Optimal Detection of Rare "sub-significant" Events in the Time-Domain

Frank J. Masci[a] and the Keck Institute for Space Studies "Digging Deeper" collaboration

[a]*Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA*
Email: fmasci@ipac.caltech.edu

Abstract. One of the challenges in current and future synoptic sky surveys is to identify reliable candidate transient sources from immense data streams that can lend themselves to follow-up and classification. To increase one's chances of discovering rare and new events will require either pushing to fainter flux levels with a "bigger" telescope to maintain a relatively high signal-to-noise ratio (an inevitable consequence of technological growth), or, probing to lower *single-epoch* signal-to-noise ratios and combining the data in such a way to maximize the reliability and statistical significance of detections. The latter is the topic of this study and is applicable to the current era of synoptic surveys where one may wish to extend discovery space to guide future exploration. However, this process will not end as new technology comes on-line. There will always be a desire to probe deeper and optimize discovery methodologies against instrumental and physical limitations. We offer some ideas on how to optimize the detection of transients from multi-epoch imaging data using various statistical measures in image-pixel space. We explore their sensitivity, compare to traditional approaches, and test them on data from the Catalina Real-time Transient Survey (CRTS).

Keywords: methods: analytical – methods: data analysis – methods: statistical – techniques: image processing – supernovae: general

1. INTRODUCTION

Time domain astronomy is now transitioning into an immensely data-rich regime made possible by advancements in instrumentation and technology. This is enabling efficient wide-field surveys and progress in data-management practices to handle peta-scale archives and databases. Current surveys include the CRTS (Drake et al. 2009), PTF (Law et al. 2009), and Pan-STARRS (Kaiser et al. 2002) in the optical, and various SKA pathfinder missions in the radio (Dewdney et al. 2009). These are setting the course for the next generation of surveys like LSST (Ivezic et al. 2011) and SKA that will yield tens of terabytes nightly over a span of at least five years. The community is faced with a challenge to process and mine this flood of data as efficiently and optimally as possible.

The first important step is the detection of candidate transients and variables at some level of significance (set by the maximum tolerable fraction of false positives) that ensures one is reasonably confident the detections are real and worthy of follow-up. However, there is always a desire to open-up discovery space and attempt to

uncover the "unknown unknowns". To increase our chances of discovering rare and new events, either through archival analyses or real time follow-up of interesting events would require probing to lower flux levels (or larger spatial volumes). This may mean detecting events with a lower signal-to-noise (S/N) ratio in *single-epoch* exposures when the physical capabilities of the instrumentation are reached since there is little scientific return when operating at nominal (more conservative) extraction thresholds. By mining deeper into the "muck", we must be prepared for a higher rate of false positives, e.g., detector glitches and artifacts, and/or contamination from a "fog" of uninteresting astrophysical transients. Consequently, we need to place more emphasis on ensuring reliability at the expense of completeness.

A popular method for detecting transients is difference imaging with prior PSF-matching against a deeper reference image (e.g., Alard 2000; Bramich 2008). This method is powerful, but is limited by the PSF-matching accuracy in a single-epoch exposure, often requiring a relatively high detection S/N threshold in the difference image to offset the residuals. Furthermore, this method can be very difficult to apply for some detectors due to atmospheric refraction effects (e.g., CRTS; Drake et al. 2009).

This paper focuses on optimizing transient detection at low to moderately low S/N levels across multiple single-exposure observations. We have explored some optimal image-combination metrics to achieve this (§2). By "optimal", we mean in the maximum-likelihood sense according to the noise-distribution followed by the input measurements. In fact, our method is optimized for optical/IR data where the underlying photon-noise is well into the Gaussian limit (setting aside systematic error sources). An important question is: how low a S/N can we reach in a series of single-epoch observations to ensure a moderately high significance in our combined metric-image space? We address this using Monte-Carlo simulations. Our focus here is *reliable* identification of faint, low S/N transient candidates, not classification, although the latter can assist the former to weed out the "known unknowns". We have implemented our method in a prototype software tool "*imtrandetect*" which we briefly describe in §3. We present some preliminary results of testing on image data from the CRTS in §4.

2. METHODOLOGY

The problem of detecting faint (usually low-significant) events in single epoch observations entails devising a statistic that combines information from multiple consecutive epochs in time that we can test for significant excursion above the null hypothesis (H0) of pure noise fluctuations. The difficulty is finding a statistic which is most sensitive to repeated (systematic) behavior in a series of measurements (different from that expected by the underlying noise) that may suggest a faint transient. This assumes the underlying noise (including any correlated behavior over time) is well characterized beforehand.

Figure 1 shows a schematic of a transient whose peak signal is shown to be relatively strong for clarity and the purposes of this discussion. The measurements may be that of a single pixel j through a stack of registered, time-ordered images, which we represent as z-scores, i.e., the number of sigma above the pixel's 'long-run' baseline level, m_j. The sigma value here (σ_j) is characteristic of the pixel over time

(see below for estimation methods). Figure 1 also shows marginal distributions by collapsing the time axis and binning the z-scores. One can see that if the time series is windowed in time, the distribution of measurements for the window containing a suspect transient will be more skewed or rather, have a relatively greater fraction of values with large excursions from the underlying baseline level above the *long-run* noise (σ_j) than if the distribution from all measurements (i.e., from all windows) is used.

Windowing a series (with some optimum window-length; see below) therefore reduces dilution from the underlying noise. For relatively faint short-lived transients (compared to the available history of measurements on a sky location) the baseline noise will dominate if many 'null' measurements over a long time-span were combined, hence rendering reliable detection difficult. Therefore, windowing increases our chances of detecting faint transients on *local* time-scales if one has a good handle on the historical noise at a given sky location. This is an improvement over traditional single-epoch image differencing (which is a point-wise process in time) since by combining multiple consecutive epochs (assuming they are relatively closely separated in time; i.e., to provide good sampling of the target transients), will increase the detection S/N. This is the crux of our method. We expand on the details and limitations below.

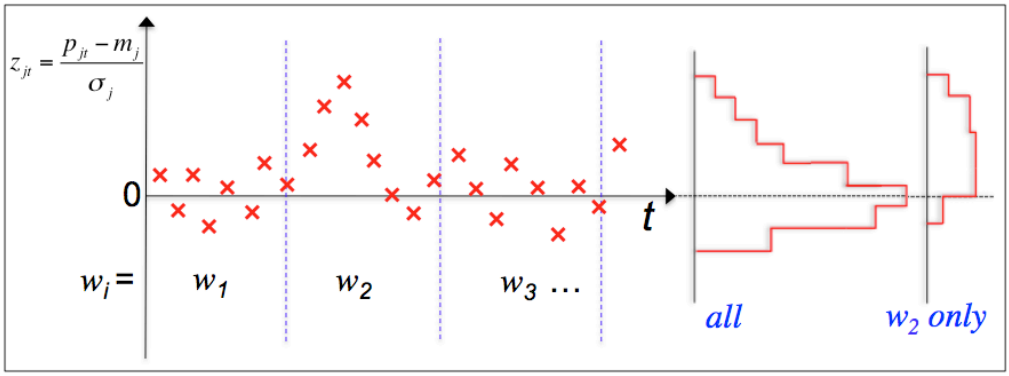

FIGURE 1. Schematic of windowing scheme for a series of noisy pixel measurements (relative to some baseline m_j and normalized by the long run noise-sigma, σ_j, over time on a sky location j). A hypothetical transient appears above the noise in window w_2. On the far right are two marginal distributions (collapsed along the time axis) formed by *all* measurements and only those in window w_2.

We have constructed several image-combination metrics for "collapsing" a series of time-ordered pixel measurements (z-scores) from a set of sky-registered images. These "metric-images" are generated for each window along the time-sequence. At first, we experimented with four metrics (per pixel stack in a window): (i) the maximum pixel z-score; (ii) the fractional excess of z-score values above some threshold relative to that expected from noise alone (e.g., a Gaussian distribution); (iii) the classic reduced chi-square; and (iv) the third central moment (which we refer to as the *skew* from now on, with symbol S). These are respectively defined as follows for a pixel stack j in window i containing N_{wi} images:

$$z_{ij,\max} = \max\left\{z_{ijt} \quad \forall\ t = 1,2,3\ldots N_{wi}\right\} \tag{1}$$

$$R_{ij} = \frac{Frac\left(z_{ijt} \geq z_{thres}\right)}{0.5\left[1 - \mathrm{erf}\left(z_{thres}/\sqrt{2}\right)\right]} \tag{2}$$

$$\chi^2_{ij} = \frac{1}{N_{wi} - 1} \sum_{t=1}^{N_{wi}} z^2_{ijt} \tag{3}$$

$$S_{ij} = \frac{\sqrt{N_{wi}\left(N_{wi} - 1\right)}}{N_{wi}\left(N_{wi} - 2\right)} \sum_{t=1}^{N_{wi}} z^3_{ijt} \tag{4}$$

where

$$z_{ijt} = \frac{p_{ijt} - m_j}{\sigma_j} \tag{5}$$

for a pixel signal p_{ijt} falling in window i, at sky location j, and measured at time t that exhibits a *long-run* "static" baseline level m_j and noise-sigma σ_j. The functions of N_{wi} pre-multiplying Eqs (3) and (4) make these metrics *unbiased* estimators of the respective *normal* population values since each is based on sample estimates of the location (baseline level) and noise-variance (e.g., Pearson 1931). A χ^2 statistic (similar to Eq. 3) was used by Szalay et al. (1999) for generic source detection by combining images across multiple passbands. It is used here in a somewhat different context. Furthermore, a related metric was explored by Babu et al. (2008) for transient detection, referred to as the *Mahalonobis* distance. This is just the square-root of the χ^2 metric that includes covariance estimates between the measurements:
$D_M = \sqrt{[(\boldsymbol{P_j} - \boldsymbol{M_j})^{\mathrm{T}}\Omega^{-1}(\boldsymbol{P_j} - \boldsymbol{M_j})]}$, where $\boldsymbol{P_j}$, $\boldsymbol{M_j}$ are the data and mean column vectors respectively and Ω is the error-covariance matrix. We ignore correlations since our method only combines pixel measurements in the temporal domain where they are expected to be largely independent, i.e., Ω is diagonal, while Babu et al. also combine measurements in the spatial domain which are not necessarily independent.

The metric-images formed by metrics (1) – (4) can then be thresholded to identify transient candidates through use of a matched filter (e.g., that appear PSF-like), or searching for spatially contiguous hi-values above some local spatial-noise threshold. As a detail, one may want to place these metrics on an equal footing for thresholding purposes, e.g., by converting them to probabilities per pixel (i.e., of getting a value larger than that observed by "chance" under a H0 of pure noise). The best approach is to use empirical null probability distributions derived from the data at hand. This will capture the noise structure and properties inherent in the data itself, including systematics, correlated-noise etc. The calibration of null empirical probability distributions is cumbersome, although it need only be done once for the detector/instrument being used. In this initial study, we opted to threshold the metric-image values directly relative to the mean and sigma of a *sample* metric expected under a H0 that measurement errors are distributed as Gaussian. Taking metrics (3) and (4) for instance, we convert these to equivalent z-scores that can be thresholded:

$$z(\chi^2_{ij}) = \frac{\chi^2_{ij} - 1}{\sqrt{2/(N_{wi} - 1)}} \tag{6}$$

$$z(S_{ij}) = \frac{S_{ij}}{\sqrt{\sigma^2_{Sij}}} \tag{7}$$

where

$$\sigma^2_{Sij} = \frac{6\left(N_{wi} - 2\right)}{\left(N_{wi} + 1\right)\left(N_{wi} + 3\right)}. \tag{8}$$

Equation (6) follows from the fact that the mean and variance of the reduced χ^2 are 1 and $2/dof$ respectively, where the number of degrees of freedom is $dof = N_{wi} - 1$. Equation (7) uses the fact that the *skew* (third moment) for a Gaussian population of errors is zero, and the expected variance for the *sample skew* as computed using the unbiased estimator in Eq. (4) for Gaussian-distributed errors is given by Eq. (8). This took some effort to verify via simulations, but discussions can be found in Cramer (1946) and Pearson (1931). Furthermore, we find that predictions for the sample mean and variance as used in Eqs (6) and (7) conform very well to empirical (histogram-derived) estimates in real optical/IR image data in noisy background regions, testament that the Gaussian-noise assumption is acceptable (when systematics and instrumental glitches are at bay!).

Experiments on real and simulated data revealed that metrics (1) and (2) did not perform as well as the reduced χ^2 and *skew* in Eqs (3) and (4) respectively. The results showed that the metric images from Eqs (1) and (2) were very noisy and generated a plethora of false positives when thresholded. The two that looked most promising (in terms of maximizing detection S/N in metric-image space; see §2.2) were the χ^2 and *skew* metrics. For the remainder of this paper, we focus on these last two metrics. Even though related, these metrics reinforce each other in that the *skew* preserves the *sign* of an event (or events), i.e., whether it is a positive or negative excursion relative to the baseline level. Negative excursions are obviously unphysical (e.g., instrumental glitches) and can be immediately flagged as unreliable. The χ^2 however, depends on the square of fluctuations and cannot be used on its own to reject negative excursions.

Earlier we fleetingly mentioned long-run estimates of the underlying baseline-level and noise-sigma per pixel-stack at sky position j (i.e., m_j and σ_j respectively in Eq. 5) for computing z-scores. By "long-run", we mean over the available history of pixel measurements, or a large number of them to properly capture the average temporal behavior of a detector's pixel when collecting real flux from the sky and possibly a *static* astrophysical source. Before using the pixels in a set of images acquired at different times to compute m_j and σ_j, we first stabilize the pixels in each image against possible temporal variations in the sky background by subtracting a local estimate of the background (at low spatial frequencies) from each respective image (details are given in §3). We do not consider possible *uncalibrated* multiplicative effects over time (e.g., changing instrumental throughput, atmospheric transparency, etc.), nor possible changes in the noise properties of a detector (including photon noise). Once the *single-epoch* images have been stabilized against local offset variations (i.e., effectively a de-trending operation), the challenge then is to estimate the m_j and σ_j images as robustly

as possible from the global image stack (over all windows). The goal is to be robust against the possible presence of transients that may bias m_j and/or σ_j for a pixel relative to that expected in the steady state, i.e., containing a static or null signal that fluctuates according to the properties of the detector and photon collection process. One may resort to using a well characterized noise model for σ_j, but from experience, we have found such models difficult to tune over the full dynamic flux range of a detector. We have decided to estimate m_j and σ_j directly from the data (with some caveats in mind, see below). We adopted a simple median for m_j and half the difference in 15.85 and 84.13 percentile values in each temporal pixel-stack:

$$\sigma_j = 0.5\left[p_j(84.13\%) - p_j(15.86\%)\right]. \quad (9)$$

This is equivalent to the standard deviation of a Gaussian population. We ignore the convergence properties of sample estimates based on Eq. (9) with respect to unbiased population estimates for now. The important thing is that this is robust and its accuracy improves appreciably as more data is used. In our software implementation (§3), we have an option to globally regularize estimates from Eq. (9) for pixels j that fall on sources whose flux varies significantly on regular (or perhaps irregular) timescales. These sources will inevitably inflate estimates of σ_j. We regularize the σ_j image by computing its *mode* and robust spatial *RMS* over all pixels j and then winsorize (reset) σ_j values exceeding some threshold: *mode* + *n*RMS* to equal this threshold itself. This is still an approximation, but it reduces the incidence of high stack sigmas due to the presence of real astrophysical variables and intermittent transients, bringing down their σ_j, increasing their z-scores (Eq. 5), and increasing their chance of detection in the metric images.

2.1 Assumptions, Caveats, and Limitations

At this stage, some limitations of the above method are worth noting. Many of these have been fleshed out during the course of testing on CRTS data (details are expanded in the software description in §3). We stress that this method is not a generic tool for detecting all flavors of astrophysical transients and variables. Aside from the limitations imposed by the data (e.g., separation of observation epochs, Earth's atmosphere, instrumental glitches), there are assumptions in our design that severely limit the physical transient phase space. Our methodology is intended to complement other more generic search methods (e.g., single-epoch image differencing), with the goal of extending discovery space.

First, the method is ideally suited to detecting *faint* (possibly intermittent) transients close to the background level, and not continuous variables (that may vary regularly or sporadically). By faint, we mean below some S/N threshold in the median-combined global stack image (m_j). Pixel signals above this threshold (e.g., typically 5 to 7) are masked and excluded from all the windowed metric images (i.e., via Eqns 3 and 4), and do not participate in the transient search. The reason for this is to minimize contamination from detector artifacts associated with bright sources (e.g., diffraction spikes, noisy PSF wings, charge bleeds, etc.). This masking is defined using the global median-combined image since the majority of sources in this image will be static, or

more precisely, will have been active for ≥ 50% of the time spanned by the single-epoch images used to compute the median. Therefore, potentially transient or variable sources with a *long-run* median signal exceeding some user-specified S/N threshold (either static or periodically varying with ≥ 50% of its "peak" phases above the threshold) will be missed. This includes bona fide transients superimposed on extended sources with high apparent surface brightness, e.g., supernovae that explode in nearby galaxies. Our design therefore severely sacrifices completeness for reliability, since the latter is of utmost importance at faint, low S/N levels when searching for rare events.

Second, a related issue is the use of a median to estimate the long-run baseline-level per pixel (m_j), which enters in computation of the z-scores (Eq. 5) and eventually the windowed metrics for transient detection. An astrophysical transient must persist for <50% of the entire historical length of the series of single-epoch images under investigation (from which m_j is computed) to stand a chance of detection. A signal persisting for ≥ 50% over the span of all epochs will be pegged to m_j (the median) and treated as static, resulting in null z-score values and metrics. Therefore it is advised that a sufficiently long historical set of observations be used in order to be sensitive to the longest transient timescales of interest, i.e., up to half the historical span.

Tied to the previous point is selection of an optimal window size for computing the metric images. One may get the impression that a specific size will bias against certain types of transients, but this is not the case. Note that the window size is defined as the *number* of single-epoch observations in a partition, regardless of their separation in time, regular or irregular. The frequency of observations obviously determines what types of transients can be detected. The window must be small enough so the metrics are sensitive to the shortest-lived transients of interest, given limitations imposed by the observing frequency. That is, as discussed above, such that dilution from noisy measurements within the window is minimized and the metric S/N is maximized (see Figure 1). However, the window size must be big enough to ensure good statistics are accumulated for the faintest longer-lived transients so the metric S/N is maximized as well. Once a window size is selected, the metrics will then be sensitive to *all* transient timescales exceeding the window size, but < 50% the full history of observations from which the baseline median m_j and noise-sigma σ_j in Eq. (5) were computed. As discussed above, transients persisting longer than this will not be detected since they'll be pegged at the value m_j resulting in a z-score of 0. Tuning of the window size may be done via simulations (e.g., see sec 2.2).

In practice, the windowed, time-collapsed metric-images from either Eq. (3) or (4) (or z-score equivalents in Eq. 6 and 7) may be generated from a historical set of images in an archive, or in real time as new observations become available and some minimum number of images is reached within the window to trigger generation of a new metric-image. Another possible caveat is that depending on the observing cadence, window size, and transient time-scale of a source, this process may incur a longer lag-time for alerting that an event has occurred (or *is* occurring) compared to the traditional single-epoch image differencing method.

We stress that the important elements for this method to work optimally are the derivation of unbiased, robust estimates of the baseline level m_j and noise-sigma σ_j to capture the *long-run* steady state behavior of the instrument, including any fluctuations in throughput (and detector gain) that controls the level of photon-noise

observed. The more data, the better, but a long enough history of observations must exist to enable these parameters to be determined in the first place. They can be refined as more observations are accumulated.

2.2 Enhancing the effective detection S/N: how deep can we go?

We have performed Monte Carlo simulations to explore the sensitivity of the χ^2 and *skew* epoch-combination metrics (Eqs 3 and 4 respectively) to the number of times a transient is measured at or above some single-epoch S/N level within a window of noisy observations. We assumed a window containing 15 hypothesized observation epochs, which could be part of a much longer history of observations from which *long-run* estimates of the baseline level m_j and noise-sigma σ_j were derived, as discussed above. Any window length would suffice, with the number of hypothesized transient events scaled accordingly to illustrate our point. We assume uncorrelated Gaussian noise throughout. Figure 2 shows a schematic of two transient signals: one reaches ~ 3σ at three epochs (left) and another reaches ~ 2σ at five epochs (left). The measurements could be of a single pixel or a source integrated over a region. Either scenario in Figure 2 would pose a challenge to the single-epoch image-differencing method, i.e., by differencing against a deeper, higher S/N template image and examining the detections above some threshold. How high a S/N can we achieve by transforming the measurements to a new space formed by combining the epochs according to the χ^2 or *skew* metric? Furthermore, what is the minimum number of times a transient must persist (or be intermittently elevated) above some single-epoch S/N within a window in order to achieve an appreciable S/N for detection in the metric-space?

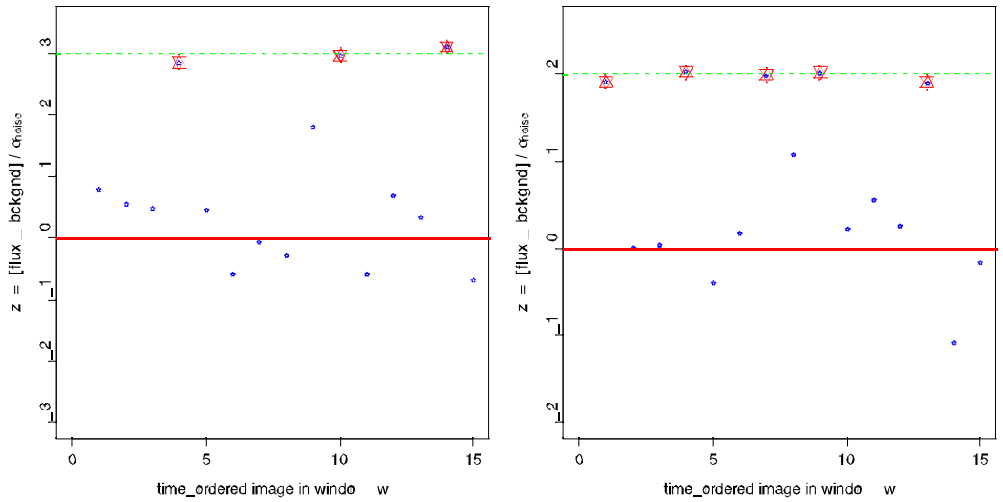

FIGURE 2. A window containing 15 simulated measurements of a transient where three are at ~ 3σ (*left*) and five are at ~ 2σ (*right*).

Figure 3 shows the results of our simulations for the χ^2 metric (Eq. 3). We considered a transient exhibiting 2, 3, 5, 7, 10, and 15 events (shown labeled) with single-epoch S/N running from 1 to 10 within a window of 15 measurements each affected by Gaussian noise ε ~ N(0,1). When the measurement is not elevated as an

"event" with some S/N, it is assigned a pure noise fluctuation at the zero baseline. We converted the single-epoch S/N and χ^2 values to equivalent probability measures assuming Gaussian statistics, i.e., as the probability of obtaining at least the observed values by chance. For a given number of events at some single-epoch significance, the χ^2 value obtained is actually a random variable, attaining a slightly different value for a new realization of the noise across the 15 measurements. Therefore, we show the mean χ^2 (high-tail) probability values (solid blue lines) and the 10 – 90 percentile ranges (error bars) obtained over 500,000 simulation trials. The single-epoch significance levels are equal to the χ^2 significance levels along the red-dashed line. Overall, the χ^2-combined measurements outperform the single-epoch measurements – effectively what one would obtain from the image-differencing method. One can see that for $N = 3$ single-epoch events hovering at S/N ~ 3 out of 15 measurements, the χ^2-combined measurements can attain a significance (probability of occurring by chance) of $\alpha <\sim 10^{-4}$. Can we do better?

Figure 4 shows our simulation results for the *skew* metric (Eq. 4) using the same method and inputs as for the χ^2. The only difference is a computational detail in how the probabilities are computed. While the distribution of a χ^2 random variable is well known, the distribution for *skew* when sampling from a *normal* population is not. We resorted to estimating probabilities from analytical fits to distributions for the sample *skew* derived from bootstrap resampling of a *normal* population. Figure 4 shows that the *skew* metric is more sensitive than the χ^2 metric (Figure 3) at detecting transients for the same range of single-epoch S/N levels and number of events that may occur at these levels. For example, an intermittent transient exhibiting S/N ~ 2 single-epoch events needs to occur on average $>\sim 5$ times out of 15 to give an average *skew*-metric significance of $\alpha <\sim 10^{-8}$. The χ^2 metric will require it to occur $>\sim 10$ times out of 15 to achieve the same level of significance. Pushing the *skew*-metric further, a S/N ~ 1 event will need occur $>\sim 7$ times out of 15 to give an average *skew*-metric significance of $\alpha <\sim 10^{-4}$. This is very encouraging. In general, the lower the single-epoch S/N, the *longer* a transient must persist at $>\sim$ S/N (or exhibit more events at or above this threshold) for it to be detected with a high significance in the epoch-combined metric space. Note that there may be other more sensitive metrics. From experimenting on several metrics, we found that the *skew* is the most sensitive at detecting low S/N transient events, presumably due to it's ability to detect slight asymmetries in an appropriately windowed, time-collapsed distribution of measurements relative to some long-run baseline.

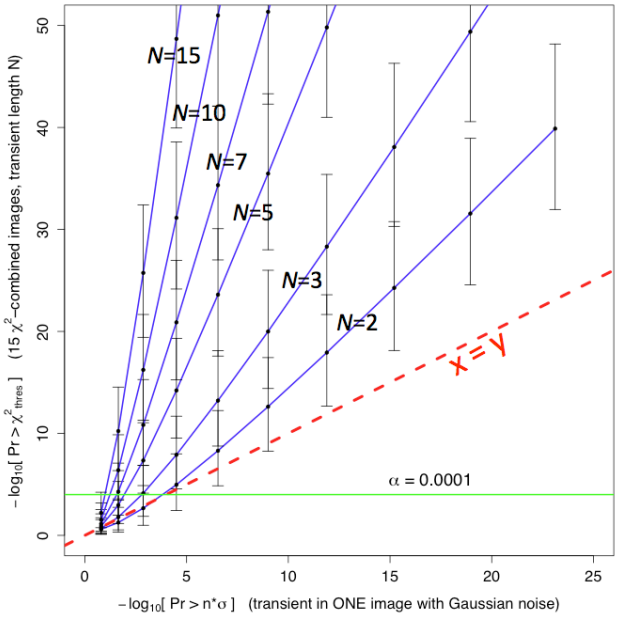

FIGURE 3. Simulation illustrating the significance (or effective sensitivity) of the χ^2 metric (Eq. 3) represented as the probability of obtaining a χ^2 value larger than that measured by chance if a sequence of 15 images contains a transient exhibiting N events with a single-epoch significance of n (= S/N) ≥ 1, 2, 3, … 10 running along the horizontal axis. Blue lines are average χ^2 probability values and the error bars span the 10 – 90 percentile range in probabilities obtained over 500,000 simulation trials for each N and n. The red dashed line is the line of equality.

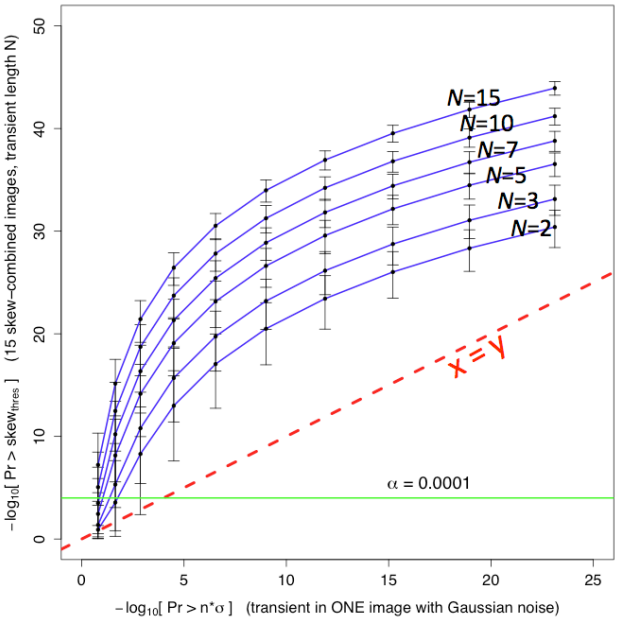

FIGURE 4. Simulation illustrating the significance (or effective sensitivity) of the *skew* metric (Eq. 4) represented as the probability of obtaining a χ^2 value larger than that measured by chance if a sequence of 15 images contains a transient exhibiting N events with a single-epoch significance of n (= S/N) ≥ 1, 2, 3, … 10 running along the horizontal axis. Blue lines are average *skew* probability values and the error bars span the 10 – 90 percentile range in probabilities obtained over 500,000 simulation trials for each N and n. The red dashed line is the line of equality.

3. PUTTING IT INTO PRACTICE

We have implemented the methodology outlined in §2 in a prototype software tool called *imtrandetect*. This tool is still in a developmental, alpha-testing phase, although it implements all crucial elements of the transient search algorithm with a few extra features to assist with reliability. As discussed, we only focus on the χ^2 and *skew* metrics, and the tool is being made flexible enough to run stand-alone on image data acquired in real-time. Details and software usage will be outlined in a future publication. Below we summarize some of the features. The rational for most of the steps was described above, with caveats and limitations outlined in §2.1.

3.1 Features and Processing Flow

The main processing steps in *imtrandetect* are shown in Figure 5. The most CPU-intensive steps are the first two: reprojection and interpolation of the input images onto a common sky grid, and the estimation/subtraction of a local background at low-spatial frequencies to ensure stationary pixel baselines versus time. The first of these may not be needed if the images are from fixed predefined survey fields and the telescope pointing is reasonably accurate. If the software is to be run on an image archive, all steps in Figure 5 are massively parallelizable, with certain steps being triggered as intermediate products become available (e.g., when a window's worth of data has been preprocessed). If processing on a incoming data stream in real-time, one will still have to pre-process a historical subset of archival data in order to obtain initial long-run estimates of the baseline-level and noise-sigma per pixel (last box on the top row of Figure 5). This "calibration" need only be done once, and perhaps refined later. The incoming image-data can then be processed serially as a new window's worth of data becomes available.

Some features of the *imtrandetect* tool are as follows.

- Overall, the tool emphasizes masking of instrumental artifacts through use of dynamic image masking of bright "static" sources and their artifacts.
- There is minimal impact from temporal and spatial PSF variations. Hence there are no spurious PSF-related residuals since no image-differencing is involved.
- There is the ability to combine images acquired simultaneously across different filters within a window, in order to further improve S/N.
- It can handle image data with irregularly-spaced observation times and large gaps provided one is aware of the limitations.
- It can handle images with non-uniform overlap (hence spatially-varying depth) across epochs, where it is assumed that images will be to be reprojected and registered prior to use.
- Generates light-curves that are photometrically calibrated if calibration information is available, otherwise internal relative photometry is performed.

- Generates image-cutouts of transient candidates through an image stack over a specifiable time-range, as well as the window-combined metric detection images.
- *Under development*: optional use of priors (e.g., light-curve templates) to assist with reliability, isolating specific transient candidates, or omitting undesired types.
- *Under development*: optional moving object (asteroid) filtering.
- Other constraints to maximize reliability: e.g., require n consecutive (or intermittent) events above some single epoch S/N spanning some Δt.

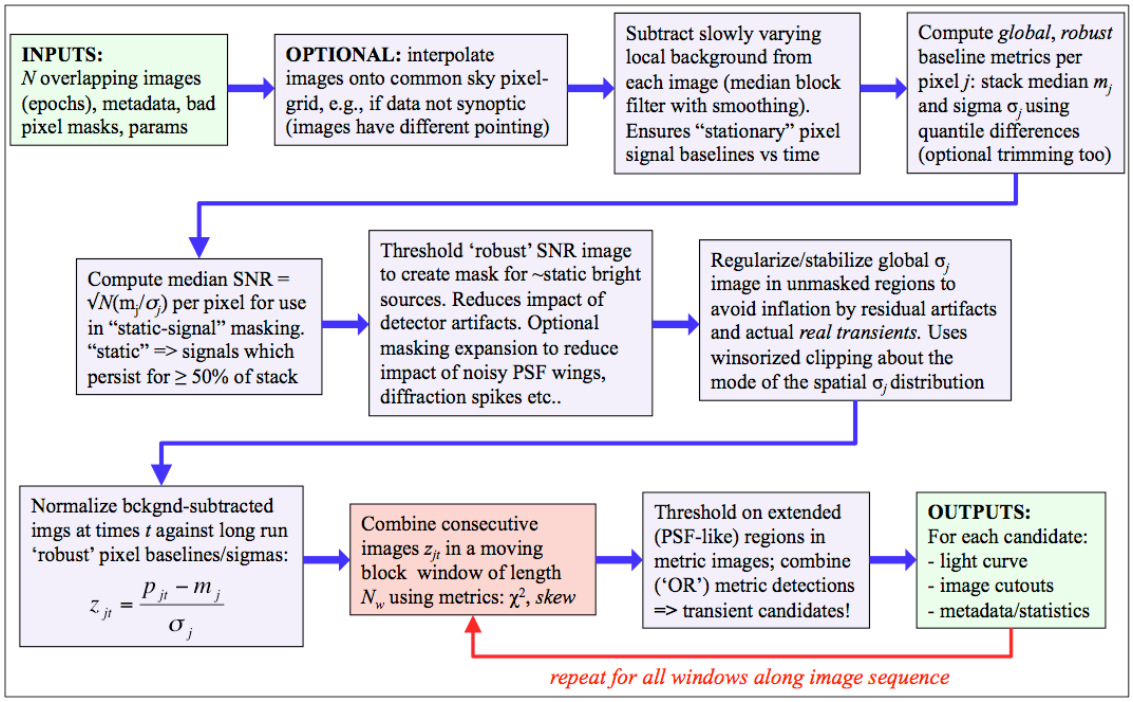

FIGURE 5. Processing flow in *imtrandetect* version 1.0. Details are expanded in §2 and 3.

4. TESTING ON CRTS DATA

We are currently testing our methodologies on optical image data acquired from the Catalina Real-Time Transient Survey (CRTS; Drake et al. 2009). The primary objective of CRTS is to search for Near-Earth Objects (asteroids), although there are parallel searches for SNe, CVs, Novae, and a wealth of other astrophysical transients and variables, both new and previously identified in other surveys. The search for SNe in particular has uncovered some rare types (Drake et al. 2011), a large fraction being extremely luminous with a tendency to favor very faint host galaxies. Our methodology is well suited to discovering these types since it relies on minimal contamination from host galaxy light, or other bright "static" underlying/nearby emission to avoid being masked for reliability purposes (see details in §2.1).

Our testing is very preliminary, although we have managed to recover several previously discovered SNe, e.g., Figures 6 and 7. Guided by the simulations presented in §2.2, a moving block window of 15 images was used throughout. We pushed down to an effective single-epoch S/N of ~ 3 and uncovered a false-positive rate of ~6%, comprising mostly instrumental glitches. This isn't too bad compared to other traditional approaches (e.g., image differencing) down to the same S/N level. We also uncovered a plethora of faint asteroids which at the time of writing, may or may not have been previously discovered. These are detected by virtue of their motion, however slight. The metrics are sensitive to differences in flux at fixed sky positions across an image stack. Objects moving at speeds of typically >~ an effective PSF width between epochs will inevitably appear "compact" and trigger a detection in the metric-image. They show up as "events" on a light-curve since the photometry is *forced* at the fixed sky location. Light-curves and thumbnails for two asteroids uncovered with *imtrandetect* are shown in Figure 8.

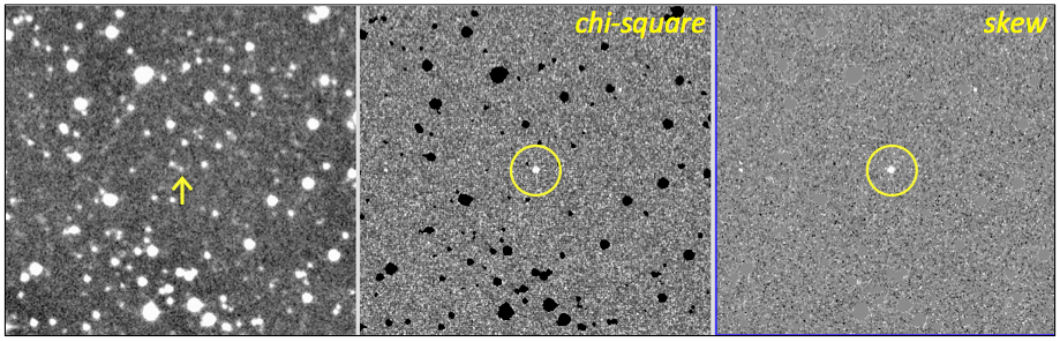

FIGURE 6. *left*: long-run median-combined image containing a barely visible galaxy, and *right*: Metric images from *imtrandetect* of a ~ 3′ × 3′ field centered on the type IIn supernova SN 2011cw discovered by CRTS on May 5, 2011. A running window of 15 images was used. See Figure 7 for light-curve and single-epoch thumbnails.

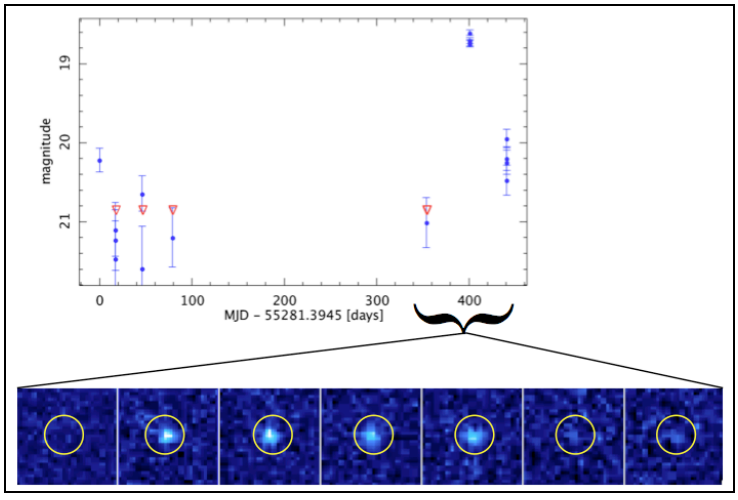

FIGURE 7. Light curve and single-epoch thumbnail images from *imtrandetect* of SN 2011cw detected off the metric images shown in Figure 6.

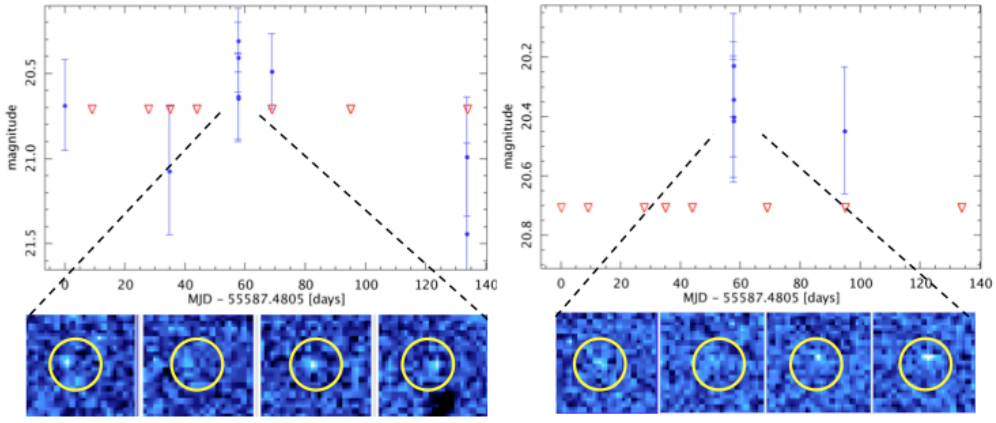

FIGURE 8. Asteroid candidates. Estimated speeds are ~15 and 9 arcsec/hr for the left and right objects respectively.

5. CONCLUSIONS AND CLOSING THOUGHTS

We have described a methodology and tool for optimally detecting low S/N (possibly intermittent) transient events from an incoming data stream or an image archive that may have escaped detection using traditional methods. The goal is not to replace existing methods but extend them (perhaps in parallel) to maximize the scientific returns of a dataset given the observational and technological limitations. We emphasize reliability over completeness since we are interested in detecting rare events on the surface of a sea of noise. Only by judiciously combining observations where a transient may be active do we stand a chance of going beyond what sequential single-epoch searches can offer.

Setting aside technological improvements, there may be other more optimal methods and metrics (in the maximal S/N sense) than what we presented here. We will continue the search. Furthermore, we plan to optimize and extend the *imtrandetect* tool with more functionality, in particular to enable the use of prior information to assist with reliability, weeding out "uninteresting" transients, and/or targeting a specific class of transient for further study.

In closing, we mention that optimizing transient searches to low S/N levels will severely strain efforts for follow-up. There is a deluge of transient candidates being generated in current synoptic surveys, even at moderately high S/N levels of >~ 10. For example, the CRTS imposes a threshold of ?X? for detection and only ~10% of transients are follow-up spectroscopically. This problem is getting worse and the number of potentially good candidates will inflate by orders of magnitude in the next generation of synoptic sky surveys. LSST for instance is expected to discover $> 5 \times 10^4$ Type Ia supernovae *per year* using multiband detection down to S/N ~ 10 (Bernstein et al. 2009). Therefore, it is unlikely one will expend (let alone be granted) valuable follow-up resources on a few low S/N events, unless of course they're truly exotic and genuine, and perhaps predicted theoretically. We advocate that the method described here is more apt to archival studies, where one has the luxury of applying contextual information, classification templates, models and matched filters to optimize transient searches of a given type or even new hypothesized types. Even

without rigorous follow-up on a per-case basis, archival research is still a very powerful way to proceed.

ACKNOWLEDGMENTS

The authors would like to thank the W. M. Keck Institute for Space Studies for establishing the initial collaborations that resulted in this work. CRTS is supported by NSF grant AST-0909182. FJM acknowledges support provided by NASA through a contract issued by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA.

REFERENCES

Alard, C. 2000, A&AS, 144, 363
Babu, G. J., Mahabal, A., Djorgovski, S. G., and Williams, R., 2008, Statistical Methodology, 5/4, 299
Bernstein, J. P., et al. 2009, LSST Science Book, ch.11
Bramich, D. M., 2008, MNRAS, 386L, 77
Cramer, H., 1946, Mathematical Methods of Statistics. Princeton: Princeton University Press
Dewdney, P. E., Hall, P. J., Schilizzi, R. T., and Lazio, T. J., 2009, IEEE, 97, 1482
Drake, A. J., et al. 2009, ApJ, 696, 870
Drake, A. J., et al. 2011, arXiv:1111.2566
Ivezic, Z., et al. 2011, arXiv:0805.2366
Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, Proc. SPIE, 4836, 154
Law, N. M., et al. 2009, PASP, 121, 1395
Pearson, E. S., 1931, Biometrika, 22, #3, p.423 (JSTOR 2332104)