# Statistical Properties of a Two-Stage Procedure for Creating Sky Flats

**Robert W. Crawford**
Rincon Ranch Observatory
Tucson, AZ

**Mark Trueblood**
Winer Observatory
Sonoita, AZ

**Abstract**: Accurate flat fielding is an essential factor in image calibration and good photometry, yet no single method for creating flat fields is both practical and effective in all cases. At Winer Observatory, robotic telescope operation and the research program of Near Earth Object follow-up astrometry favor the use of sky flats formed from the many images that are acquired during a night. This paper reviews the statistical properties of the median-combine process used to create sky flats and discusses a computationally efficient procedure for two-stage combining of many images to form sky flats with relatively high signal-to-noise ratio (SNR). This procedure is in use at Winer for the flat field calibration of unfiltered images taken for NEO follow-up astrometry.


**References**:
1. See URL http://www.winer.org for additional information
2. URL http://www.minorplanetobserver.com/astlc/PhotometryGuide.htm
3. R. Berry and J. Burnell, "The Handbook of Astronomical Image Processing" (Willmann-Bell, Inc., Richmond, VA, 2000).
4. P. Massey, "A User's Guide to CCD Reductions with IRAF" (NOAO, Tucson, AZ). http://www.noao.edu/kpno/manuals/dim/dim.html (1997).
5. M. Newberry, "Signal-to-Noise Considerations for Sky-Subtracted CCD Data", PASP, 103, 243 (1991).
6. T. Spahr, private communication (2004).
7. H. Cramer, "Mathematical Methods of Statistics" (Princeton University Press, Princeton, NJ, 1961, p. 369).

## 1. Introduction

Good flat fielding is important to the science programs, both ongoing and planned, at the Winer Observatory. The Observatory was incorporated in 1983 in Maryland under Internal Revenue Service rules for non-profit public charity organizations performing scientific research, and its research program initially focused on lunar and minor planet occultations. The Observatory moved to Arizona in 1990 and, with completion of the Sonoita Field Station in October 1997, now provides facilities and services under dark southeastern Arizona skies for robotic telescopes operated primarily by universities for faculty and graduate student research and undergraduate instruction [1].

A collaboration involving the authors is establishing a new research program (not yet fully operational as of this writing) directed toward astrometric follow-up of Near Earth Objects (NEOs) listed on the Minor Planet Center's (MPCs) Near Earth Object Confirmation Page. Once adequate experience is gained with NEO follow-up, we intend to expand into other research areas. Topics of mutual interest include differential photometry for minor planet rotational studies, eclipsing binary star lightcurves, and exoplanet searches using photometric techniques.

Our NEO astrometry program has available to it two telescopes (see Table 1). The Rincon telescope is a 0.5-m f/9 corrected Newtonian on a robotic German equatorial mount using a commercial CCD camera with a thinned, backside illuminated detector. The Winer telescope is a 0.5-m f/8 Ritchey-Chretien on a robotic altitude-azimuth mount, designed and built at the University of Iowa and acquired from them in 2003 after installation of a new Iowa telescope. The Winer telescope also features a commercial CCD camera with a thinned, backside illuminated detector.

**Table 1. Telescopes and Equipment**

| | **Rincon** | **Winer** |
|---|---|---|
| **Optical Configuration** | f/9 Corrected Newtonian | f/8 Ritchey-Chretien |
| Aperture | 0.5 m | 0.5 m |
| Focal length | 4.57 m | 4.06 m |
| Image Scale | 21.5 $\mu$m/arc second | 19.6 $\mu$m/arc second |
| **Mount** | German Equatorial | Alt-Azimuth |
| **CCD camera** | commercial | commercial |
| Detector | thinned, back illuminated non-antiblooming | thinned, back illuminated non-antiblooming |
| Image Format | 1k x 1k pixels | 1k x 1k pixels |
| Pixel Scale | 1.09"/pixel | 1.22"/pixel |
| Field of View | 19 arc minutes | 21 arc minutes |

Image calibration is covered extensively in the literature, from which merely three references are the Collaborative Asteroid Lightcurve Link (CALL) photometry tutorial [2], Berry [3] and Massey [4]. Newberry [5] has shown that a flat field must have a substantially higher signal-to-noise ratio (SNR) than any pixel of interest if the image calibration is to avoid degrading the SNR of the data. Thus, to obtain 1% photometry the flat field must have a precision[1] in the range of 0.25% to 0.5%. Obtaining accurate flat fields of high precision is notoriously difficult. Doing so requires illuminating the telescope pupil with a source of illumination that varies spatially (over the detector) by less than the desired precision of the result and that has a spectral distribution appropriate to the intended targets.

At Winer, broadband photometry will be done initially only to obtain NEO magnitude estimates for submittal with astrometric positions to the MPC. NEO photometry of this type is often done in unfiltered "white" light and with only approximate calibration to a broadband color magnitude (often of uneven quality itself) found in astrometric catalogs. The MPC's stated goal

[1] The term precision is used in the sense of the repeatability of a measurement. A precision stated as a percentage (e.g., 1%) is equivalent to a signal-to-noise ratio equal to the reciprocal (e.g., 100).

is to obtain photometry at the 5% level [6], but the quality of the available observations is often less. Such a tolerance is considerably less demanding than that of stellar photometry, where 2%, 1%, or even greater precision is often needed. For example, the CALL Web site states a minimum requirement of SNR ~ 50 (2% precision) and a preference for SNR ~ 100 (1% precision) in the differential photometry used for determining rotational periods. The goal at this stage in our research program is to work toward the capability of performing 1% photometry, for which flat fields of 0.25% to 0.50% precision will be required.

Many methods are used to create high SNR flat fields, although no single method appears to be both practical and effective in all circumstances. Dome flats are a common solution to the flat field illumination problem at professional observatories. Great efforts are made to achieve uniform field illumination, using special screen materials or paints, and special lamps are selected to provide the correct spectral distribution. In return, flat fields of very high SNR can be obtained in a few images. Light boxes are a conceptually similar, but less sophisticated, solution to the flat field problem at amateur facilities. Twilight flats may also be used, as are sky flats in some research programs, and there are methods for combining dome flats with twilight or sky flats to improve the uniformity of field illumination. Massey (op cit) discusses a range of alternatives for obtaining satisfactory flat fields at the community-use telescopes at Kitt Peak National Observatory.

At Winer Observatory, we are experimenting with sky flats created from the science images taken in our NEO astrometry program. Sky flats are nothing more than composites of many images that are formed using a median-combine process to reject pixels that are illuminated by bright objects (e.g., stars and galaxies) in the individual images. Sky flats are similar in principle to dome and twilight flats – i.e., illuminate each pixel with the same light intensity, in this case generated by the sky background, and record the response of the telescope and detector system. The assumption made is that, over the field of view, the sky background does not vary by more than the precision we seek. Unlike twilight and dome flats, however, the pixel counts from the sky background are small, and their Poisson variation relatively high, so the SNR of the sky background in a single image is too low to be of use as a flat field. To ensure that we achieve a sufficient SNR at every pixel, without contamination from stars, several to many individual images are combined to form the sky flat.

Sky flats have several advantages from our current perspective, in that they avoid the need for dome flat equipment and can be created in a data processing pipeline using the science images taken robotically during a night. Another advantage is that their color automatically matches the characteristics of the night sky, giving us the best color match for images taken to detect faint objects against the sky background. Twilight flats are basically sky flats taken at twilight with much higher illumination levels, but the twilight sky has a different color than the night sky.

We acknowledge that sky flats may be less suitable for color filter photometry. The spectrum of the night sky is not continuous, but is characterized by naturally occurring emission lines and by urban light pollution in both narrow (low pressure sodium vapor lighting) and broad (high pressure lighting) spectral bands. Thus, the night sky may not provide spectrally flat illumination across the filter passbands of interest, and the resulting sky flats may be a poor match to the colors of astronomical targets or of comparison and check stars. The problem of non-uniform spectral distribution should be of less concern the narrower the filter passband of interest.

An obvious disadvantage of sky flats is that the SNR of the sky background in a single image will be low. Thus, many images must be combined to boost the SNR to a precision of 0.5% or better. This problem will be exacerbated by the reduction in photon counts when observing through broadband filters, but the reduction can be offset by increasing the number of images used in the creation of the flats. The study summarized in this paper was conducted to understand the statistical properties of sky flats and to develop practical computational techniques for creating sky flats from large numbers of images.

## 2. Basic Properties of Sky Flats

Sky flats are conventionally created using the median operator to combine N astronomical images in a single step. Each pixel value in the output image is defined as the median value of the N individual values taken on by that pixel in the input images. Before the combine, the input images are first normalized to the same brightness level to account for variations in the brightness of the sky background across the fields. Under suitable conditions, the resulting output image will record the response of telescope and detector to uniform illumination by the sky background.

Any of several procedures can be used for the normalization before images are combined, including equalizing the average pixel value across the image (or within the central portion of the image) or equalizing the median or modal (most frequent) pixel value. Because pixels illuminated by sources other than the sky are to be rejected by the median operator, a normalization method (such as median or modal) that equalizes the sky background brightness is to be preferred over methods based on average values that are affected by the pixel counts of astronomical objects.

The basic properties of sky flats are easily understood. As a simplification, let us consider that images are composed of two types of pixels – *sky pixels*, which are illuminated by sky background, and *non-sky pixels*, which are illuminated by astronomical objects or cosmic rays. Sky and non-sky pixels are distinguished by the fact that non-sky count levels typically far exceed sky count levels. A particular output pixel in the composite image will be assigned a sky value by the median-combine process whenever a majority of the input images have a sky value in that position. It will be assigned a non-sky value whenever a majority of the input images has a non-sky value in that position.

As an example, if there are 21 images to be combined, the output pixel will take on a sky value if that position is a sky pixel in 11 of the input images (or more). If there are 22 images, then the output pixel will take on a sky value if it is a sky pixel in 12 images or more (because the median of an even number of values is defined as the average of the two central values – the $11^{th}$ and $12^{th}$ in this case). On the other hand, chip defects such as hot or dead pixels will flow through to the output image, because such pixels will be consistently high or low in the input images, and appropriately characterize the (non-linear) detector response at those positions.

**Rejection of Non-Sky Pixels**. An important property and desired of the median-combine process is that, under suitable circumstances, it will reject astronomical objects and cosmic ray hits with high statistical confidence and yield a composite image composed of the system response to a spatially uniform sky background. The requisite circumstances are that:

1. The images are sufficiently displaced from each other spatially that the values taken on at a particular pixel position are uncorrelated across the input images.
2. A minority of the values at each pixel position are non-sky pixels.

3. The gradient in the brightness of the sky background across each image must be less than the precision desired in the final result. (Differences among images in the average sky background level can be removed by the pre-normalization.)

The first requirement (spatial displacement) can be satisfied when the image set consists of star fields imaged after independent pointings, even when the fields are of the same target, because random pointing errors of ~1 arc minute or larger in amateur equipment will be sufficient to break pixel-scale correlations. The technique is much less likely to reliably reject non-sky pixels when the images are composed of extended objects or contain bright objects consistently positioned near the centers. Nor will the technique reject non-sky pixels when the image set follows the same field throughout the night without repointing.

The second requirement (a minority of non-sky pixels) must be met at every pixel position if the output image is to be completely free of contamination from non-sky pixels. Although it is minimally sufficient that a pixel position take on a bare minority of non-sky values in the image set, the rejection process has much greater power when the non-sky pixels comprise a small minority in each image.

The third requirement (small brightness gradient) will be easier to meet in images that have small fields of view, are not taken near bright objects, and are positioned away from sources of light pollution on the horizon. When this requirement is violated, the sky flat will suffer from systematic errors caused by the gradient in field illumination. It may be possible to correct for this problem, but it would be necessary to deconvolve the gradient in sky illumination in the images from the potential gradient in pixel response across the chip. We have not, as yet, found a treatment of this issue in the literature, and we plan to reassess this issue empirically after a suitable dataset of sky flat frames has been obtained.

The basic statistical properties of sky flats are not difficult to assess quantitatively as an application of the binomial theorem, under the simplifying assumption that images are composed of sky and non-sky pixels that occur with equal probability at each pixel position. For a single output pixel, the probability $prob_{NS}$ that a non-sky pixel survives to the output image is given as follows. Let p be the fraction of non-sky pixels in an image and 1-p the fraction of sky pixels; let N be the number of images combined, where N = 2n + 1; and let m be the number of sky values that may occur. Then:

$$prob_{NS} = \sum_{m=1,\ldots,n} C_{N,m} (1-p)^m p^{N-m} \qquad (1)$$

where $C_{N,m}$ is the combinatorial operator giving the number of ways that N things can be chosen m at a time. The terms in the summation give the probability of encountering m sky pixels and N-m non-sky pixels at a given pixel location. The summation runs from m=1 through m=n because these are the cases in which the sky pixels are a minority of the values present in the input images and the median operator selects a non-sky value for the output.

The statistical reliability of the process can be measured by the probability $prob_{NS}$ that the median-combine fails to reject non-sky values at a given pixel. If there are k pixels in an image, then k * $prob_{NS}$ is the expected number of non-sky pixels present in the output image. This value can be thought of as a contamination rate and used as an overall measure of the reliability of the process.

Figure 1 shows that the expected number of non-sky pixels surviving the median combine operation is highly sensitive to p, the fraction of non-sky pixels in an image. The values given are for a 1k x 1k CCD image, but may be scaled to other image formats in proportion to the total number of pixels. When 11 images are combined, approximately 100,000 pixels in the output image will be non-sky pixels if p is as large as 0.30. For 11 images, the number of non-sky pixels falls to 10,000 when p equals 0.20, and to 300 (< 0.05%) when p equals 0.10. The contamination rates fall very quickly as the number of images increases. As long as $p < 0.50$, the number of non-sky pixels in the output can theoretically be driven to low values by combining a sufficient number of images. However, the median-combine process is most effective and reliable when p is small.

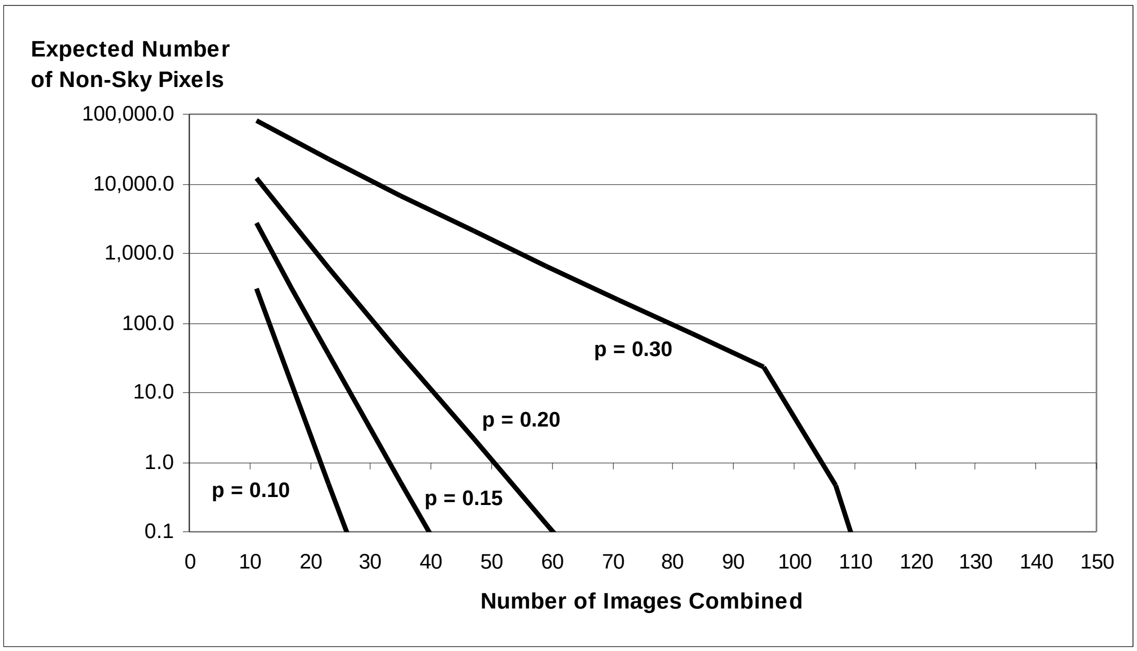


**Figure 1. Statistical Reliability for Rejecting Non-Sky Pixels (1k x 1k image)**

The fraction of non-sky pixels in an image will vary with a number of factors including the astronomical target (star field, cluster, extended object, etc.), galactic latitude, image scale, seeing conditions, and depth of exposure. While we cannot provide general guidelines on the values of p that will be encountered in astronomical imaging, based on our experience to date we believe that p values in the range 0.10 to 0.15 are typical of the non-cluster star fields imaged in the course of NEO studies, using our equipment and exposure times of 60 to 120 seconds. Even if p is as large as 0.20 in some cases, we are easily able to acquire many more than the 60 images in a night that are needed to reduce the contamination rate below 1 pixel per frame. Therefore, we should be able to acquire sufficient images to form sky flats composed only of sky pixels.

**Variance Reduction**. A second important property of the median-combine process is its ability to attenuate the variance that is present in the input values. As a result, the output image has reduced variance and higher SNR than was present in the sky backgrounds of the input images. Variance reduction occurs because the median operator is an estimator for the central tendency of the data, much like the mean value operator. If $s^2$ is the variance of the input values, then the variance of the median is given asymptotically by [7]:

$$\sigma^{2}_{median} = (\sigma\ / 2\ )\ \pi^{2} / N \tag{2}$$

Compare this to the familiar result for the mean value operator:

$$\sigma^{2}_{mean} = \sigma^{2} / N \tag{3}$$

As a result, the standard deviation $\sigma$ of a median value will be about 25 percent higher than the standard deviation of an average value computed from the same data. Equation (2) is valid asymptotically; there is no precise point at which the asymptote is approached, but a sample of 30 or more is often considered to be sufficient.

Combining more images will reduce the standard deviation of the result in proportion to $1/\sqrt{N}$ and increase the SNR in proportion to $\sqrt{N}$. To demonstrate this, we conducted a Monte Carlo simulation in which pixel values were chosen randomly from a Poisson distribution to fill a pixel sample of size N. The mean and median values were computed for each sample and the results stored. The process was then repeated 3000 times so that the standard deviation of the median and mean values could be accurately determined for the sample size. Sample sizes from N=5 to N=300 were investigated. The SNR is estimated for each sample size as the ratio of the mean (or median) pixel count divided by the standard deviation of the count across the repeated samples. Only the Poisson variation in the sky background was considered, and noise sources related to read noise and dark current in CCD images were excluded.

Figure 2 presents the results of the Monte Carlo simulation, showing the increase in SNR as a function of the number of images combined and the method of image combination. The vertical axis measures the SNR relative to that for a single image. For example, the lower line in the figure shows an increase in SNR of approximately 13 for N=300 images. Therefore, if the sky background in one image has an SNR of 10, the sky background of a median-combined sky flat formed from N=300 images would be approximately 130. The upper line shows the result if images were combined using the mean value operator (averaging). The mean value operator is more efficient at variance reduction, but it would not reject non-sky pixels[2]. For samples of N=30 images or more, the variance of the median-combined image is 25 percent greater than would result from averaging, in accordance with Eqs. (2) and (3). This gap can be thought of as the price one pays to reject non-sky pixels.

The sky background in unfiltered images taken from our location can have a SNR ~ 50. Therefore, we can create sky flats with SNR ~ 200 (0.5% precision) by combining 25 images, and flats with SNR ~ 400 (0.25% precision) by combining 100 images. When imaging through broadband filters (or with smaller apertures), the SNR in single images will be much less, and we must be prepared to combine larger image sets to achieve comparable results. These results suggest to us that we should be able to achieve satisfactory SNRs in sky flats for both current and future programs. However, such flats will still fall short of the very high SNRs that can be achieved with dome or twilight flats (not withstanding issues regarding uniform illumination and color match).

---

[2] There are iterative techniques (often called "sigma clipping") in which outliers (both high and low) are rejected based on their distance from the mean value until all remaining values are statistically consistent. Such techniques would allow image combination and rejection of non-sky pixels using an averaging process. We do not consider these techniques here.

**Computational Complexity**. Tests conducted during the development of a data processing pipeline for our research program showed that combining more than about 100 images was a computational bottleneck. When the number of images being combined exceeds a threshold that is specific to the configuration of each computer system, slow virtual memory (disk) is used in place of fast physical RAM. Beyond that threshold, the time required to complete the median combine operation escalates rapidly as increasing use is made of virtual memory.

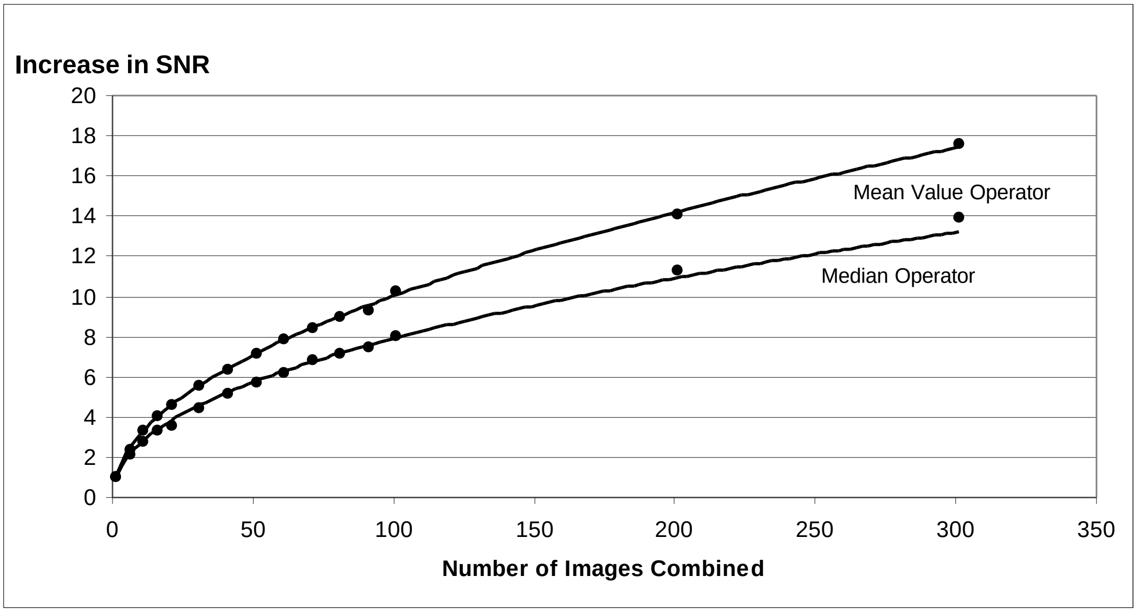


**Figure 2. Increase in SNR Through Image Combination (relative to a single image)**

The data in Table 2 suggest that the threshold for our computer system (1.2 GHz processor, 512 MB RAM) is approximately 70 images. The processing time per image is essentially constant through at least 70 images, but the processing time and disk access activity escalate rapidly as substantially more images are processed. These values are specific to our processing computer, and each computer will have its own curve with a threshold that depends primarily on the ratio of available RAM to image size.

## 3. Properties of a Two-Stage Process

While sky flats can give acceptable SNRs when formed from many images, a single-stage median-combine process may be impractical. This discovery led to development of a two-stage process for performing the median combine that is more computationally efficient. In the first stage, one combines images in n sets of smaller size, with the set size chosen to be large enough to reject non-sky pixels with high probability, but small enough to avoid use of virtual memory. If non-sky pixels are reliably rejected in the first-stage, then the n first-stage images may be combined in a second-stage by averaging (or summation), without loss of precision in the final result compared to a one-stage combine. If non-sky pixels are not rejected with sufficient reliability in the first stage, the n first-stage images can be combined again using the median operator, although with additional loss of precision as shown by simulation later in this section.

| Table 2. Dependence of Processing Time on Number of Images Combined | | |
|---|---|---|
| **Number of Images** | **Total Processing Time** | **Processing Time per Image (minutes)** |
| 30 | 1 min 30 secs | 0.05 |
| 70 | 3 min 15 secs | 0.05 |
| 100 | 8 min 4 secs | 0.08 |
| 150 | 22 min 45 secs | 0.15 |
| 300 | 1 hr 57 mins | 0.39 |

Based on characteristics of our images and the capability of our processing computer, we elected to combine images in batches not smaller than 35 and not larger than 70. Figure 3 shows how this so-called "35-70 Rule" divides the total number of images into an increasing number of groups of size always greater than 35. When the probability of non-sky pixels does not exceed p = 0.15, median combining a group of 35 images produces a frame that, on average, will contain 0.5 pixel contaminated with a non-sky value (for a 1k x 1k chip). We found this result to be acceptable as a worst-case condition, and we expect a better outcome (lower probability of contamination) in most circumstances. When p = 0.10, only 1 frame in 1000 resulting from the median-combine of 35 images will be contaminated with a non-sky pixel, and in most cases, more than 35 images will be combined in each group.

The n frames produced in the first stage are combined in a second stage to produce a final sky flat. Under conditions typical of our asteroid images, first stage frames produced using the 35-70 Rule can be combined by averaging while running only a small risk that the final sky flat is contaminated with non-sky pixels. If the second stage averages n=5 frames, each of which has a 1-in-1000 chance of 1 contaminated pixel, then the final sky flat will, on average, have a 1-in-200 chance of 1 contaminated pixel. Under worst-case conditions (p=0.15 and 35 images), the final sky flat would, on average, have 5 * 0.5 = 2.5 contaminated pixels out of ~ 1 million pixels in the image. We judged these outcomes to be acceptable. Under less favorable circumstances, one can median-combine the frames in the second stage to increase the rejection of non-sky pixels, but the SNR of the final sky flat will be reduced.

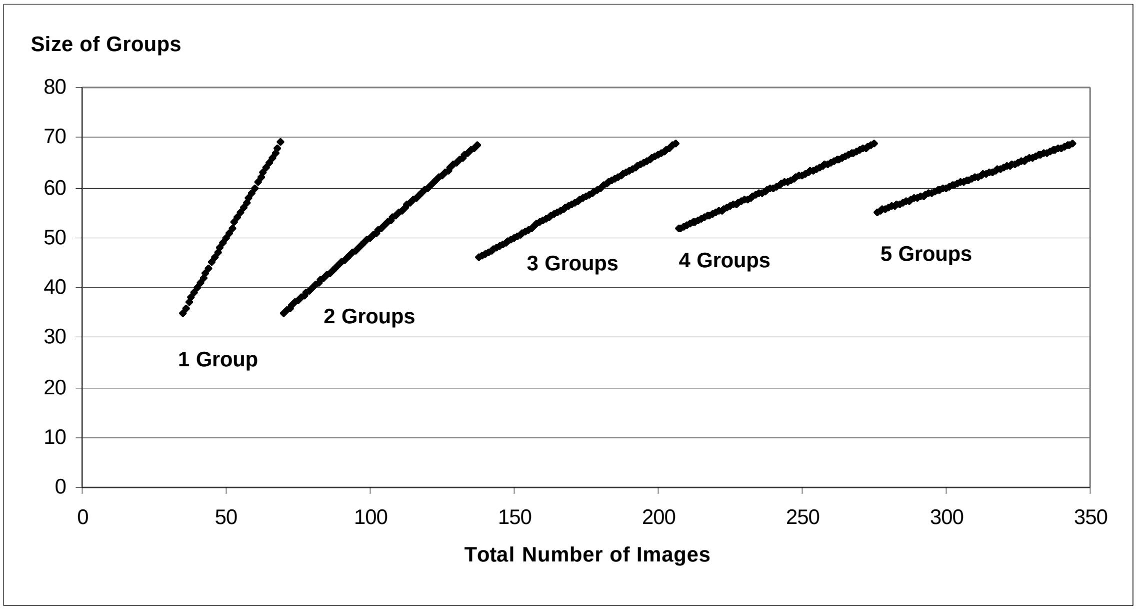


**Figure 3. Size of Groups in first Stage Median Combine (based on 35-70 Rule)**

One motivation for the 2-stage approach was the realization that a median-combine followed by averaging in the second stage would produce a final sky flat with the same SNR as if all the images were median-combined in a single step. This seemed intuitive because the median-then-average approach applies the median operator once, and only once, to all of the data. A second Monte Carlo simulation was conducted to confirm this property. In the simulation, pixel values were combined in groups according to different methods (see Figure 4). The SNR ratio of the 2-stage median-then-average combination process is found to be 25 percent less than that for simple averaging of the images, which was the result seen previously for a 1-stage median combine (Figure 1). Applying the median operator twice in a median-then-median combination process results in another 25 percent reduction in the SNR of the final sky flat.

## 4. Conclusions

This paper has presented the considerations and analysis that led us to experiment with sky flats for out science program at Winer Observatory. Sky flats have the considerable practical advantage to us that they can be created from images already available without the need for additional equipment or manual intervention. With a research program that acquires a large number (N = 100 to 300) of 60- to 120-second exposures in an evening, it appears possible to achieve sufficiently high SNR values for unfiltered observations. We are also hopeful of obtaining satisfactory SNR values when images are taken in a few (2 or perhaps 3) broadband filters during a night. The color match of sky flats to the nighttime sky background should also be advantageous to the sky-limited detection of faint objects in NEO astrometry and for the differential photometry of relatively faint objects against the sky background, as in asteroid light curve studies.

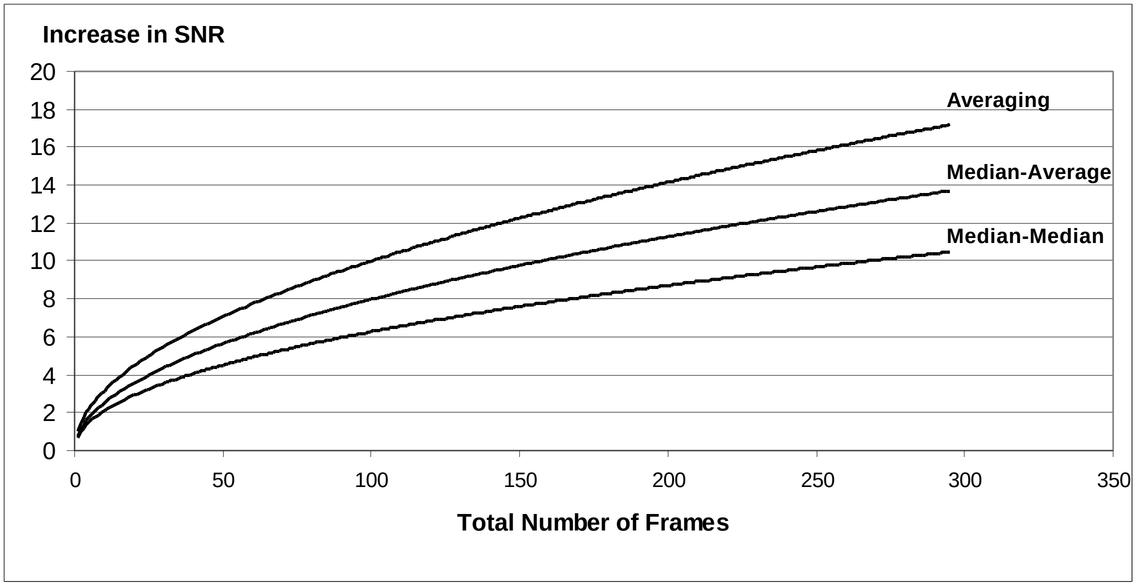


**Figure 4. SNR Increase in a Two-Stage Image Combination Process (relative to a single image)**

The information we present on the statistical properties of sky flats was developed using a simplified analytical treatment and is not a definitive treatment of the problem. It is intended to provide guidance on the selection of parameters to other amateurs whose observing programs might benefit from experimentation with sky flats. Because the treatment is not definitive, users of the results are encouraged to build in a safety margin on the number of images that will be combined. Those engaged in color filter photometry should also take care to achieve a satisfactory match to the colors of objects of interest and to ensure sufficient sample sizes in each passband so that reliable transformation coefficients and color terms can be derived.

## Acknowledgements

The authors wish to acknowledge a fruitful discussion on the topic of sky flats with Dr. Mark Dickinson of the National Optical Astronomy Observatory. Author M. Trueblood wishes to thank the many generous supporters of the Winer Observatory listed on the observatory's website Honor Roll of Donors.