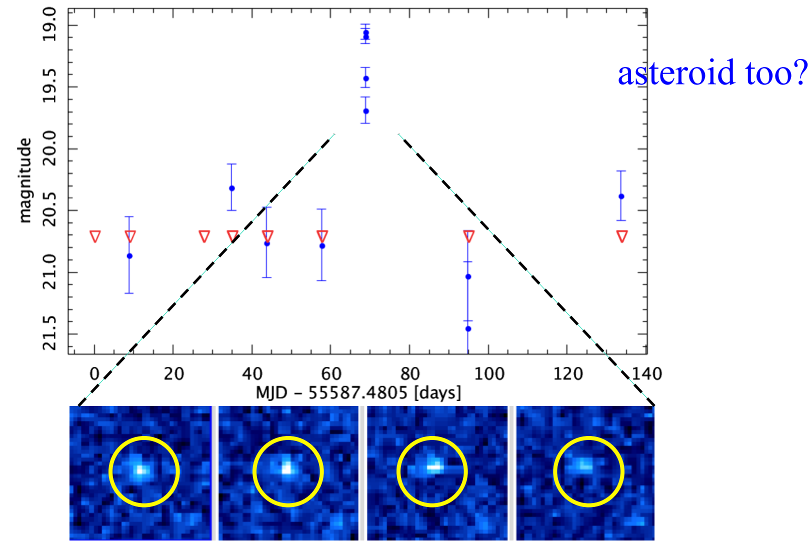


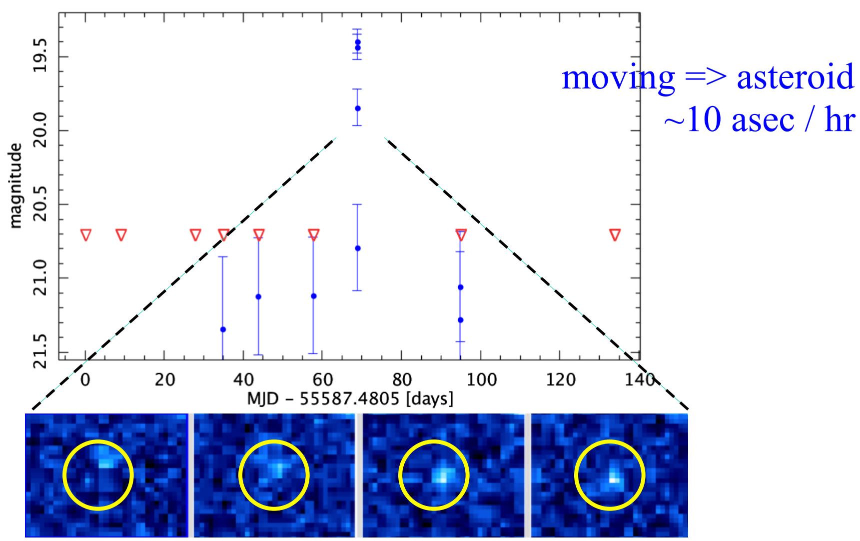
4 image metrics from page 8

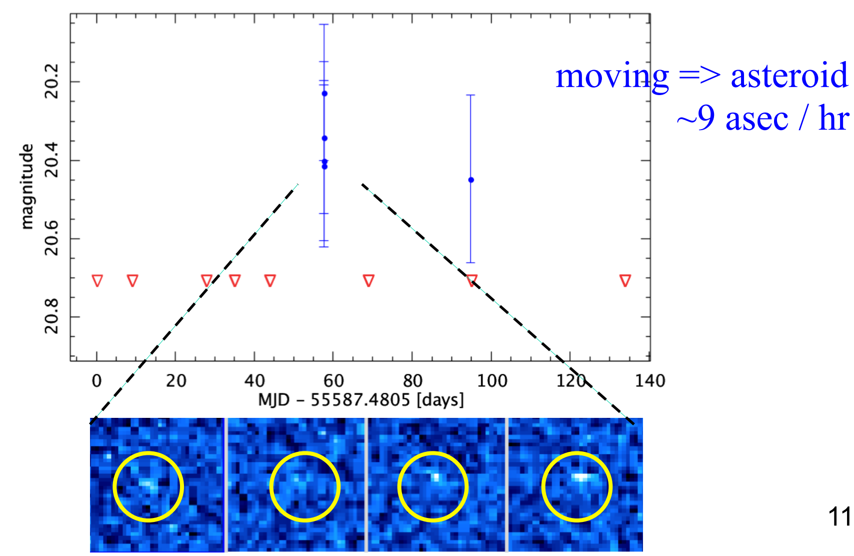
Some new(?) CRTS transient candidates

Pushed to S/N = 3, found spurious transient rate of ~8% (glitches, glints, CRs) and <u>tons of faint asteroids</u>!





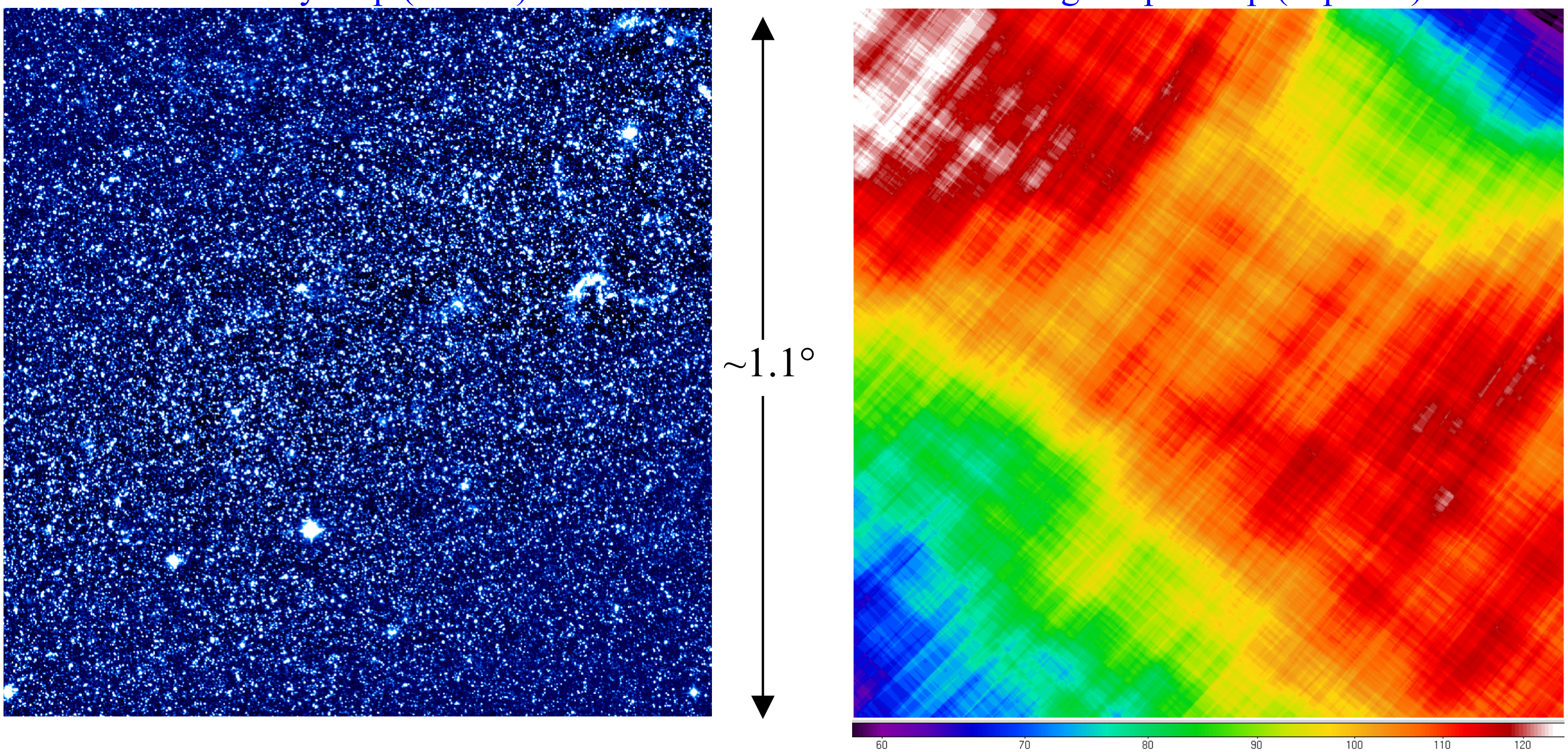




Wide-field Infrared Survey Explorer
Core of LMC (3.4μm)

- **Not a synoptic survey**, but have multiple irregularly spaced epochs of non-uniform depth over sky
- Performed a blind search for variable stars at 3.4 μm in the core of the LMC using *imtrandetect*

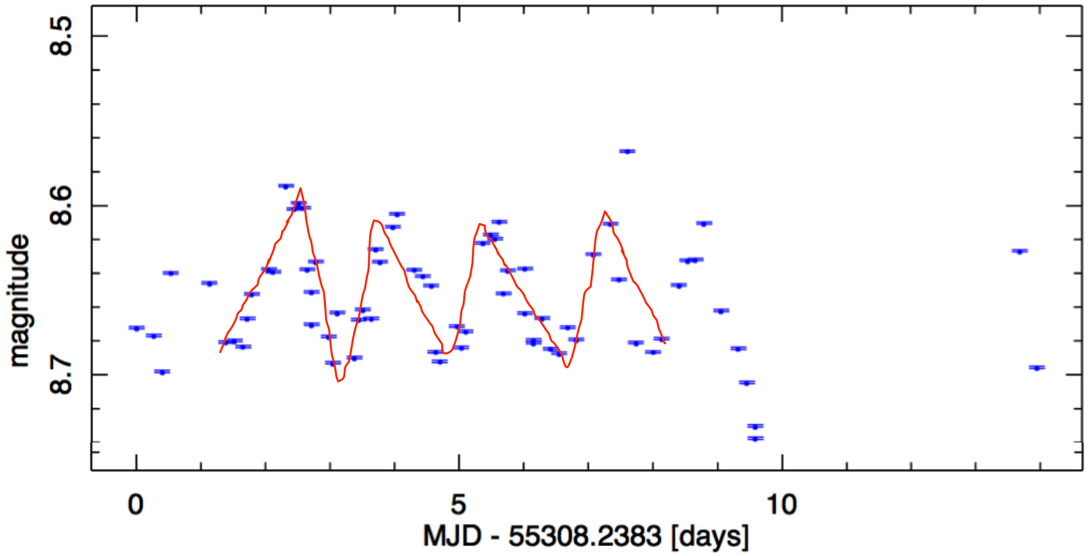
intensity map (co-add) image-depth map (#epochs)



~60 deep ~125 deep

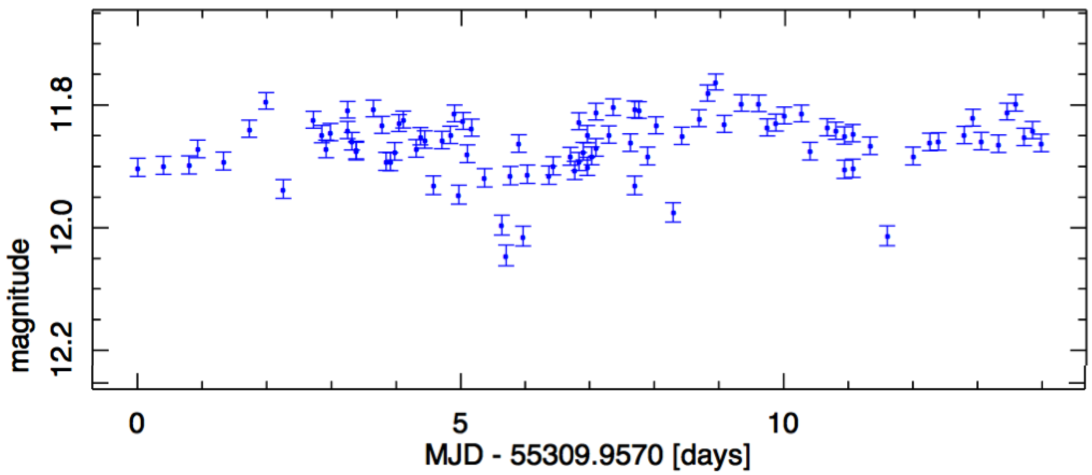
⇒ sample baselines of ~ 15 – 20 days

Example light-curves of variable candidates in LMC (3.4μm)



New RR-Lyrae? Period ~1.58 days

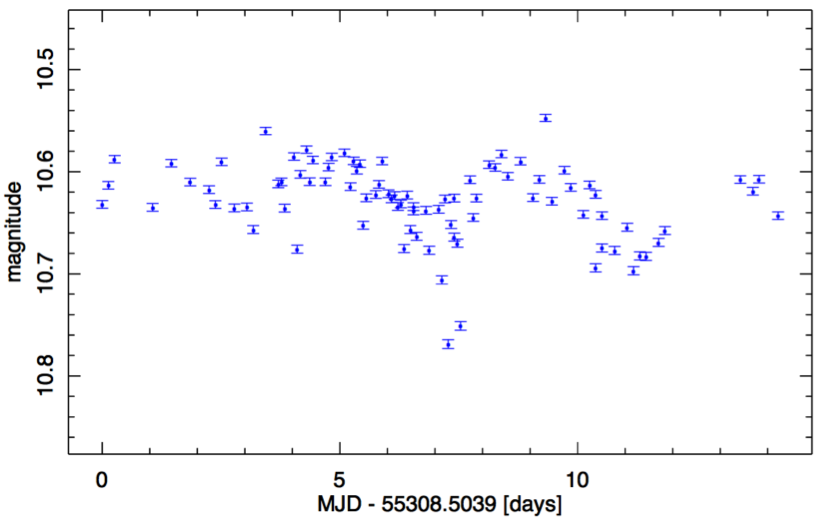
Red curve => to guide the eye

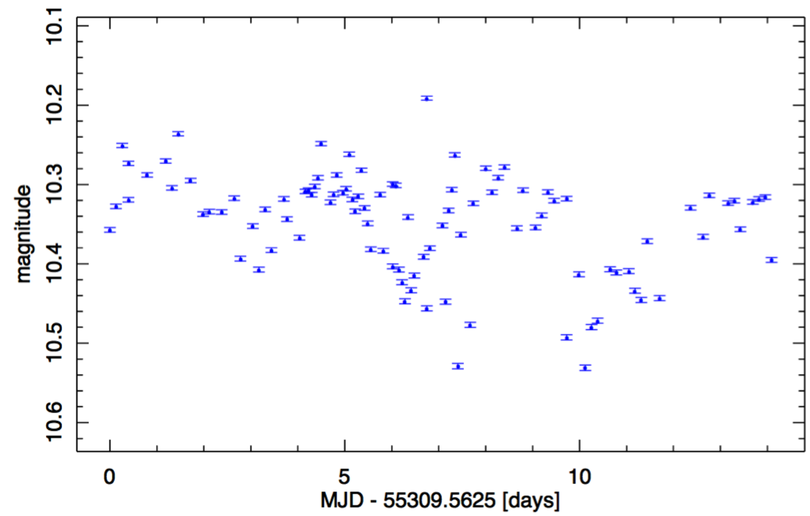


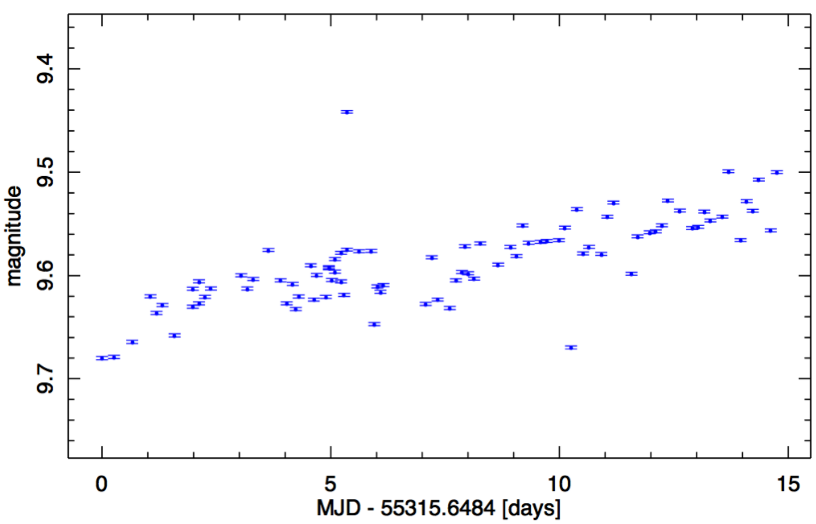
Classified as a short period Cepheid in optical by OGLE

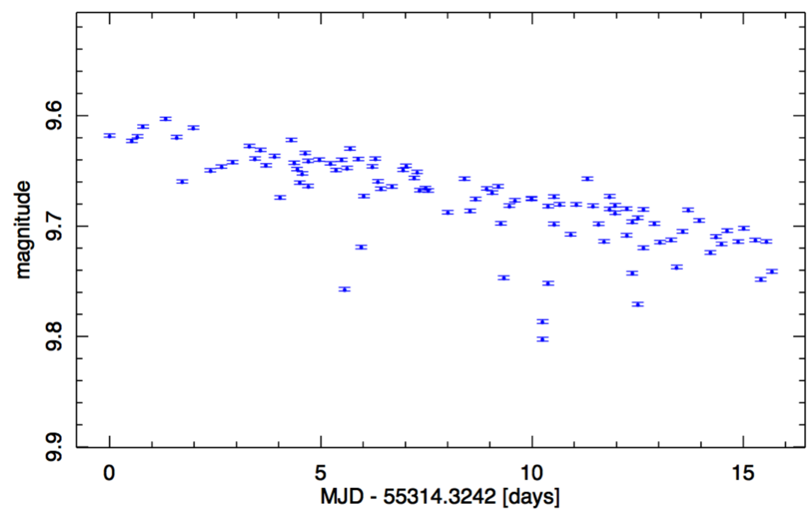
Example light-curves of variable candidates in LMC (3.4μm)

Some **“uncertain unknowns”**. Main limitation for IDs are short noisy baselines









Summary, closing thoughts..

- Described some optimal metrics for detecting transients, implemented in a new standalone tool: *imtrandetect* (work in progress; plan to make publically available)

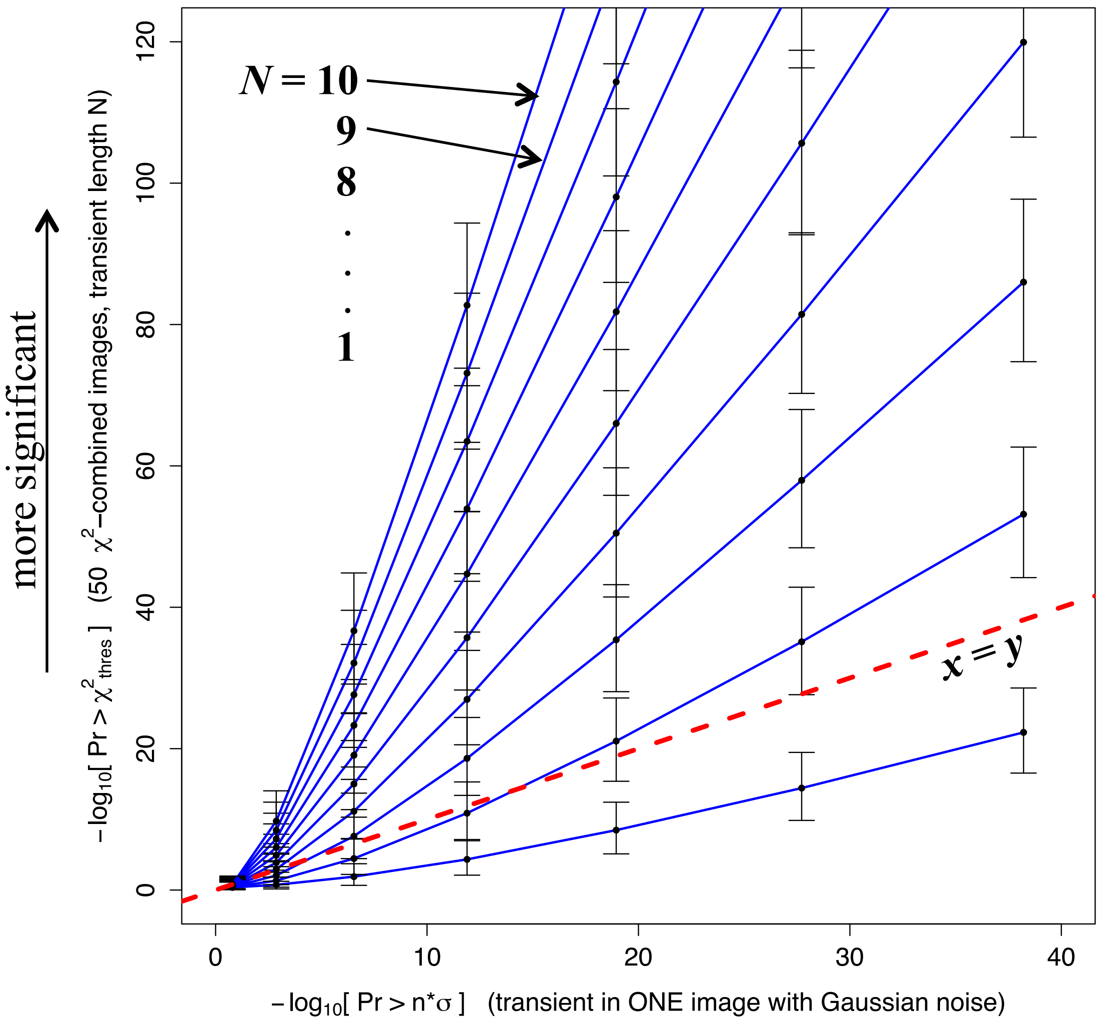
- **Goal:** probe down to low S/N levels to maximize chances of discovering rare/new events
  - ⇒ place more emphasis on reliability than completeness, e.g., through static-source filtering to reduce detector and bright-source artifacts, varying PSF, …

- Optimization methods are easy to apply for classical noise-distributions, and the CLT usually saves us. But to apply the *most optimal* method, need to understand how all ***your*** noise sources are distributed, especially systematics. These control the degree of reliability.
  - ⇒ future panchromatic surveys will vastly improve reliability against artifacts and uninteresting transient "fog"

- **Dilemma still remains:** how and which candidates do we follow-up for classification from the deluge of detections to low S/N (which is still a problem today at high S/N)?
  - is rigorous, exhaustive follow-up always needed? Statistical studies on archives of reliable light-curves using contextual info, classification templates, models, matched filters… will still be very powerful
  - still strive to make unsupervised, real-time classification an integral component of synoptic surveys

Backup Slides

Sensitivity of χ^2 metric to transients

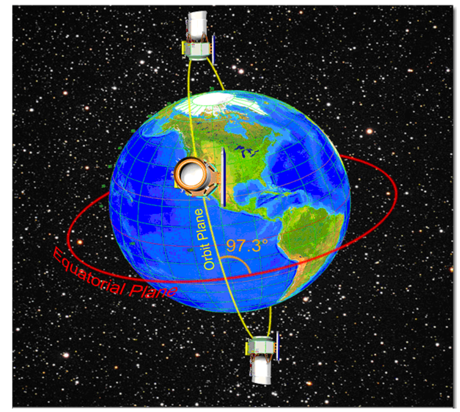


S/N = *n* = 1, 3, 5, 7, 9, 11, 13

- Different representation for the χ^2 image metric: probability that observed flux sequence will occur by chance under H0 of no transient in the presence of pure Gaussian noise (see page 9)
- Distribution for a χ^2 random variable is well known, but not for *skew*
- In general, the lower the single image epoch S/N, the *longer* the transient must persist at >~ S/N for it to be detected using these image-combination metrics
- These results assume one can robustly measure a long-run baseline level per pixel (stationary background; e.g., a stacked median) and a robust σ_{noise} with no contamination from actual transient

Testing on WISE image data

- The **W**ide-field **I**nfrared **S**urvey **E**xplorer:
  - Performed an all-sky survey in 2010 in four IR bands: ~ **3.4**, **4.6**, **12**, **22 μm**
  - **Not a synoptic survey**, but have multiple epochs of non-uniform depth over sky
  - Sun-synchronous Earth-polar orbit
    => image-depth (#epochs) increases towards ecliptic poles
    => Large Magellanic Cloud (LMC) fits the bill
  - Observation times irregularly spaced, Δt >~ 95 minutes
  - Baselines sampled: ~ 2 days to > 6 months over sky



- Performed a blind search for variable stars at 3.4 μm in the core of the LMC using *imtrandetect*. Postdoc Doug Hoffman is assisting with classifications

- *imtrandetect* is most optimal for detecting transients but can be tuned to detect variables or anything with a flux variance in excess of that expected for a static source of similar flux

- Have multi-band information to improve reliability of transient/variable-source detections

WISE photometric repeatability (3.4μm)