A METHOD FOR INDIVIDUAL SOURCE BRIGHTNESS ESTIMATION IN SINGLE- AND MULTI-BAND DATA.

T. M. Crawford,1,2 E. R. Switzer,1 W. L. Holzapfel,3 C. L. Reichardt,3 D. P. Marrone,1,2,4 and J. D. Vieira1,5,6

ABSTRACT

We present a method of reliably extracting the flux of individual sources from sky maps in the presence of noise and a source population in which number counts are a steeply falling function of flux. The method is an extension of a standard Bayesian procedure in the millimeter/submillimeter literature. As in the standard method, the prior applied to source flux measurements is derived from an estimate of the source counts as a function of flux, $dN/dS$. The key feature of the new method is that it enables reliable extraction of properties of individual sources, which previous methods in the literature do not. We first present the method for extracting individual source fluxes from data in a single observing band, then we extend the method to multiple bands, including prior information about the spectral behavior of the source population(s). The multi-band estimation technique is particularly relevant for classifying individual sources into populations according to their spectral behavior. We find that proper treatment of the correlated prior information between observing bands is key to avoiding significant biases in estimations of multi-band fluxes and spectral behavior, biases which lead to significant numbers of misclassified sources. We test the single- and multi-band versions of the method using simulated observations with observing parameters similar to that of the South Pole Telescope data used in Vieira et al. (2009).

Subject headings: submillimeter — radio continuum: galaxies — methods: data analysis — methods: statistical

1. INTRODUCTION

The bias incurred when using noisy measurements to estimate source counts as a function of flux has been a recognized issue since at least the time of Eddington (1913). The seminal work of Scheuer (1957) on the topic was crucial to reconciling apparently conflicting radio source count measurements and establishing the model of an evolving universe (e.g., Longair 1966).

Accounting for this bias becomes particularly important when the source population under investigation has counts that are a steeply falling function of flux, as is the case for sources in the 1-to-100 mJy range at millimeter/submillimeter (mm/submm) wavelengths. Sources in this flux and wavelength range have been of particular recent interest to astronomers, astrophysicists, and cosmologists, primarily due to the recently discovered population of high-redshift starburst galaxies (see Blain et al. (2002) for a review), which has been shown to make up a substantial fraction of the cosmic infrared background (Lagache et al. 2003) and which provides strong tests for theories of galaxy and star formation. Both Bayesian and frequentist methods have been developed for dealing with this bias in mm/submm surveys (e.g., Coppin et al. 2005, Maloney et al. 2003). However, as will be discussed in Sec. 2, the methods developed so far are not appropriate for estimating properties of individual sources.

Dealing with this bias is further complicated when source properties are estimated from data in multiple wavelength bands. Mason et al. (2009), for example, used a maximum-likelihood technique to make an unbiased estimate of the spectral index distribution of a population of sources from two-band centimeter-wave data, but they did not attempt to estimate spectral properties of individual sources. When multi-band data have been used to estimate properties of individual sources, the data in different bands have generally been treated as independent, despite the fact that the information used to correct the biased flux measurements is highly correlated between bands (e.g., Greve et al. 2008). As will be shown in Sec. 3, ignoring these correlations can result in severely misestimated source properties.

In this work, we propose a reliable, minimally biased estimator for the single-band brightness of individual sources. We then extend this formalism to the simultaneous estimation of individual source properties over many bands, parameterized as either the source brightness in each band or as the brightness in one band and the spectral behavior between bands. The multi-band implementation of the method explicitly accounts for correlations in the information used to correct for flux bias in the individual bands. This work was motivated by the extragalactic sources detected in multi-band South Pole Telescope (SPT) data, which are described in Vieira et al. (2009, hereafter V09); however, we believe the method is directly applicable to other current or planned mm/submm experiments. We also believe the method should be appropriate for use in single- and multi-band surveys at other wavelengths, with the caveat that source variability could compromise the multi-band
2. SINGLE-BAND FLUX ESTIMATION

The estimation of source brightness in mm/submm surveys is complicated by the fact that, at these wavelengths, the number density of sources as a function of source flux is expected to be very steep. As a result, the measured flux of a source detected in a mm or submm survey will almost certainly suffer a positive bias, often referred to as “flux boosting.” In this work, we define flux boosting as the increased probability that a source we measure to have flux $S$ is really a dimmer source plus a positive noise fluctuation over the probability that it is a brighter source plus a negative noise fluctuation. This asymmetric probability distribution means that naive estimates of source fluxes will be biased high.\footnote{This phenomenon is closely related to what is referred to in the literature as “Eddington bias” (e.g., Teerikorpi 2004); however, the consensus use of that term in the literature is to describe the bias introduced to estimation of source counts vs. brightness, not on the estimated brightness of individual sources. This usage is consistent with the original work of Eddington (1913), and the distinction we draw was also pointed out in (among others) Hogg & Turner (1998) and (Copin et al. 2000).} In the literature of mm/submm source surveys, one standard method for dealing with this problem (e.g., [Coppin et al. 2003]) is to construct a posterior probability distribution for the intrinsic flux of each detection, as suggested by (among others) Jauncey (1968) and Hogg & Turner (1998). According to standard Bayesian reasoning, this posterior distribution is given by

$$P(S_i|S_m) \propto P(S_m|S_i)P(S_i),$$  \hspace{1cm} (1)

where $S_i$ is the intrinsic, true flux of the detected object, and $S_m$ is the measured flux. $P(S_m|S_i)$ is the likelihood of measuring a flux $S_m$ given a true flux $S_i$, which in the simplest case is a Gaussian distribution centered on $S_i$ with width $\sigma_n$, determined solely by the noise in the maps from which sources are being extracted. $P(S_i)$ is the prior probability of a source with intrinsic flux $S_i$, which is proportional to the differential number counts vs. flux, $dN/dS$.\footnote{In the standard choice of prior in the literature, which is the pixel distribution of pixel fluxes, which is also dependent on the survey instrument. When calculating source counts using this method, it is possible to use simulations to account for this instrumental dependence or demonstrate that it has a negligible effect (e.g., Austermann et al. 2009); however, the detailed shape of the posterior flux distributions for individual sources can still be affected, particularly at low significance.}

However, the formulation in Eqn. (1) implicitly assumes that each detection corresponds to (at most) one real source. This is equivalent to assuming that the instrument used in the survey has infinite resolution. In reality, there will always be some possibility that a resolution element containing a source above the detection threshold will also contain one or more fainter sources which will contribute to the measured flux in that resolution element. As a consequence, even in a noiseless measurement, the probability of measuring a particular value of flux within a finite resolution element is not identical (in general) to the probability that a single source of that flux exists within the resolution element. In much of the mm/submm literature (e.g., Coppin et al. 2003 and Austermann et al. 2009), $P(S_p)$ (the probability of the total astrophysical flux $S_p$ within a pixel) and $P(S)$ (the probability of finding a source of flux $S$ within the solid angle of a pixel) are used interchangeably. In other words, what is being reported when this method is used is not the distribution of intrinsic source fluxes (which is a property solely of the source population) but rather a distribution of pixel fluxes, which is also dependent on the survey instrument. When calculating source counts using this method, it is possible to use simulations to account for this instrumental dependence or demonstrate that it has a negligible effect (e.g., Austermann et al. 2009); however, the detailed shape of the posterior flux distributions for individual sources can still be affected, particularly at low significance. One approach which avoids these particular complications is to find the underlying (intrinsic) number counts model that is consistent with the observed pixel flux distribution (as in, e.g., Maloney et al. 2005 and Patanchon et al. 2009) or with the observed counts as a function of raw flux (as in “Reduction C” in Coppin et al. 2009). However, these methods are incapable of estimating properties of individual sources. For this purpose, we extend the traditional Bayesian method to describe the properties of only one source in a given resolution element. Instead of defining the posterior with respect to the (intrinsic) total flux in a resolution element or pixel, we define the posterior with respect to the intrinsic flux of the single brightest source in the resolution element. The posterior probability distribution for this quantity is

$$P(S_{\text{max}}|S_{p,m}) = P(S_{p,m}|S_{\text{max}})P(S_{\text{max}}),$$  \hspace{1cm} (2)

where $P(S_{\text{max}}|S_{p,m})$ is the posterior probability that the true flux of the brightest source in a pixel is $S_{\text{max}}$ given that we have measured the total flux in that pixel to be $S_{p,m}$; $P(S_{p,m}|S_{\text{max}})$ is the likelihood of measuring flux $S_{p,m}$ in a pixel given that the brightest source in that pixel has flux $S_{\text{max}}$; and $P(S_{\text{max}})$ is the prior probability that the brightest source in that pixel has flux $S_{\text{max}}$.

2.1. Prior Probability $P(S_{\text{max}})$

The prior probability $P(S_{\text{max}})$ can be expressed as the probability that one source of flux $S_{\text{max}}$ exists in that pixel times the probability that zero sources brighter than $S_{\text{max}}$ exist in that pixel. As mentioned previously, the probability that within a pixel there exists a source of flux $S$ is proportional to the differential number counts (per unit flux per unit solid angle) $dN/dS$. Under the assumption of purely Poisson statistics, the probability that zero sources above $S_{\text{max}}$ exist in a pixel is

$$P(N > S_{\text{max}} = 0) = \frac{\mu^n}{n!}e^{-\mu}; \quad n = 0$$

$$= e^{-\mu},$$

where $\mu$ is the mean number of sources above $S_{\text{max}}$ in pixels of size $\Delta\Omega_p$:

$$\mu(S_{\text{max}}) = \Delta\Omega_p \int_{S_{\text{max}}}^{\infty} \frac{dN}{dS}dS.$$  \hspace{1cm} (4)

In other words, $P(S_{\text{max}})$ should look like $dN/dS$ with an exponential suppression at low $S$ (where $\mu(S)$ will be large):

$$P(S_{\text{max}}) \propto \left. \frac{dN}{dS} \right|_{S=S_{\text{max}}} \exp \left( -\Delta\Omega_p \int_{S_{\text{max}}}^{\infty} \frac{dN}{dS'}dS' \right).$$  \hspace{1cm} (5)

We note that this choice of prior is natural in the context of characterizing individual source properties and avoids certain complications associated with the standard choice of prior in the literature, which is the pixel
probability distribution that would be obtained from analyzing noiseless observations of a sky with the assumed underlying source distribution (e.g., Coppin et al. 2003; Austermann et al. 2009). In particular, because point-source analyses of mm/sub-mm data almost inevitably involve spatially high-pass-filtering the data to remove large-scale noise and astronomical signals, the standard prior can take on negative flux values. This calls into question how such a prior can be describing the probability of the intrinsic flux of an astrophysical object.

2.2. Likelihood \( P(S_{p,m} | S_{\text{max}}) \)

The total flux in a pixel given that the brightest source in that pixel has flux \( S_{\text{max}} \) will be a sum of three contributions: the source at \( S_{\text{max}} \), a contribution from instrumental or atmospheric noise in the survey, and a contribution from sources fainter than \( S_{\text{max}} \).\(^8\) We can think of these each having its own probability distribution, and the probability of the sum of their contributions to the flux in a pixel will be the convolution of the individual distributions (by the addition theorem for probabilities \( P(u = x + y) = \int P_y(x) P_x(u - x) \, dx \)). We will assume that the contribution from instrument noise and atmosphere is well approximated by a Gaussian distribution:

\[
P(S_{p,m}, \text{noise-only}) = \frac{1}{\sqrt{2\pi \sigma^2_n}} e^{-S_{p,m}^2 / 2\sigma_n^2},
\]

where \( \sigma_n \) is the width of the combined instrument and atmospheric noise distribution. The contribution from sources fainter than \( S_{\text{max}} \) is given by the probability of total flux within a pixel, knowing that there are no sources in that pixel brighter than \( S_{\text{max}} \). The probability for flux in a pixel given \( dN/dS \) — the so-called “deflexion probability” or \( P(D) \) — is worked out in Scheuer (1957), and, under the assumption of pure Poisson statistics is

\[
P(S_{p,m}, \text{noise-free}) = FT\{e^{[r(\omega) - r(0)]}\}, \quad (7)
\]

where \( FT\{\} \) denotes Fourier Transform, and \( r(\omega) \) is the characteristic function of the probability of finding a source of flux \( S \) in a pixel of solid angle \( \Delta \Omega_p \):

\[
r(\omega) = FT\left\{ \frac{dN}{dS} \Delta \Omega_p \right\}. \quad (8)
\]

To calculate the contribution of sources fainter than \( S_{\text{max}} \) under the constraint that there are no sources in a pixel greater than \( S_{\text{max}} \), we apply the above result using a \( dN/dS \) distribution that is truncated at \( S_{\text{max}} \).

To summarize, the flux in a pixel given that we have exactly one source of flux \( S_{\text{max}} \) and zero sources above that flux will be sum of the contribution from the source at \( S_{\text{max}} \), the contribution from sources fainter than \( S_{\text{max}} \), and the contribution from noise. The probability distribution for this sum is the convolution of the individual distributions, so that

\[
P(S_{p,m}, S_{\text{max}}) = \delta(S_{\text{max}}) \ast \text{FT}\{e^{[r(\omega) - r(0)]}\} \ast \frac{1}{\sqrt{2\pi \sigma_n^2}} e^{-S_{p,m}^2 / 2\sigma_n^2}, \quad \text{where “\ast” denotes convolution.}
\]

\(^8\) In reality, there will also be contributions from diffuse signals such as galactic dust or primary CMB anisotropy. However, these diffuse contributions will likely have been filtered out of maps used for source detection, so we neglect them here.

2.3. Beam and filtering effects

The exact formulations in Eqs. 5 and 9 only hold for a survey in which the only spatial filtering done involves binning into pixels. For most real instruments, the situation is complicated by the instrument beam (or point-spread function) and by the time-domain and map-domain filtering performed on the data. The likelihood in Eqn. 9 needs to be modified to reflect the difference in how sources fainter than \( S_{\text{max}} \) contribute to a real instrumental beam compared to the top-hat pixel considered earlier. Scheuer (1957) shows that Eqn. 9 can be modified for finite-resolution experiments by defining:

\[
r(\omega) = FT\left\{ \int \frac{dN}{dS_{\text{beam}}} \frac{d\Omega}{B(\theta, \phi)} \right\}, \quad (10)
\]

where \( B(\theta, \phi) \) is the angular response pattern of the instrument, and

\[
S_{\text{beam}} = \frac{S}{B(\theta, \phi)}. \quad (11)
\]

This formulation can be extended to account for any filtering in the data analysis by modifying the angular response function to include the filtering.

The prior in Eqn. 5 and the \( \delta(S_{\text{max}}) \) term in Eqn. 9 must also be modified for a finite-resolution experiment. We choose to define \( S_{\text{max}} \) for a finite-resolution experiment as the source that contributed the most flux in a resolution element. That means that both equations must be replaced by integrals over the beam with \( S_{\text{max}} \) replaced by \( S_{\text{max,p}} \) times the beam response. Also, for each value of the integrand in Eqn. 10 the value at which \( dN/dS \) is truncated in the \( r(\omega) \) calculation will be different. This quickly becomes computationally intractable, but here we are actually helped by the expected steepness of the source distribution. If we define \( S_{\text{max,p}} \) as the largest flux contribution in a resolution element, the probability that the source contributing that flux is a source of intrinsic flux very close to \( S_{\text{max,p}} \) that lies near the center of the beam is far larger than the probability that the contributing source is a much brighter source far off beam center. This means we can approximate the correct versions of Eqs. 5 and 9 (that include integrals over the beam) with the original, infinite-resolution versions, using a value for \( \Delta \Omega_p \) that is roughly the area of the beam over which its response is near unity. In comparing to the simulated observations described in Sec. 2.4, we use 1 arcmin\(^2\), which is roughly the square of the full width at half maximum of the beam used in the simulations.

2.4. Comparison with simulations

We use simulated observations of mock skies including point-source populations drawn from model \( dN/dS \) distributions and Gaussian-distributed noise to test the formalism developed in the previous sections. Simulated observations were performed using three sets of observing parameters, each roughly consistent with the 2.0 mm maps presented in V09, which we use as a concrete example. The three sets of observing parameters used are: 1) No spatial filtering beyond binning into 1-arcmin pixels (and subtracting off the mean value of the map), noise consistent with the 2.0 mm SPT map shown in V09 (~ 1.4 mJy rms); 2) Noise as in 1, but with spatial filtering similar to the real SPT 2.0 mm maps in V09; 3) Filtering as in 1, but with the noise level halved. The source
count model that was used to create the simulated maps is the sum of the Negrello et al. (2007) 850 \textmu m counts for dust-dominated sources (scaled to 2.0 mm as in V09) and the De Zotti et al. (2007) counts for synchrotron-dominated sources.

From these simulated observations, we extract the true, underlying posterior flux PDF, $P(S_{\text{max}}|S_{p,m})$, for three different values of $S_{p,m}$ in each observing configuration. The true, underlying posterior flux PDF for a given value of $S_{p,m}$ was estimated by taking each source detected in the simulated maps with measured flux within $\delta S$ of $S_{p,m}$, finding every source in the true, underlying source population that was associated with that detection, and recording the flux of the brightest associated source as $S_{\text{max}}$ for that detection. In the pixel-only cases, sources were considered associated with a detection if they were in the same pixel as the detection; in the beam-and-filtering case, they were considered associated if they were within 1 arcmin of the position of the detection. The true, underlying $P(S_{\text{max}}|S_{p,m})$ for that value of $S_{p,m}$ is then simply the histogram of the $S_{\text{max}}$ values assigned to all the detections with measured flux within $\delta S$ of $S_{p,m}$. A total of 200 simulated 100 deg$^2$ maps were used to construct the histogram.

Fig. 1 shows the posterior flux PDFs extracted from the simulated observations and the calculated values of those posterior flux PDFs (using the formalism developed in the previous sections). The source model used to calculate the prior probability $P(S_{\text{max}})$ is the same model used to generate the mock point-source skies. The values of $S_{p,m}$ for which posterior flux PDFs are shown correspond to detection significance values of 4.5$\sigma$, 5.5$\sigma$, and 6.5$\sigma$ in each set of simulated observations. These detection significance values correspond to raw flux values of 6.3, 7.7, and 9.1 mJy in the full-noise simulations and 3.4, 4.1, and 4.9 mJy in the halved-noise simulations. (The detection significance in the halved-noise simulations for a given raw flux value are not exactly twice those in the full-noise simulations because of the contribution of background sources to the map rms.)

For comparison with the output of the pixel-only simulated observations, the posterior flux PDF was calculated exactly using the equations in Sec. 2.1 and Sec. 2.2. The top left and bottom left panels of Fig. 1 show that the calculated posterior PDF values for these cases are consistent with the simulation output to within the assumed Poisson errors. In the beam-and-filtering case, $r(\omega)$ was calculated exactly, but the approximation described in the previous section was used for the prior and for the $\delta(S_{\text{max}})$ term in Eqn. 1. The small but statistically measurable discrepancies between the simulated and calculated PDFs in the full-beams-and-filtering case (middle left panel of Fig. 1) are due to the imperfect nature of this approximation.

2.5. Gaussian likelihood approximation

The most computationally intensive step in estimating the posterior flux probability distribution $P(S_{\text{max}}|S_{p,m})$ is calculating the contribution from the other sources in the resolution element. For deep, single-dish mm/submm surveys, in which the angular resolution varies from $\sim$ 10 arcseconds to a couple of arcminutes, this contribution will be dominated by the background of sources below the confusion limit (the regime in which there are many sources per resolution element). By the central limit theorem, as the number of sources per resolution element becomes large, the distribution of total flux from these posterior sources will approach a Gaussian. If we use this fact to approximate the background source contribution to the posterior as another Gaussian-distributed noise source, the calculation of the likelihood in Eqn. 4 becomes trivial:

$$ P(S_{p,m}|S_{\text{max}}) = \frac{1}{\sqrt{2\pi\sigma_{\text{tot}}^2}} e^{-(S_{p,m} - S_{\text{max}} - \bar{S}_p)^2 / 2\sigma_{\text{tot}}^2}, \tag{12} $$

where $\bar{S}_p$ is the mean contribution from sources in the map, and $\sigma_{\text{tot}}$ is the quadrature sum of the instrumental and atmospheric noise and the fluctuations in the source background:

$$ \sigma_{\text{tot}}^2 = \sigma_n^2 + \sigma_{\text{sources}}^2. \tag{13} $$

There will always be a high-flux tail to the $P(D)$ distribution, because some resolution elements in the survey will have more than one source brighter than the confusion limit. However, for steep source populations, the mean number of sources contributing to the total flux in a pixel will be large, and the tails will be less important. Furthermore, because the background source contribution adds in quadrature to the instrument and atmospheric noise contribution (which is likely to be very well approximated by a Gaussian), non-Gaussianity in the source contribution will manifest itself in the total $P(D)$ distribution only if it is the dominant source of noise in a survey. For a survey like the SPT, this is not the case, and the Gaussian likelihood approximation holds very well, as shown in the top and middle right panels of Fig. 1. However, a survey that was a factor of two deeper at the same wavelength and resolution would incur much more significant errors by adopting the Gaussian likelihood approximation, as shown in the bottom right panel of Fig. 1.

It is worthwhile to note that adopting the Gaussian likelihood approximation removes one of the differences between the flux estimation method presented here and the standard method in the literature. In the Gaussian likelihood approximation, the implicit assumption is made that the only contributions to the measured flux in a pixel are instrumental and atmospheric noise, a single bright source, and a source background that acts as another symmetric noise source. In this case, the only difference between the likelihood of measuring some flux in a pixel given the flux of the brightest source in that pixel and the likelihood of measuring some flux in a pixel given the intrinsic total flux in that pixel is the mean contribution from the source background. Since all of the mm/submm experiments under consideration here are differential (i.e., not sensitive to the mean brightness on the sky), the two likelihoods are effectively identical. However, a significant difference remains between the priors used in the two methods, as discussed in Sec. 2.2.

2.6. Estimating source counts with a source-count prior

If the individual source fluxes estimated with this method (or with the standard method in the literature) are used in estimating source counts (as they are in V09 and most submm analyses), one might object that a prior probability has been assumed for the very quantity that
A method for individual source brightness estimation

Fig. 1.— Left Panels: True, underlying posterior flux PDF $P(S_{\text{max}}|S_{p,m})$ (extracted from simulated observations and shown by symbols with error bars) and calculated values for that PDF (using the procedure outlined in Sec. 2 and shown by lines). In the top two rows, the raw flux values for which the posterior PDFs are calculated are 6.3 mJy (diamond symbols and dashed line), 7.7 mJy (triangle symbols and dot-dashed line), and 9.1 mJy (square symbols and triple dot-dashed line). In the bottom row, those raw flux values are 3.4, 4.1, and 4.9 mJy. In both cases, these raw flux values correspond to detection significance levels of 4.5$\sigma$, 5.5$\sigma$, and 6.5$\sigma$. Vertical error bars on the extracted PDFs are from Poisson statistics; horizontal error bars show $\delta S = 0.2$ mJy, the width of the flux bins used to construct the $P(S_{\text{max}}|S_{p,m})$ histogram from the simulated observations (see text for details). Right Panels: Gaussian likelihood approximation to the posterior PDF calculation. Solid lines are from the full calculation and are identical to the lines in the left panel; dashed lines are calculated using the Gaussian approximation to the likelihood $P(S_{p,m}|S_{\text{max}})$ shown in Eqn. 12.

Top Row: binning into 1-arcmin pixels only, 1.4 mJy noise rms; Middle Row: full beam and filtering, 1.4 mJy noise rms; Bottom Row: binning into 1-arcmin pixels only, 0.7 mJy noise rms;
one is attempting to measure. It is important to remember that the source-count prior is applied probabilistically to determine each source’s flux, so the estimated source counts will only resemble the prior if the measurements have no constraining power. From the posterior flux PDFs shown in Fig. 4 it is clear that for the combination of data and prior used in these simulated observations, the data are providing the bulk of the information in the flux measurement down to the lowest detection significance shown (4.5σ). For a real survey, plots of individual posterior flux PDFs such as these are useful for determining where the reported source counts contain significant new information and where they simply reproduce the prior. Another useful check on the constraining power of the data is to vary the source count prior and confirm that the estimated counts do not change significantly (as in Scott et al. 2008 and V09).

3. MULTI-BAND SOURCE FLUX ESTIMATION

The posterior flux estimation method described in Sec. 2 implicitly assumes that both the data and the prior knowledge of the source distribution are restricted to a single observing band. The situation will inevitably be more complicated when both data and priors are available in multiple bands. In full generality, the task at hand is now to simultaneously estimate the posterior probability of the intrinsic source flux in multiple bands given the measured flux in those bands and any prior information. Using the formalism from Sec. 2 we can write the two-band case as

\[ P(S_{\text{max},1}, S_{\text{max},2}|S_{p,m,1}, S_{p,m,2}) \propto P(S_{p,m,1}, S_{p,m,2}|S_{\text{max},1}, S_{\text{max},2})P(S_{\text{max},1}, S_{\text{max},2}). \]

In the mathematically simplest case, both the conditional and prior probabilities on the right-hand side of Eqn. 14 are uncorrelated between bands, so that the posterior probability distribution for flux in the two bands is simply the product of the two single-band distributions:

\[ P(S_{\text{max},1}, S_{\text{max},2}|S_{p,m,1}, S_{p,m,2}) \propto P(S_{p,m,1}|S_{\text{max},1})P(S_{p,m,2}|S_{\text{max},2})P(S_{\text{max},1}, S_{\text{max},2}). \]

However, the assumption of uncorrelated prior information in the two bands will very rarely be valid. The prior in each band will be derived from previous estimates of source counts at wavelengths and flux levels as near as possible to that band, using assumptions about spectral behavior to extrapolate to the bands of interest. For bands that are reasonably close to each other in wavelength, the existing data that go into estimating the priors in the two bands will almost certainly have some overlap. This is not simply a matter of imperfect priors; in the limit of perfect prior knowledge of the source counts in both bands, the priors in the two bands would only be independent if the source populations measured in the two bands had zero overlap. In the example of the SPT data at 1.4 and 2.0 mm analyzed in V09, the prior probability \( P(S_{\text{max}}) \) in both bands is dominated in the few-mJy flux region by the dusty starburst population, with some contribution from synchrotron-dominated AGN. It is clearly a poor approximation to assume that the priors in these two bands are uncorrelated.

It is possible to construct the full two-dimensional prior probability \( P(S_{\text{max},1}, S_{\text{max},2}) \); however, it will often be more convenient to change variables and cast this prior probability in terms of the flux in one band and the expected spectral behavior of the sources contributing to the prior. We define the spectral index \( \alpha \) through the relation

\[ S(\lambda_2) = S(\lambda_1) \left( \frac{\lambda_2}{\lambda_1} \right)^{-\alpha} \]

and write the two-dimensional prior as \( P(S_{\text{max},1}, \alpha) \). The only requirement for this prior to factor into independent priors \( P(S_{\text{max},1}) \) and \( P(\alpha) \) is that the spectral index distribution between the two bands not depend on flux. Of course, in reality \( P(\alpha) \) will depend somewhat on flux — for example, in the SPT case, the brightest sources are mostly AGN, but near the confusion limit the source population is dominated by dusty galaxies. In cases such as the analysis of SPT data in V09, in which estimating the spectral index distribution of the sources is one of the key goals of the analysis, the prior used will be sufficiently weak (V09 use a flat prior from \(-3 \leq \alpha \leq 5\)) that the slight dependence on flux of the real \( P(\alpha) \) is of no consequence. Alternatively, we can accept the added complexity of using the full two-dimensional prior on \( S_{\text{max}} \) and \( \alpha \) — assuming sufficient data exist to meaningfully construct such a distribution.

No matter how we choose to construct the spectral index prior, we now have the two-dimensional posterior PDF of flux in one band and \( \alpha \):

\[ P(S_{\text{max},1}, S_{\text{max},2}|S_{p,m,1}, S_{p,m,2}) \propto P(S_{p,m,1}|S_{\text{max},1})P(S_{p,m,2}|S_{\text{max},2}) \]

If we choose, we can then transform this into a two-dimensional posterior probability distribution for the flux in both bands:

\[ P(S_{\text{max},1}, S_{\text{max},2}|S_{p,m,1}, S_{p,m,2}) = P(S_{\text{max},1}, S_{\text{max},2}) \frac{d\alpha}{dS_{\text{max},2}}, \]

where \( d\alpha/dS_{\text{max},2} \) is derived from Eqn. 10.

3.1. Choice of detection band

We note that as the prior on \( \alpha \) is relaxed to a flat prior between \(-\infty \) and \( \infty \) (independent of source flux), the posterior PDF for \( S_{\text{max},1} \) in Eqn. 17 reduces to the single-band posterior of Eqn. 2. Meanwhile, the posterior PDF for \( S_{\text{max},2} \) becomes equal to the likelihood \( P(S_{p,m,2}|S_{\text{max},2}) \) — i.e., there is effectively no prior on \( S_{\text{max},2} \). This points out an apparent asymmetry between bands in our approach, namely that the two-band posterior flux PDF will depend on which band you choose to use prior source count information from — call this the “detection band” — and which band you apply prior information to only through the combination of the prior on \( \alpha \) and source-count prior on the first band.

The real issue is that \( dN/dS \) in each band and a distribution in \( \alpha \) are not three independent pieces of information; rather, the choice of \( dN/dS \) in one band and a distribution in \( \alpha \) uniquely specifies \( dN/dS \) in the other band. If the assumptions regarding \( dN/dS \) in both bands and the distribution of \( \alpha \) are internally consistent, then the two-band posterior flux PDF will be identical regardless of which band is chosen as the detection band. However, if one of the goals of the analysis is to measure the
distribution of spectral indices, such a prior on \( \alpha \) may be too restrictive. In V09, the choice was made to adopt a non-restrictive prior \((-3 \leq \alpha \leq 5)\) and analyze the data using each band in turn as the detection band. The differences in the final derived source counts between the two analyses were small compared to the statistical uncertainties.

### 3.2. Variable sources

The members of at least one of the populations expected to contribute to mm/submm source counts — namely AGN — are expected to have highly time-variable flux. This complicates any estimate of source properties from data taken at different epochs. For instruments that observe simultaneously in multiple wavelength bands — as the SPT does — this is not an issue, and we have ignored any effects of variability on the method presented here. This would be an issue, however, in extending this method to radio survey data.

### 3.3. Gaussian likelihood approximation in the multi-band case

As in the single-band case, the calculation of the likelihood \( P(S_{\text{max},1}, S_{\text{max},2}|S_{\text{max},1}, \alpha) \) is greatly simplified by the assumption that the instrumental and atmospheric noise and the contribution from sources below \( S_{\text{max}} \) are Gaussian-distributed. While the full, non-Gaussian two-dimensional likelihood is in principle calculable — by extending the result of Scheuer (1957) into a multivariate Fourier operation — that derivation is beyond the scope of this work. It is also possible to estimate the full, non-Gaussian two-dimensional distribution by simulated observations. For now, we will adopt the Gaussian likelihood approximation and use simulated observations to check its validity, as we did for the single-band results.

Under the Gaussian likelihood approximation, the two-band likelihood (analogous to the single-band likelihood in Eqn. (12)) is given by

\[
P(S_{p,m,1}, S_{p,m,2}|S_{\text{max},1}, \alpha) = \frac{\exp\left(-\frac{1}{2} r^T C^{-1} r\right)}{2\pi \sqrt{\text{det} C}},
\]

where \( C \) is the noise covariance between the bands (including contributions from instrument noise, atmosphere, and sources fainter than \( S_{\text{max}} \)), and \( r \) is the residual vector

\[
\begin{align*}
S_{p,m,1} - S_{\text{max},1} - S_{p,1} \\
S_{p,m,2} - S_{\text{max},2}(\alpha) - S_{p,2} \\
S_{p,m,1} - S_{\text{max},1} (\lambda_2 - \lambda_1) - S_{p,2}
\end{align*}
\]

### 3.4. Comparison to simulations

Analogous to Fig. 1 for the single-band case, Fig. 2 shows comparisons between calculated versions of the two-dimensional posterior distribution \( P(S_{\text{max},1}, S_{\text{max},2}|S_{p,m,1}, S_{p,m,2}) \) and the true values of those distributions (extracted from simulated observations). The simulated observations were performed by populating fake skies at two observing wavelengths (1.4 mm and 2.0 mm) with a single population of sources with an underlying Gaussian distribution of spectral indices \( \alpha = 2.7 \pm 0.3 \), similar to the spectral index distribution for high-redshift, dusty galaxies derived in Knox et al. (2004). The number counts as a function of flux for the source population come from the dusty galaxy counts at 850 \( \mu \)m; this model was also used to construct the source flux prior.
Noise was added to the fake skies in each band at a level similar to that in the corresponding SPT bands in V09. To avoid confusing effects of slight misestimations of priors due to beam and filtering with fundamental issues in the two-band implementation, this set of simulated observations involved no spatial filtering beyond binning into 1-arcmin pixels. Posterior two-band flux PDFs were extracted from the simulated observations as in the single-band case, namely by finding the brightest source associated with each detection in the simulated maps and constructing a two-dimensional histogram of \( \{S_{\text{max,1}}, S_{\text{max,2}}\} \) for every pair of measured flux values.

True and calculated values of the posterior PDF are shown in Fig. 2 for two values of measured two-band flux. These raw flux values at 1.4 mm and 2.0 mm — \{14.2, 7.0\} mJy in the top panel of Fig. 2 and \{17.2, 6.0\} mJy in the bottom panel of Fig. 2 — correspond to detection significances of \{4.2, 4.9\} and \{4.9, 4.1\}. These values were chosen to illustrate the importance of using the full two-band information as opposed to calculating each band’s posterior flux PDF individually and ignoring correlations in the two bands’ prior information.

Three different calculated values of the posterior are shown in each panel. Two versions of the calculated posterior use the two-band implementation described in Sec. 2; one using a flat prior on \( \alpha \) \((-3 \leq \alpha \leq 5\) and one using a prior on \( \alpha \) that is equal to the true underlying \( \alpha \) distribution (a Gaussian with \( \bar{\alpha} = 2.7 \) and \( \sigma_\alpha = 0.3 \)). In each of these cases, the band in which the source was detected more significantly was used as the detection band. The last version of the two-band posterior flux PDF shown in Fig. 2 is the product of the individual single-band posterior PDFs, each calculated using the procedure outlined in Sec. 2 and assuming no correlations between the prior information in each band.

The large amount of information in Fig. 2 can be boiled down to three main points:

1. If one had perfect prior information on \( dN/dS \) and \( \alpha \), one could calculate the posterior two-band flux PDF for every source perfectly.

2. With far less restrictive priors on the \( \alpha \) distribution, one can make an estimate of the posterior two-band flux PDF for every source that has no strong bias but has somewhat less constraining power.

3. Calculating the posterior two-band flux PDF by de-boosting each source individually and assuming no correlation between the priors in each band can result in highly biased posterior distributions. The posterior flux estimate in the band in which the source is detected strongly is reasonable, but the flux estimate in the other band is de-boosted to the confusion limit. The spectral index inferred from this calculation is actually a worse estimate of the true index than using the measured flux in each band uncorrected for boosting (as shown by the black crosses in Fig. 2). This highlights the importance of accounting for the correlations in the prior information used to de-boost multi-band fluxes, particularly when dealing with a source population for which the number counts are a steeply falling function of flux.

3.5. Classification of sources through their spectral behavior

![Figure 3](image_url)

**Fig. 3.** — Fraction of sources in simulated observations which would be misclassified by a posterior spectral index PDF criterion which divides sources into synchrotron- or dust-dominated populations based on the value of \( P(\alpha \geq 1.5) \). See Sec. 3.5 for details of the simulated observation. Diamonds and solid lines show the misclassification fraction obtained using the two-band posterior flux estimation described in Sec. 2 independently in both bands and ignoring the correlations in the two bands’ priors. **Top Panel:** fraction of all sources misclassified as a function of 2.0 mm detection significance. **Bottom Panel:** fraction of sources misclassified which had \( P(\alpha \geq 1.5) > 0.9 \) or \( P(\alpha \geq 1.5) < 0.1 \), also as a function of 2.0 mm detection significance. The feature at \( \sim 4.7 \sigma \) in the dashed curve in the bottom panel is an artifact of the specific noise levels chosen for the two bands (for details, see Sec. 3.5).

One key use of spectral information is to separate sources into different populations. For example, V09 use the posterior probability distribution of \( \alpha \) for every detected source to classify it as either synchrotron- or dust-dominated. This allows the source counts in each population to be compared to models of that population and to counts at other wavelengths where that population is dominant. Any bias in the posterior two-band flux PDF of a source could cause sources to be misclassified, leading to a bias in the estimation of source counts for both populations.
We have used simulated observations very similar to those used in Sec. 3.4 to investigate how often sources are misclassified using the two-band posterior flux PDF estimated with and without accounting for correlations in the prior information between bands. The only difference in this set of simulated observations and those of Sec. 3.4 is that there are two populations used to create the fake skies that are observed by our instrument. The two populations are the dust-dominated population used in Sec. 3.4 plus a synchrotron-dominated population with source counts as in De Zotti et al. (2003) and a Gaussian spectral index distribution with \( \alpha = -0.7 \pm 0.5 \), roughly consistent with the spectral behavior of the brightest synchrotron-dominated sources in V09.

For each detection in the simulated maps, the brightest source associated with that detection was identified, and the posterior spectral index PDF for the source was calculated by initially calculating the two-dimensional posterior PDF for flux in one band and \( \alpha \) and then marginalizing over the flux variable to create a one-dimensional spectral index PDF. For this exercise, we only use the broad, flat \( \alpha \) prior from Sec. 3.4 \( (-3 \leq \alpha \leq 5) \). The posterior \( \alpha \) PDF was then compared to the true spectral index of the brightest source associated with that detection. We classified detections in the simulated maps as synchrotron-dominated if \( P(\alpha \geq 1.5) < 0.5 \) and dust-dominated if \( P(\alpha \geq 1.5) > 0.5 \); similarly, we classified the brightest source associated with the detection as synchrotron- or dust-dominated according to whether its true spectral index was greater than or less than 1.5. The posterior \( \alpha \) PDF for each detection was also estimated by calculating the posterior flux PDF in each band independently using the procedure outlined in Sec. 2 — ignoring any correlations between the prior information in the two bands — and combining the two flux PDFs to create a PDF for \( \alpha \). This \( \alpha \) PDF was used to classify sources similarly to the two-band \( \alpha \) posterior.

The fraction of sources misclassified (labeled as synchrotron-dominated using the posterior \( \alpha \) PDF when the brightest associated source was in fact dust-dominated, or vice-versa) in each case is shown as a function of single-band detection significance in Fig. 3. The two-band implementation — even with the weak prior on \( \alpha \) — shows a clear improvement in misclassification fraction over combining independent single-band PDFs at all significance levels up to 7\( \sigma \). Particularly striking is the difference in “high-confidence” misclassifications — instances in which the classification based on the posterior \( \alpha \) PDF was at the 90% confidence level or greater, but was wrong. As shown in the bottom panel of Fig. 3, the \( \alpha \) estimation based on independent single-band PDFs has up to a 25% rate of high-confidence misclassifications, but rate for the two-band implementation is effectively zero.

The turnover at \( \sim 4.7 \sigma \) in the rate of high-confidence misclassifications using the single-band information and ignoring correlations (Fig. 3 bottom panel, dashed curve) is due to the relative noise level in the two bands. Because the noise at 1.4 mm is more than twice that at 2.0 mm, there are many sources that are intrinsically dust-dominated which are nevertheless detected more significantly at 2.0 mm. If such a source is detected above 4.5\( \sigma \) at 2.0 mm but below 4.5\( \sigma \) at 1.4 mm, the posterior flux PDF for that source will be centered near the raw, detected flux at 2.0 mm but de-boosted to the confusion limit at 1.4 mm. This results in a robust “measurement” of a negative spectral index for this source and, hence, a high-confidence misclassification. If the source is detected below 4.5\( \sigma \) in both bands, there is effectively no constraint on \( \alpha \) using this procedure, so the source may be misclassified, but not at high confidence.

3.6. More than two bands

The two-band formalism laid out in Sec. 3 is sufficient for the V09 analysis of two-band SPT data. However, data in three or more bands at mm/submm wavelengths and mJy flux levels have been or are currently being collected by the Balloon-borne Large-Aperture Submillimeter Telescope (BLAST, Devlin et al. 2009), the SPT, and the Atacama Cosmology Telescope (ACT; Fowler et al. 2007). Furthermore, the recent launch of Herschel\(^9\) and Planck\(^{10}\) means that we will soon have simultaneous measurements of mm/submm sources in as many as seven bands (depending on where you choose to define the limits of the mm/submm spectral region).

Fortunately, the two-band formalism laid out in Sec. 3 is easily extended to more than two bands, although the calculation necessarily becomes more complex. For the case in which the Gaussian approximation holds and each source’s spectral behavior can be described by a single power-law index across all bands, then the multi-band calculation is a trivial extension of Eqs. 17-20. A first step in relaxing the assumption of a single spectral index for each source would be allowing a break in the spectrum such that each source would have a single spectral index between any two bands. The spectral index prior would then be a function of \( N_{\text{bands}} - 1 \) variables \( \alpha = \{\alpha_2, \alpha_3, ..., \alpha_{N_{\text{bands}}}\} \). These variables would necessarily be highly correlated, so we would require the full \( (N - 1) \)-dimensional prior. A full treatment of the \( N \)-band version of the method, including tests on simulated observations, will be the subject of future work.

4. CONCLUSIONS

We have constructed a method for reliable, minimally biased estimation of single-band and multi-band properties of individual sources from noisy data. We find that proper treatment of correlated prior information in the multi-band version of the method is crucial to avoid significant biases in estimates of multi-band fluxes and spectral behavior. The single- and multi-band implementations of the method have been verified through simulated observations of mm data, and the two-band implementation has been used to estimate source fluxes and spectral behavior in SPT data (Vieira et al. 2009). This method, or an extension thereof to more than two bands, is directly applicable to source analyses for most current and upcoming mm/submm experiments, includ-

\(^9\) De Zotti et al. (2003)
\(^{10}\) BLAST, Devlin et al. 2009
\(^\) ACT; Fowler et al. 2007
\(^\) Herschel\(^{12}\) and Planck\(^{13}\)

\(^{11}\) Millimeter Telescope
\(^{12}\) http://www.blastexperiment.info
\(^{13}\) http://www.ess.int/SCIENCE/Herschel
ing BLAST, ACT, Planck, and Herschel and should also be applicable to data taken at other wavelengths.

This work was supported by the National Science Foundation through grants ANT-0638937 and ANT-0130612 and the NSF Physics Frontier Center grant PHY-0114422 to the Kavli Institute of Cosmological Physics (KICP) at the University of Chicago. ERS acknowledges support from a KICP fellowship. The authors would like to thank members of the SPT collaboration, Jason Austermann, and the anonymous referee for useful discussions and helpful suggestions.
REFERENCES


