Principled point-source detection in collections of astronomical images

Dustin Lang[1,2,3], David W. Hogg[4,5], & Others

ABSTRACT

We review the well-known *matched filter* method for the detection of point sources in astronomical images. This is shown to be optimal (that is, to saturate the Cramér–Rao bound) under stated conditions that are very strong: an isolated source in background-dominated imaging with perfectly known background level, point-spread function, and noise models. We show that the matched filter produces a maximum-likelihood estimate of the brightness of a purported point source, and this leads to a simple way to combine multiple images—taken through the same bandpass filter but with different background levels and point-spread functions—to produce an optimal point source detection map. We then extend the approach to images taken through different bandpass filters, introducing the *SED-matched filter*, which allows us to combine images taken through different filters, but requires us to specify the colors of the objects we wish to detect. We show that this approach is superior to some methods traditionally employed, such as the *chi-squared coadd*, and that other traditional methods can be seen as instances of SED-matched filtering with implied (and often unreasonable) priors. We suggest a set of SED filters and how one ought to proceed: lower the detection threshold to ensure that all sources of interest are detected, then fit for their properties and keep or discard sources based on their utility for the question at hand.

Subject headings: methods: statistical — techniques: image processing

FIXME – are we going to talk about detecting galaxies? Seems like we should!

FIXME – should we talk about the probability of selecting the wrong pixel as the peak?

FIXME – need some consistent way of labelling image quantities, and sensible naming and careful use of "flux" and "counts". Some notation for "estimates for" or "maximum-likelihood point estimate of" would be useful too. Hats?

---

[1]To whom correspondence should be addressed; dstn@cmu.edu

[2]McWilliams Center for Cosmology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

[3]Department of Physics & Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

[4]Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, New York, NY 10003, USA

[5]Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

1. Introduction

There are few operations in astronomy more important than the detection of stars or point sources. Indeed, many astronomical discoveries come down to point-source detection. What is the best method for performing such detection? Here we answer that question, in the limited context of isolated sources, uniform sky-limited noise, and well-understood point-spread function. Even in this limited context, the subject is rich and valuable; more general—and more difficult—cases will be illuminated if we can understand the simplest case first.

Fundamentally, when much is understood about a signal latent in noisy data, the best detection methods are (or look like) *matched filters*. A matched filter is a model of the expected signal with which the data are *cross-correlated*. Peaks in the cross-correlation are candidate signal detections. In astronomical point-source detection, the expected signal is the point-spread function (PSF), and the cross-correlation operation is often wrongly called "convolution by the PSF". Matched filters are well used in astronomy, in contexts ranging from spectroscopy (CITE) to ultra-faint galaxies (CITE) to gravitational radiation (CITE). [ADD ADDITIONAL REFS]

In what follows, we will argue for matched filtering for point-source detection. This is not new (CITE); what is new is that we consider the common context of heterogeneous (in point-spread function and sensitivity) multi-epoch, multi-band imaging. While the optimality of matched filtering for single-image point source detection is well known by astronomers, the straightforward mathematics behind it is often not, leading to a misconception that it is simply an algorithmic choice. To our knowledge, the optimal* method of combining multiple images with different properties to obtain the greatest possible sensitivity to point sources has not been written down in the astronomical literature. The mathematics are straightforward and the resulting procedure is simple and computationally inexpensive.

Perhaps more controversially, we go on to argue that when imaging in multiple bands is available, one should again use a matched filter, now matched to the spectral-energy distribution (SED) of the sources to be detected. This SED-matched filtering, as we will call it, makes explicit the assumptions that are implicitly embedded in any method that attempts to detect sources by combining imaging from multiple bands.

In the Real World, astronomers never precisely know their point-spread function, their noise model, their flat-field (or other calibration parameters), nor the spectral-energy distributions of the sources of greatest interest. Also, often, the sources of interest aren't point sources or perhaps vary with time. In these cases, we advocate parameterizing ignorance, and operating with the union of all possibly appropriate matched filters. We will fully execute this idea here when it comes to spectral-energy distributions, but there are natural extensions to deal with point-spread function, noise-model, calibration, and time-domain uncertainties. DSTN: WE SHOULD RETURN TO

*In this paper we use "optimal" in a technical sense, of saturating the Cramér–Rao bound under stated assumptions. We encourage other authors to do the same!

THESE ISSUES IN THE DISCUSSION.

A "traditional" approach for detecting sources in multi-epoch imaging is to co-add the images and then run a detection algorithm on the resulting coadd. When the images have different point-spread functions or noise properties, this method results in needless loss of sensitivity; producing a coadd effectively forces the use of a *mismatched filter* rather than a matched filter. We will show that the correct procedure involves creating a weighted co-addition of matched-filtered (smoothed) images.

That is, in what follows, we will detect sources as above-threshold pixels or regions in a weighted co-add of PSF-correlated input images. We will call this object a "detection map". This detection map is the best thing to use for source detection. Once sources are detected, of course, the detection map should be put aside, and source properties (positions, colors, and so on) ought to be measured (inferred) from the raw pixels in the collection of input images via a likelihood function. That measurement and likelihood function is beyond the scope of this paper, but the subject of a parallel research program (CITE).

2. Our image model

We consider astronomical images such as those obtained from an (idealized) CCD. Aside from noise, each pixel contains a signal proportional to the number of photons incident upon it during the exposure, modulated by a bandpass filter. We will consider a discrete image made up of a square array of pixels $\boldsymbol{j}$, each of which has signal (or counts) $C_{\boldsymbol{j}}$, where we use $\boldsymbol{j}$ as a two-dimensional focal-plane position, measured in integer pixel units.

We will make many strong assumptions:

- that the noise (coming from such sources as the Poisson distribution of the sky background, dark current, and read noise) is zero-mean, Gaussian, pixelwise independent, and of known variance. The zero-mean assumption can be seen as assuming perfect background subtraction (sky estimation);
- that the image is well sampled;
- that the (perhaps tiny) image contains *only* one point source with some (unknown) flux and position;
- that the source is centered within a pixel (we will relax this later);
- that the point-spread function is spatially constant (across the possibly small image patch of interest) and known perfectly;
- that the photometric calibration of the image is known perfectly; that is, that it is possible to map from "count" units back to physical units or a photometric standard;

- that the image is not contaminated by cosmic rays, stray light, bad pixels, bad columns, electronic artifacts, or any other of the many defects in real images.

Throughout this paper we assume a "pixel-convolved" PSF;[†] we consider the point-spread function to *include* the effects of pixelization. In well-sampled images, we think of the image as being a continuous function which, after being correlated by the PSF, is sampled at delta-function pixel locations. There is no need to think of pixels as "little boxes"; they are simply samples of an underlying smooth two-dimensional signal.

With these strong assumptions, we can write down the probability distribution of each pixel value, which allows us to prove the optimality of the method. Specifically, if the image contains a single point source, centered on the pixel at position $\boldsymbol{k}$ and with constant flux resulting in a total number of counts F, then the image is

$$C_{\boldsymbol{j}} \quad = \quad F\,\psi_{\boldsymbol{j}-\boldsymbol{k}} + e_{\boldsymbol{j}} \quad , \tag{1}$$

where $\psi(\boldsymbol{j}-\boldsymbol{k})$ is the point-spread function evaluated at offset $\boldsymbol{j}-\boldsymbol{k}$ and $e_{\boldsymbol{j}}$ is per-pixel noise drawn from a zero-mean Gaussian with known per-pixel variance σ_1^2. We can also write this as

$$C_{\boldsymbol{j}} \quad \sim \quad \mathcal{N}\big(F\,\psi_{\boldsymbol{j}-\boldsymbol{k}}\,,\,\sigma_1^2\big) \quad , \tag{2}$$

meaning that $C_{\boldsymbol{j}}$ is drawn from a Gaussian distribution with mean $F\,\psi(\boldsymbol{j}-\boldsymbol{k})$ and variance σ_1^2.

3. Detecting a point source in a single image

Here we consider a point source k in a single image, as in equation 1.

The *matched filtering* operation, also known as "smoothing by the PSF" or "convolving by the PSF" can be written as

$$M_{\boldsymbol{j}} \quad = \quad = \sum_{\boldsymbol{i}\ \mathrm{in}\ \mathcal{A}} \psi_{\boldsymbol{i}}\,C_{\boldsymbol{i}+\boldsymbol{j}} \quad , \tag{3}$$

where $\mathcal{A}$ is the support of the PSF and $\psi_{\boldsymbol{i}} = \psi(\boldsymbol{i})$ is an image of the PSF evaluated at integer pixel positions $\boldsymbol{i}$. This operation can be seen as "gathering up" the signal that is dispersed into many pixels by the PSF. In Appendix A.1 we derive the matched filter and

We define the *detection map* $D_{\boldsymbol{j}}$ as the matched filter, scaled to be in convenient units:

$$D_{\boldsymbol{j}} = \frac{1}{\|\boldsymbol{\psi}\|^2} \sum_{\boldsymbol{i}\ \mathrm{in}\ \mathcal{A}} \psi_{\boldsymbol{i}}\,C_{\boldsymbol{i}+\boldsymbol{j}} \quad , \tag{4}$$

[†]Which should probably be called the "pixel-correlated" PSF!

The PSF norm $\|\boldsymbol{\psi}\|$ is

$$\|\boldsymbol{\psi}\| = \sqrt{\sum_{\boldsymbol{i} \text{ in } \mathcal{A}} \psi_{\boldsymbol{i}}^2} \quad , \tag{5}$$

and as shown in Appendix A.1.2, a Gaussian PSF with standard devation w pixels has a norm approximately:

$$\left\|\boldsymbol{\psi}^{\boldsymbol{G}}\right\| \simeq \frac{1}{2\sqrt{\pi}w} \quad . \tag{6}$$

The per-pixel error in the detection map is given by

$$\sigma_D = \frac{\sigma_1}{\|\boldsymbol{\psi}\|} \quad . \tag{7}$$

We have scaled the detection map so that each pixel contains the maximum-likelihood estimate of the total counts of a source centered at that pixel. That is, if we compute at pixel $\boldsymbol{j}$ the total source counts $F_{\boldsymbol{j}}^*$ to minimizes the chi-squared (χ^2) residual within the support of the PSF:

$$F_{\boldsymbol{j}}^* = \arg\min_f \sum_{\boldsymbol{i}} \left(\frac{C_{\boldsymbol{i}+\boldsymbol{j}} - f\,\psi_{\boldsymbol{i}}}{\sigma_1} \right)^2 \tag{8}$$

we find

$$F_{\boldsymbol{j}}^* = \frac{\sum_{\boldsymbol{i}} C_{\boldsymbol{i}+\boldsymbol{j}}\,\psi_{\boldsymbol{i}}}{\sum_{\boldsymbol{i}} \psi_{\boldsymbol{i}}^2} \tag{9}$$

$$F_{\boldsymbol{j}}^* = D_{\boldsymbol{j}} \tag{10}$$

as defined above. That is, a significant peak in the detection map indicates the likely position of a point source, and the value of the detection map at a pixel is the maximum-likelihood estimate of the total source counts for a source centered at that pixel.

FIXME – the paragraph below is repetitive.

The intuition behind creating a detection map is that the point-spread function spreads the signal from a point source over a number of pixels. By correlating with the PSF, we "gather up" that spread-out signal, weighting the pixels over which the signal has been spread by the fraction of the signal that went into each pixel. By scaling the detection map by the inverse PSF norm, we convert it to units of total source counts; this will make it easier to combine detection maps in what follows.

3.1. Comments

Convolution. It is common to hear the statement that *convolving* an image by its PSF model yields an optimal detection filter. This is only true if the PSF is defined to be "flipped" in both axes relative to the image coordinate system. We prefer to define the PSF in unflipped coordinates, so correlation rather than convolution is the correct operation (they are equivalent when the PSF model is symmetric).

Peaks. We have not yet said anything about *how* to detect peaks at integer pixel or sub-pixel locations; we discuss this in **??**.

FIXME – should I talk about peaks and detection thresholds here?

Sufficient statistic. Correlation by the PSF summarizes all relevant information regarding the presence of a source at each pixel; the detection map and its variance are sufficient to describe our knowledge. In some astronomical source detection packages (including SourceExtractor), there is a notion of requiring more than one neighboring pixel to exceed a detection threshold. This is not necessary or useful; in effect it imposes a larger detection threshold that varies based on the source morphology and PSF, which is undesirable.

Optimality FIXME – write this!

PSF model FIXME – how sensitive is this to the PSF model? (This is related to the way the peak value dips as the source shifts across a pixel, and ties in with the practicalities of what one does with a detection map.)

What would Bayes do? FIXME – We haven't really explained that we are producing and evaluating an estimate for each pixel. Could also point out that one could instantiate and test the generative model (with position and flux) at each pixel, and that this basically just short-cuts that process, telling you places where that test is going to give significant results.

4. Detecting a point source in multiple images

In this section we will assume we have a stationary point source whose flux is constant over time, and a series of images taken through different bandpass filters and with different noise levels, exposure times, point-spread functions, and telescope pointings. We can achieve optimal detection of the source by building a detection map for each image and combining them with weights as described below.

4.1. Identical bandpass filters

We first present the simpler case where all the images are taken through identical bandpass filters.

As we have seen, the detection map defined in equation 4 is a maximum-likelihood estimate of the total *counts* contributed by the source, in the units of the original image. In order to combine

information from multiple images, we must calibrate them so that they are in the same units. Since this calibration is simply a linear scaling, it can be applied to the original image or to the detection map. Similarly, if the images are on different pixel grids—either from different pointings of the same CCD, or from different CCDs—then we must *resample* the detection maps to a common pixel grid. If the original image is well-sampled, then the detection map (which has been further smoothed by PSF correlation) will also be well-sampled, so resampling to a pixel grid of the same or finer resolution results in no loss of information. Since the pixel values in the detection map represent the *total* flux from the source, the detection map does not need to be rescaled when resampled to a different pixel scale.

Once the detection map for each image has been calibrated and resampled to a common pixel grid, we have multiple *independent* maximum-likelihood estimates of the source flux in our chosen filter, each with a known standard deviation and Gaussian statistics. That is, we have multiple Gaussian likelihood functions that we wish to combine. Since they are independent, the combined likelihood is the product of the individual likelihoods. For Gaussian distributions, the resulting aggregate maximum likelihood estimate is simply the inverse-variance-weighted sum of the individual estimates.

If the calibration factor κ_i scales image i to flux in common units, and R_i represents resampling to the common pixel grid, then the flux estimate E_i is

$$E_i \quad = \quad R_i(\kappa_i \, D_i) \tag{11}$$

with per-pixel error

$$\sigma_{E,i} \quad = \quad \frac{\kappa_i \, \sigma_{1,i}}{\|\boldsymbol{\psi}\|_i} \tag{12}$$

and we combine the estimates from multiple images via

$$E^\star \quad = \quad \frac{\sum_i E_i \, \sigma_{E,i}^{-2}}{\sum_i \sigma_{E_i}^{-2}} \tag{13}$$

which has per-pixel error

$$\sigma_{E,\star} \quad = \quad \left(\sum_i \sigma_{E,i}^{-2} \right)^{-\frac{1}{2}} \quad . \tag{14}$$

This is simply the maximum-likelihood estimate of the flux based on a set of independent Gaussian estimates.

In summary, the procedure to produce an optimal detection map given multiple images (in the same filter) is:

1. *correlate* each image by its PSF model
2. *calibrate* each resulting detection map (and its variance) to common units
3. *resample* each calibrated detection map to a common pixel grid
4. *coadd* the calibrated detection maps weighted by their inverse variances.

Assuming well-sampled images, the *correlation*, *calibration*, and *resampling* steps can occur in any order. Importantly, however, the *coaddition* stage must occur *after* correlation by the point-spread functions of the individual images; each image must be correlated by its own matched filter to produce *detection maps* which are coadded.

4.2. Comments

Optimality FIXME – write this!

Dumb things that people do FIXME – how exactly do coadds suck?

4.3. Different bandpass filters

As we saw in the single-bandpass case, we can combine multiple individual exposures into an aggregate estimate of the flux of a point source. In order to do this, it was essential to calibrate the images so that each one was an estimate of the same underlying quantity. The multiple-bandpass case is similar: For each bandpass, we first combine the images taken in that bandpass into an aggregate estimate. Then, to combine the bandpasses we must scale them so that they are estimates of the same quantity. This requires *knowing* the spectral energy distribution, or at least the *colors* in the filters of interest, of the source to be detected; this allows us to scale the various bandpasses so that they are estimates of a common quantity: perhaps the flux in a canonical band, or some other linear quantity such as the integrated intensity.

The intuition here is that if we know that our sources of interest are twice as bright in bandpass A as in bandpass B, then we can convert an estimate of the brightness in band B into an estimate of the brightness in band A by multiplying by two. The variance of the scaled estimate increases appropriately (by a factor of four), so a bandpass in which a source is expected to be faint will contribute an estimate with a large variance and will be downweighted when the estimates are combined. We can also view the problem as one of estimating a total flux that has been split into the different bandpasses, and in that view it is analogous to the way flux is spread into pixels by the point-spread function.

Assume we have detection maps D_j, with per-pixel standard deviation σ_j for a number of different bandpasses. Assume each bandpass has a known conversion factor f_j to the canonical band; that is,

$$D_j \sim \mathcal{N}\left(f_j F\,,\,\sigma_j^2\right) \tag{15}$$

for flux in the canonical band F. Given a number of such detection maps, we first scale them so they are all estimates of the same quantity, by dividing by f_j:

$$F_j = \frac{D_j}{f_j} \tag{16}$$

$$F_j \sim \mathcal{N}\left(F\,,\,\frac{\sigma_j^2}{f_j^2}\right) \tag{17}$$

and assuming that the images are independent, we combine them by inverse-variance weighting to produce the maximum-likelihood estimate for F:

$$\hat{F} = \frac{\sum_j \frac{D_j}{f_j}\frac{f_j^2}{\sigma_j^2}}{\sum_j \frac{f_j^2}{\sigma_j^2}} = \frac{\sum_j D_j\, f_j\, \sigma_j^{-2}}{\sum_j f_j^2\, \sigma_j^{-2}} \tag{18}$$

with per-pixel error

$$\hat{\sigma}_F = \left(\sum_j f_j^2 \sigma_j^{-2}\right)^{-\frac{1}{2}} \quad . \tag{19}$$

For example, if we treat r band as the canonical band and our objects of interest have color $r - i = 1$, then we expect the flux in i to be a factor of 2.5 greater than the flux in r; $f_i = 2.5$, and we will scale our i-band detection map D_i by $1/f_i = 0.4$. Since the sources are expected to be brighter in i band so we must *scale down* the i-band estimate to produce an r-band estimate. The i-band variance is also reduced in a corresponding way, so this does not dilute the weight of high-precision measurements.

FIXME – we need to talk here about this idea of "priors", and perhaps at this point say that we recommend running with a number of SEDs and taking the union of detections.

4.4. Comments

Chi-squared coadd And how it is different than this.

You always have a prior And if you don't think you have a prior, you're probably using a dumb prior.

How much does it hurt if the SED is wrong? (It's just like having a wrong PSF model.)

5. Experiments

5.1. Sloan Digital Sky Survey

Here we present some proof-of-concept experiments on Sloan Digital Sky Survey (SDSS) data from the Stripe 82 region **?**.

Input data selection. We select input exposures to include as follows. Using the "window_flist" table for SDSS Data Release 9 **?**, we select photometric exposures from camera column 3, with RA in the range $[45.15, 45.85]$ and Dec in the range $[-0.2, -0.1]$. This results in 109 fields from 23 unique runs, which are listed in Table 1.

Detection map and resampling. For each input image, we take the PSF to be the double-Gaussian model computed by the SDSS *Photo* pipeline **?**, and compute the detection map. Since SDSS DR9 images are already in sky-subtracted, calibrated "nano-maggies", we do not need to apply a calibration factor to convert them to common units, nor do we need to remove a sky background level. We set to zero any image pixels that are masked as saturated, interpolated, cosmic ray, ghost, or "not checked" (close to the image boundaries). We compute the per-pixel noise (σ_1) based on a Gaussian approximation of the Poisson errors from the mean sky level, plus dark current and read noise, as recommended by the SDSS team.[‡]

For each of the u, g, r, i and z bandpass filters ("bands"), we resample to a common pixel grid and then compute a per-band detection map as per equation 13. The common pixel grid is defined by a WCS tangent-plane projection centered on our patch of interest with the nominal SDSS pixel scale (0.396 arcseconds per pixel) and image size 4096×2048 **?**. Since the SDSS pipeline adds a 128-pixel overlap between neighboring fields in a run, we first build a detection map for each run, averaging values in the overlapping regions to avoid double-counting, then combine these per-run detection maps as per equation 13. For resampling we use a third-order Lanczos kernel.

Background correction. We find that there is considerable spatially-coherent structure in our combined detection maps, which we suspect is due to slightly biased background subtraction in the

[‡] *http://data.sdss3.org/datamodel/files/BOSS_PHOTOOBJ/frames/RERUN/RUN/CAMCOL/frame.html*

SDSS pipeline code. We estimate and remove a residual background as follows. After masking all pixels above 5 sigma, we compute the median value within each cell of a $\sim 100 \times 100$-pixel grid, and interpolate between these values with a third-order B-spline. The resulting background-subtracted detection maps have no apparent background structure (by eye) and their pixel distributions are closer to Gaussian. See Figure 1.

Detection In this paper we are not concerning ourselves with blended objects, so we identify peaks in the detection map naively. We find pixels above the desired threshold, expand by a 2-pixel radius, and find the connected components. For each component, we return the peak pixel, ignoring other peaks within the component.

We construct a series of spectral energy distributions (SEDs), building the combined detection map and running the peak detection algorithm for each. A standard approach for detecting objects in imaging from multiple bands is to weight each band by its signal-to-noise. This is equivalent to assuming a flat spectrum, or zero color. Our SED-matched filters weight each band according to both the amount of flux we expect to see in the band, and its signal-to-noise. The SEDs we use here include "Flat" (expected color $g = r = i$), "Red" ($g - r = r - i = 1$), "Blue" ($g - r = r - i = -1$), and four points lying approximately along the stellar locus ($(g - r, r - i)$ colors $(0.5, 0.2)$, $(1.0, 0.4)$, $(1.4, 0.7)$, and $(1.4, 1.4)$. In the experiments here we will compare the detection strengths in the different SEDs. In practice we would return the union of peaks detected in this way.

Results In Figure 2 we show that objects with red colors in the SDSS catalog are detected more strongly by our SED-matched Red filter than a Flat SED (which corresponds to the traditional multi-band detection approach). There are very few objects that the Blue SED-matched filter detects more strongly than the Flat or Red filters; astrophysical objects with blue colors (bluer than 0 in the AB system) are rare.

In Figure 3 we show, in color-color space, the sources detected by each of our SED-matched filters. As expected, each filter is most sensitive to sources in a given region of color-color space.

In Figure 4 we show postage-stamp images of sources that are detected by our SED-matched filters and not detected by a traditional approach. In Figure 5 we show in color-color space the additional objects that are detected as we apply a sequence of SED-matched filters.

5.2. Simulations

—Single exposure, single band

—Multiple exposures (different PSFs, sky), single band

—Multiple exposures, multiple bands

Run	Min field	Max field
1752	163	166
1887	107	111
2578	112	116
2589	200	204
2700	168	171
2738	220	224
2820	168	172
2861	86	90
2873	211	215
3362	169	173
3384	671	675
3461	20	24
4128	425	429
4157	178	181
4198	664	668
4207	673	677
4849	757	760
4858	656	660
4874	723	727
4905	304	307
4927	654	658
4933	665	668
4948	245	249

Table 1: SDSS fields included in the experiments. All fields are from camera column ("camcol") 3. The minimum and maximum (inclusive) fields are listed.

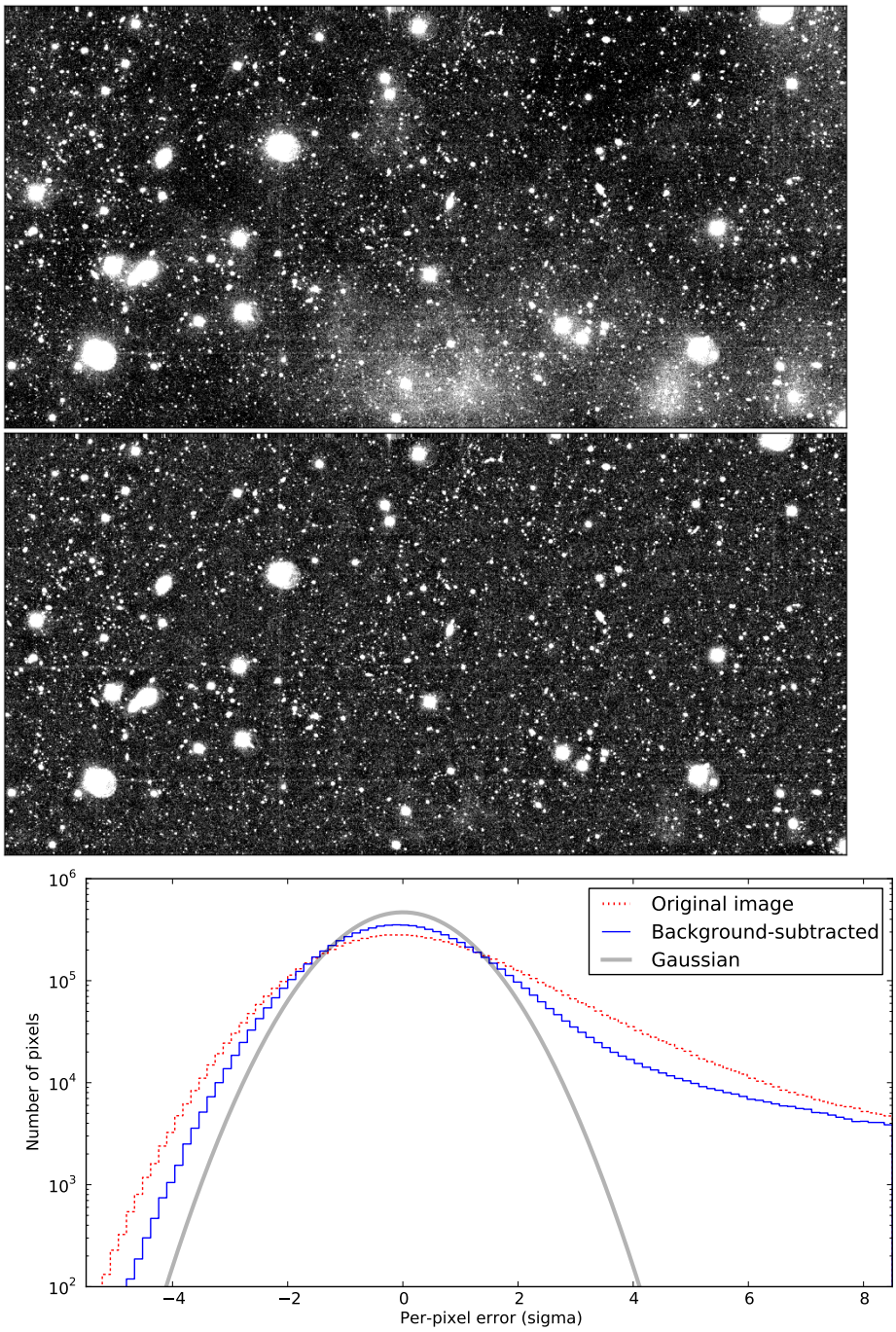

Fig. 1.— Residual background subtraction. **Top:** combined detection map for the SDSS r band, combining images from 23 runs. There are clear patterns in the background level. **Middle:** after estimating and removing the residual background as described in the text, the background appears more flat. **Bottom:** after removing the background, the pixel error statistics are closer to Gaussian, though the distribution is still significantly broader than expected.

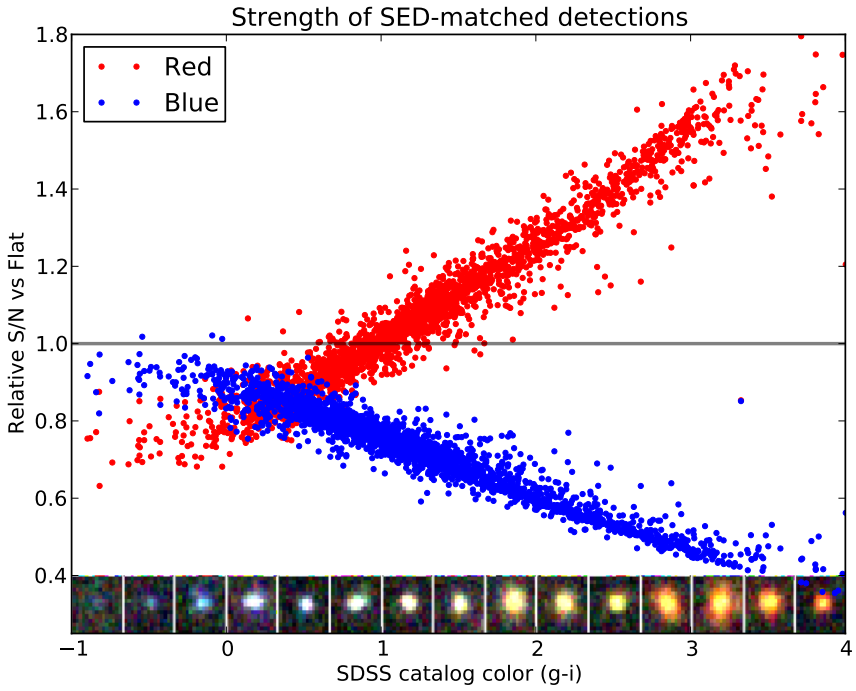

Fig. 2.— Detection strengths in the Red, Flat, and Blue SED-matched filters, versus SDSS-measured colors. We performed a spatial match between our detected sources and SDSS sources (from the DR7 CasJobs database "Stripe82"; **?**) within 1 pixel. Redder objects in $g - i$ are detected more strongly in the Red SED-matched filter than Flat or Blue. The handful of objects with significantly blue colors are detected more strongly with the Blue filter.

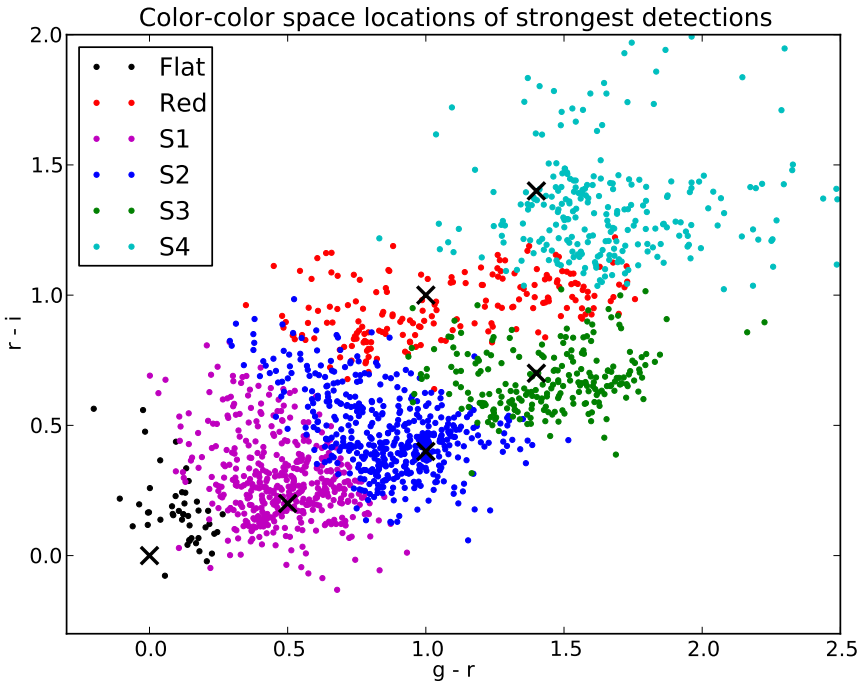

Fig. 3.— Color-color space locations of sources detected most strongly by each of the filters. The color measurements are from the SDSS Stripe82 coadd catalog **?**, matched to our detections using a 1-pixel matching radius. It is clear that our different SED-matched filters are tuned to detect objects in particular regions of color-color space. The "X" marks indicate the color each SED-matched filter is tuned for. Note that many of the sources are detected above a given detection threshold by several of the SED-matched filters; here each source is assigned to the filter by which it is detected most strongly.

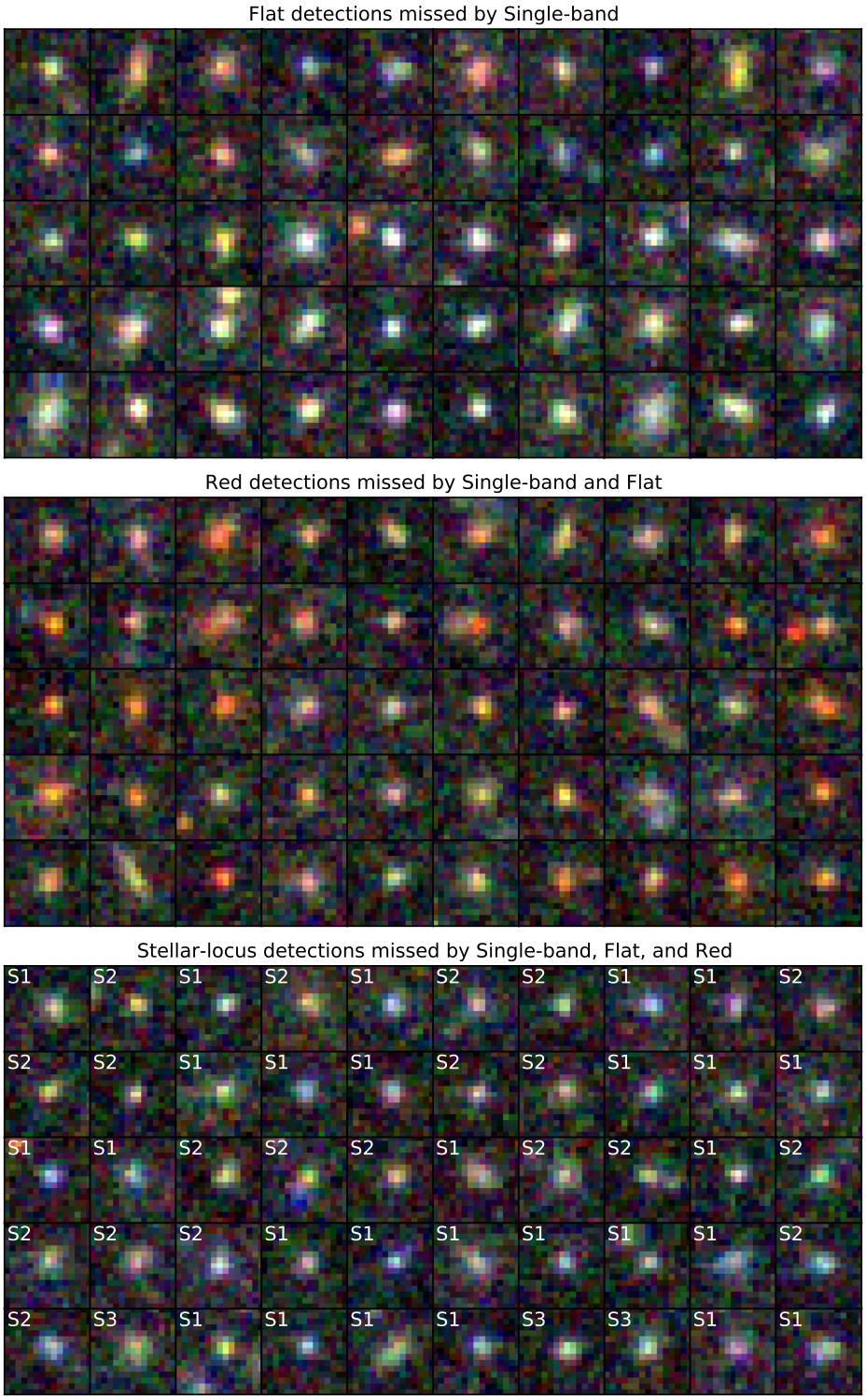

Fig. 4.— Sources detected by our SED-matched filters and not by a "traditional" approach; full caption on next page

Fig. 4.— [figure previous page] Sources detected by applying a sequence of SED-matched filters. We start with sources detected by a "traditional" approach that takes the union of single-band detections in each of the g, r, and i filters. We then apply, in sequence, our Flat, Red, and Stellar locus SED-matched filters. Each panel shows the sources detected by the new filter and not detected by previous filters, in order of decreasing signal-to-noise. In order to make the objects easier to see, we have increased the detection threshold to 20σ.

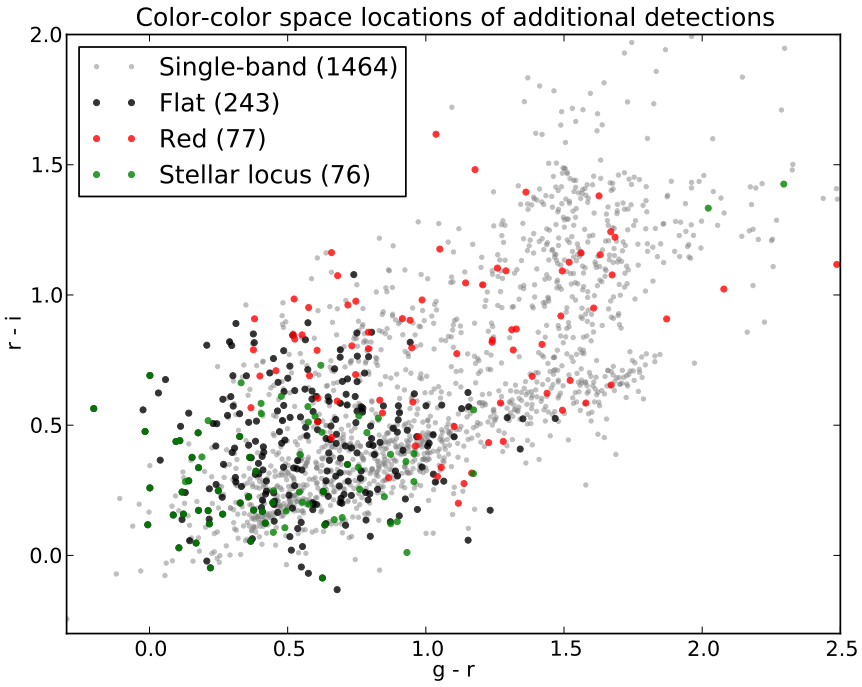

Fig. 5.— Color-color space locations of additional sources detected as additional SED-matched filters are applied. We start by taking all single-band detections in the g, r, and i-band filters. Next, we add sources detected by the Flat SED-matched filter, then the Red filter, then a set of filters along the stellar locus. Each subsequent SED-matched filter detects sources that are missed (at a fixed detection threshold) by the previous filters.

6. Discussion

–similarities and differences from Szalay et al 19999 – SED-matched filters $\sim$ "optimal subspace filtering"

–galaxy detection?

–No weights go to zero – all images contribute *something*, in proportion to t he information they bring. Contrast with lucky imaging and some co-adding schemes.

–sub-pixel shifts: lower threshold so that max pixel is above threshold; post-process to cut out peaks that really were below threshold.

7. Conclusions

[[What to do with a detection image; how to proceed. Build models, fit simultaneously to all images in the "stack".]]

[[Practicalities: bad pixels / masks, interpolation, etc]]

Ben Weiner (University of Arizona), Julianne Dalcanton (University of Washington), Brad Holden (UCO/Lick Observatories), Micha Gorelick (Fast Forward Labs), Robert Lupton, Steve Bickerton, Paul Price, and Craig Loomis (Princeton)

A. Appendices

A.1. Optimal linear detection

Let us assume that there exists a linear filter whose output allows optimal detection of isolated point sources. That is, we seek a (two-dimensional) set of coefficients $a_{\boldsymbol{i}}$ that, when correlated with the image $C_{\boldsymbol{j}}$, produces a map $M_{\boldsymbol{j}}$ whose peak is the likely location of the point source. See Figure **??**.

The linear filtering (correlation) operation is

$$M_{\boldsymbol{j}} = \sum_{\boldsymbol{i} \text{ in } \mathcal{A}} a_{\boldsymbol{i}} \, C_{\boldsymbol{i}+\boldsymbol{j}} \tag{A1}$$

where $\mathcal{A}$ is the support of $\boldsymbol{a}$ (integer pixel positions), and the center of $\boldsymbol{a}$ is $(0, 0)$. We will demand that the elements of $\boldsymbol{a}$ are non-negative and sum to unity.

Inserting equation **??**, we get

$$M_{\boldsymbol{j}} \sim \sum_{\boldsymbol{i} \text{ in } \mathcal{A}} C_k \, a_{\boldsymbol{i}} \, \psi_{\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}} + \mathcal{N}\big(0\,,\,\sigma_1^2\big) \tag{A2}$$

$$\sim \mathcal{N}\left(C_k \sum_{\boldsymbol{i} \text{ in } \mathcal{A}} a_{\boldsymbol{i}} \, \psi_{\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}} \,,\, \sum_{\boldsymbol{i} \text{ in } \mathcal{A}} a_{\boldsymbol{i}}^2 \, \sigma_1^2\right) \tag{A3}$$

and the per-pixel signal-to-noise in the map is

$$\mathbb{SN}(M_{\boldsymbol{j}}) = \frac{C_k \, \sum a_{\boldsymbol{i}} \, \psi_{\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}}}{\sigma_1 \sqrt{\sum_{\boldsymbol{i}} a_{\boldsymbol{i}}^2}} \quad . \tag{A4}$$

We want to choose coefficients $a_{\boldsymbol{i}}$ to maximize the signal-to-noise at the true pixel position of the source, $\boldsymbol{k}$. Rewriting the expression using dot-products and norms[§], treating the two-dimensional images $a_{\boldsymbol{i}}$ and $\psi_{\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}}$ as vectors indexed by $\boldsymbol{i}$, we have:

$$\mathbb{SN}(M_{\boldsymbol{j}}) = \frac{C_k \, \boldsymbol{a} \cdot \boldsymbol{\psi_{j-k}}}{\sigma_1 \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}}} \tag{A5}$$

$$= \frac{C_k \, \|\boldsymbol{a}\| \, \|\boldsymbol{\psi_{j-k}}\| \cos\theta}{\sigma_1 \, \|\boldsymbol{a}\|} \tag{A6}$$

$$= \frac{C_k \, \|\boldsymbol{\psi_{j-k}}\| \cos\theta}{\sigma_1} \tag{A7}$$

where θ is a generalized angle between $\boldsymbol{a}$ and $\boldsymbol{\psi_{j-k}}$. At the pixel position of the source, $\boldsymbol{k}$,

$$\mathbb{SN}(M_{\boldsymbol{k}}) = \frac{C_k \, \|\boldsymbol{\psi_0}\| \cos\theta}{\sigma_1} \tag{A8}$$

Clearly this is maximized when $\theta = 0$, *i.e.*, when $\boldsymbol{a}$ is a multiple of $\boldsymbol{\psi_0}$, the PSF evaluated at a grid of integer pixel positions. Since we have defined both the PSF and coefficients a to sum to unity, we find that the optimal linear filter for detection is given by:

$$\boldsymbol{a} = \boldsymbol{\psi_0} \quad , \tag{A9}$$

which means that the operation of *correlating* the image with its PSF produces a map with optimal signal-to-noise. Repeating equation A1, we have found that the map $M_{\boldsymbol{j}}$ can be computed by correlating the image with its PSF:

$$M_{\boldsymbol{j}} = \sum_{\boldsymbol{i} \text{ in } \mathcal{A}} \psi_{\boldsymbol{i}} \, C_{\boldsymbol{i}+\boldsymbol{j}} \quad , \tag{A10}$$

where, as before, $\mathcal{A}$ is the support of the PSF.

The signal-to-noise in this map at the true source pixel position $\boldsymbol{k}$ is

$$\mathbb{SN}(M_{\boldsymbol{k}}) = \frac{C_k \, \|\boldsymbol{\psi}\|}{\sigma_1} \quad . \tag{A11}$$

[§]All norms in this paper are ℓ_2 norms: $\|X\| = \sqrt{\sum_i X_i^2}$.

A.1.1. Optimality

We compute the variance of the detection map estimator (equation 4), and show that is equal to the Cramér–Rao bound. Substituting our image model into the detection map,

$$D_{\boldsymbol{j}} = \frac{1}{\|\boldsymbol{\psi}\|^2} \sum_i \psi_{\boldsymbol{i}}\, C_{\boldsymbol{i}+\boldsymbol{j}} \tag{A12}$$

$$\sim \frac{1}{\|\boldsymbol{\psi}\|^2} \sum_i \psi_{\boldsymbol{i}}\, \mathcal{N}\big(F\,\psi_{\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}}\,,\,\sigma_1^2\big) \tag{A13}$$

and at the true source position, $j = k$;

$$D_{\boldsymbol{j}} \sim \frac{1}{\|\boldsymbol{\psi}\|^2}\mathcal{N}\left(F\,\sum_i \psi_{\boldsymbol{i}}^2\,,\,\sum_i \psi_{\boldsymbol{i}}^2\sigma_1^2\right) \tag{A14}$$

$$D_{\boldsymbol{j}} \sim \mathcal{N}\left(F\,,\,\frac{\sigma_1^2}{\|\boldsymbol{\psi}\|^2}\right) \tag{A15}$$

A.1.2. Norm of a Gaussian PSF

For a Gaussian PSF with standard deviation w pixels,

$$\psi^G(x,y) = \frac{1}{2\pi w^2}\exp\left(-\frac{x^2}{2w^2}\right)\exp\left(-\frac{y^2}{2w^2}\right) \tag{A16}$$

the norm is

$$\|\boldsymbol{\psi}^{\boldsymbol{G}}\| = \sqrt{\sum_x\sum_y\left(\frac{1}{2\pi w^2}\exp\left(-\frac{x^2}{2w^2}\right)\exp\left(-\frac{y^2}{2w^2}\right)\right)^2} \tag{A17}$$

$$\|\boldsymbol{\psi}^{\boldsymbol{G}}\| \simeq \sqrt{\iint \frac{1}{4\pi^2 w^4}\exp\left(-\frac{x^2}{w^2}\right)\exp\left(-\frac{y^2}{w^2}\right)\mathrm{d}x\,\mathrm{d}y} \tag{A18}$$

$$\|\boldsymbol{\psi}^{\boldsymbol{G}}\| \simeq \frac{1}{2\sqrt{\pi}w} \tag{A19}$$

so the detection map has signal-to-noise at the true source position $\boldsymbol{k}$,

$$\mathbb{SN}(D^G_{\boldsymbol{k}}) = \frac{C_k}{2\sqrt{\pi}w\sigma_1} \quad . \tag{A20}$$

Note, however, that we have defined the point-spread function $\psi_(\cdot)$ to be the *pixel-convolved* response, so it cannot be exactly Gaussian if the pixel response is assumed to be a boxcar function. In practice, however, a two-dimensional Gaussian with variance v^2 correlated with a two-dimensional boxcar function is well approximated by a Gaussian with variance $v^2 + \frac{1}{12}$, as long as $v \gtrsim \frac{1}{2}$.

A.1.3. Why not signal-to-noise-squared?

In correlating the image with the PSF, it looks like the detection map weights pixels by their signal-to-noise, rather than signal-to-noise *squared*. This apparent conflict can be resolved by scaling the pixel values so that each pixel is an estimate of the same quantity. That is, we want to estimate the total source counts C, but the pixels contain estimates of the source counts scaled by the PSF, $C\psi$; we must undo this scaling by multiplying the pixels by $1/\psi$.

Given a source at position $\boldsymbol{k}$, we define the image K whose pixels each contain an estimate of the total source counts:

$$K_{\boldsymbol{j}} = \frac{1}{\psi_{\boldsymbol{j}-\boldsymbol{k}}} S_{\boldsymbol{j}} \tag{A21}$$

$$K_{\boldsymbol{j}} \sim \frac{1}{\psi_{\boldsymbol{j}-\boldsymbol{k}}} \mathcal{N}\left(C_k\, \psi_{\boldsymbol{j}-\boldsymbol{k}}\,,\, \sigma_1^2\right) \tag{A22}$$

$$K_{\boldsymbol{j}} \sim \mathcal{N}\left(C_k\,,\, \frac{\sigma_1^2}{\psi_{\boldsymbol{j}-\boldsymbol{k}}^2}\right) \quad . \tag{A23}$$

The signal-to-noise remains the same, since we have just scaled the values:

$$\mathbb{SN}(K_{\boldsymbol{j}}) = \frac{C_k\, \psi_{\boldsymbol{j}-\boldsymbol{k}}}{\sigma_1} \tag{A24}$$

$$\mathbb{SN}(K_{\boldsymbol{j}}) = \mathbb{SN}(S_{\boldsymbol{j}}) \quad . \tag{A25}$$

As before, the detection map pixels are a linear combination of the (shifted) pixels of the K image with weights $b_{\boldsymbol{i}}$:

$$D^\star_{\boldsymbol{j}} = \sum_{\boldsymbol{i}} b_{\boldsymbol{i}}\, K_{\boldsymbol{i}+\boldsymbol{j}} \tag{A26}$$

$$\sim \sum_{\boldsymbol{i}} b_{\boldsymbol{i}}\, \mathcal{N}\left(C_k\,,\, \frac{\sigma_1^2}{\psi_{\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}}^2}\right) \tag{A27}$$

$$\sim \mathcal{N}\left(C_k \sum_{\boldsymbol{i}} b_{\boldsymbol{i}}\,,\, \sum_{\boldsymbol{i}} \frac{b_{\boldsymbol{i}}^2 \sigma_1^2}{\psi_{\boldsymbol{i}+\boldsymbol{j}-\boldsymbol{k}}^2}\right) \tag{A28}$$

and the signal-to-noise in that detection map at pixel $\boldsymbol{k}$ is

$$\mathbb{SN}(D^\star_{\boldsymbol{k}}) = \frac{C_k \sum_{\boldsymbol{i}} b_{\boldsymbol{i}}}{\sigma_1 \sqrt{\sum_{\boldsymbol{i}} \frac{b_{\boldsymbol{i}}^2}{\psi_{\boldsymbol{i}}^2}}} \tag{A29}$$

which is maximized by setting the $b_{\boldsymbol{i}}$

$$b_{\boldsymbol{i}} \propto \psi_{\boldsymbol{i}}^2 \quad : \tag{A30}$$

proportional to the signal-to-noise *squared*, as expected.