Simulation Pipeline for Velocity Field Measurements using the Kinetic Sunyaev-Zel'dovich Effect

Thesis by
Yubo Su

In Partial Fulfillment of the Requirements for the
degree of
Bachelors of Science

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
Defended May 18, 2016

Yubo Su
ORCID: N/A

ACKNOWLEDGEMENTS

The completion of this thesis is greatly indebted to many mentors old and young and of various qualifications. Without them, I would not have had the rewarding experience of bringing together such a lengthy oeuvre, which even as it bears the clear imprint of undergraduate work is still an experience upon which I look back fondly.

I would like to thank Professor Sunil Golwala for patiently mentoring my research over the last year and a half. Many times when I either lacked the prior knowledge or the time to carry on, he has given me both the time and insight to push forward at whatever pace I could muster alongside a heavy courseload, and I am deeply grateful for his unwavering guidance.

I would like to thank Dr. Jack Sayers and fellow lab member Ted Macioce for their advice throughout my project. Being held to the standard of those far more advanced than me and obtaining their illuminating insight has pushed this project and my personal development forward in ways that cannot be set down on paper.

The various professors and mentors that I have had over the course of my four years at Caltech have also indescribably shaped my passion for the vast discipline that falls under the umbrella term "physics." The excitement of all the faculty with whom I have had the pleasure of taking classes and the inspiring atmosphere of working on problem sets with incomprehensibly talented classmates are facets of Caltech that I can never forget and that have left an indelible impression on my thinking and growth.

Finally, I would like to thank my ever-supportive family. All parents nag, and all kids get annoyed by it, but I hope you know that I, somewhere deep down on the inside, do appreciate everything you have done for me since the day I was born ☺.

ABSTRACT

The kinetic Sunyaev-Zel'dovich (kSZ) effect is of great cosmological interest for providing precision measurements of peculiar velocity fields independent of systematics present in other cosmological probes. The high precision kSZ measurements on the horizon are expected to yield valuable constraints distinguishing between competing cosmological models. Instrumental to kSZ astromony is the removal of contaminating point sources, primarily radio and dusty, star-forming galaxies that have bright emissions in kSZ spectral bands. As the precision of measurements improve, the source removal residuals may become significant contributions to the overall kSZ error budget. A full simulation is essential to characterizing these induced errors. This study develops a procedure for contamination removal from first principles, verifying its optimality against information theoretic limits. Models for contaminating sources of increasing complexity are considered, at each step characterizing the optimality of the subtraction compared to theoretical bounds. The final, currently unfinished objective is to apply this subtraction procedure to a realistic source distribution and understand the incurred systematics.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 The Sunayev-Zel'dovich Effect

The Sunyaev-Zel'dovich (SZ) effect describes how the intra-cluster media (ICM) in galaxy clusters scatters the cosmic microwave background (CMB) signal (Sunyaev and Zeldovich, 1972; Birkinshaw, 1999; Glenn et al., 1998a). Incoming CMB photons are inverse Compton scattered off energetic electrons and acquire a higher energy (are blue-shifted). This results in a deficit of electrons at wavelengths below the CMB blackbody peak and a surplus at higher wavelengths, as depicted in Figure 1.1.

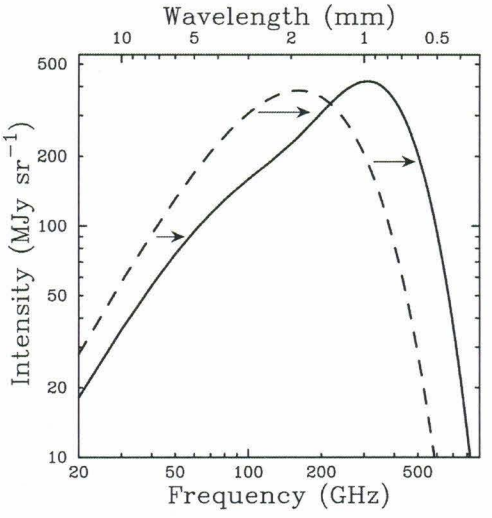


Figure 1.1: Exaggerated depiction of SZ wavelength shift (Carlstrom, Holder, and Reese, 2002).

Three SZ effects are distinguished by the contribution to the electrons' energy in each:

- The Thermal SZ effect (tSZ) describes the SZ effect due to thermal component of electron energy.
- The Kinetic SZ effect (kSZ) describes the SZ effect due to bulk motion contribution to electron kinetic energy.

- The Relativistic SZ effect (rSZ) describes the SZ effect due to relativistic corrections to electron thermal energy. Generally considered as a perturbation on the tSZ effect.

The great importance of SZ effects is that they provide probes into cosmological parameters that both are independent of redshift and carry distinct spectral signatures. This makes them both easy to identify and useful at a wide range of cosmological distances. The SZ effects are also independent of many other observational techniques and are subject to orthogonal systematics, providing complementary estimates on a variety of measurements (Carlstrom, Holder, and Reese, 2002). The relative magnitude of the tSZ and kSZ effects are presented in Figure 1.2.

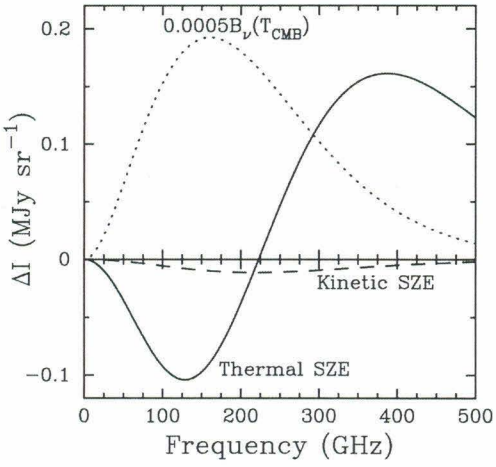


Figure 1.2: Depiction of relative scales of the tSZ, kSZ effects, with 0.0005 the CMB blackbody flux depicted for scale (Carlstrom, Holder, and Reese, 2002).

Thermal SZ Effect

The thermal SZ effect, derived from thermal contributions to electron energy, fluctuates CMB photons by the following equation

$$\frac{\Delta T_{CMB}}{T_{CMB}} = f(\nu, T_e) \int n_e \sigma_T \frac{k_B T_e}{m_e c^2} \mathrm{d}l \tag{1.1}$$

with electron density n_e, σ_T Thomson cross section, k_B the Boltzmann constant, T_e the electron temperature, m_e the electron mass, c the speed of light integrated dl along the line of sight (Sayers et al., 2013). $f(\nu, T_e)$ is the frequency dependence which is denoted in terms of $x \equiv \dfrac{h\nu}{k_B T_{CMB}}$

$$f(x) = \left(x \frac{e^x + 1}{e^x - 1} - 4 \right) (1 + \delta_{SZE}(x, T_e)) \tag{1.2}$$

where δ_{SZE} denotes the relatistic corrections typically of order a few percent for most of the spectrum (Carlstrom, Holder, and Reese, 2002; Itoh, Kohyama, and Nozawa, 1998). Among other useful applications, the tSZ effect is of great use in conjunction with X-ray techniques in distance determinations to clusters e.g. (Cavaliere, Danese, and Zotti, 1979; Birkinshaw, 1979).

Kinetic SZ Effect

The change in the temperature of CMB photons due to the kSZ effect when passing through some galaxy cluster is given

$$\frac{\Delta T_{CMB}}{T_{CMB}} = -\frac{v_z}{c}\tau_e \tag{1.3}$$

with v_z the ICM peculiar velocity along the line of sight and τ_e the total electron optical depth (Sayers et al., 2013)

$$\tau_e = \int n_e \sigma_T \, \mathrm{d}l. \tag{1.4}$$

The kSZ effect is particularly interesting as a cosmological probe of peculiar velocities, velocity deviations from Hubble's Law or equivalently velocities relative to CMB rest frame. It is the only known technique to measure large scale velocity fields at high redshift (Carlstrom, Holder, and Reese, 2002).

Research Uses of the kSZ Effect

In contrast with established measurement techniques, the kSZ effect provides a direct measurement of line-of-sight velocities v_z of hot electrons within galaxy clusters (Sunyaev and Zeldovich, 1972), useful for constraining cosmological parameters. While spectroscopy is already an actively used technique for constraining such parameters (e.g. the total matter density Ω_m and the normalization of density fluctuations σ_8 (Feldman, Watkins, and Hudson, 2010; Ma, Branchini, and Scott, 2012; Nusser and Davis, 2011)), kSZ measurements are still invaluable as these two techniques exhibit vastly different systematic uncertainties. For instance, a kSZ measurement is independent of redshift and depends only on the electron line-of-sight velocity and optical depth, local characteristics, while spectroscopy requires averaging over many galaxies which decreases the resolution of the measurement. Studies conclude that in efforts to constrain to dark energy parameters or deviations from general relativity, kSZ measurements would probe different mass scales and exhibit different systematic uncertainties while still maintaining a similar level of accuracy. By including kSZ measurements, combining constraints from

various measurements can then produce a tighter overall constraint on interesting parameters (Albrecht et al., 2006).

Recent observations of the kSZ effect have proven difficult but feasible. Data from the *WMAP* and *Planck* satellites, the two most recent CMB datasets, have been used to place upper limits on the bulk flows and root mean square (RMS) variations of v_z via the kSZ signal (Kashlinsky et al., 2008; Osborne et al., 2011). A combination of other data was also used to constrain the mean pairwise momentum of clusters that is inconsistent with noise at a confidence level of 99.8% (Hand et al., 2012). Most recently, kSZ measurements using Bolocam and *Herschel* Spectral and Photometric Imaging Receiver (SPIRE) data have been able to constrain v_z measurements of galaxy cluster MACS J0717.5+3745 to 3450 $\pm$ 900km/s (Sayers et al., 2013).

1.2 Contamination

These above results demonstrate both the feasibility of kSZ measurements and their utility in constraining cosmological parameters. With the next generation of detectors, kSZ measurement precision is expected to improve as statistical uncertainties decrease. Accordingly, systematic uncertainties on the separation of the kSZ signal from other astronomical signals will become more important.

In particular, extragalactic infrared point sources such as sub-millimeter galaxies and radio galaxies have significant emissions in kSZ frequencies and are the major source of contamination in SZ surveys (Carlstrom, Holder, and Reese, 2002). Since these galaxies remain point-like down to scales far below present SZ resolution, these contaminating sources can effectively be treated as point sources. As such, they exhibit significantly higher spatial flux density than the SZ signal and must be subtracted to examine the underlying phenomena.

Sub-millimeter Galaxies

A large portion of sub-millimeter galaxy luminosity is dominated by dusty, star-forming galaxies (DSFGs), reviewed extensively by (Casey, Narayanan, and Cooray, 2014). These galaxies are completely optically obscured, yet the dust heated by ultraviolet radiation of newly forming stars produces infrared luminosities in excess of $10^{13} L_\odot$. This implies a star formation rate that is in the thousands of $M_\odot$/yr compared to the Milky Way's paltry $2M_\odot$/yr (Robitaille and Whitney, 2010). Even so, they are extremely spatially compact, approximately 2 ± 1kpc across corresponding to an angular width of $\lesssim 1''$ and so exhibit point-source like beam profiles (Casey, Narayanan, and Cooray, 2014).

This heated dust forms the primary infrared emission of these DSFGs, and the resulting SED is well described by a modified black-body, or 'greybody' SED, for wavelengths $\gtrsim 50\mu$m. The greybody SED as a function of spectral frequency ν is given by

$$S(\nu|\beta,T) = \times \left(1 - e^{-\tau(\nu)}\right) \times B(\nu,T)$$
$$\tau(\nu) = \left(\frac{\nu}{\nu_0}\right)^{\beta}$$

where T is the redshifted temperature $T = T_{dust}\dfrac{1}{1+z}$ and τ is is the optical depth, fitted as a function of ν_0 the frequency at which the optical depth equals unity (Draine, 2006). Typical redshifts for DSFGs with peak emissions in the far infrared range from 2.2–2.7 (Casey, Narayanan, and Cooray, 2014). β is referred to as the emissivity and is generally found to be $\in [1,2]$ and commonly assumed to be 1.5 (Casey, Narayanan, and Cooray, 2014). ν_0 is estimated from laboratory experiments to be 3THz which is an order of magnitude above the model instrument $\nu \leq 400$GHz. Experimental studies indicate its value is closer to $\nu_0 \approx 1.5$THz (Conley et al., 2011; Rangwala et al., 2011). Often, we make the optically thin approximation $\left(1 - e^{-\tau(\nu)}\right) \approx \left(\dfrac{\nu}{\nu_0}\right)^{\beta}$, valid for observation wavelength $\gtrsim 450\mu$m, and obtain for SED

$$S(\nu|\beta,T) = \left(\frac{\nu}{\nu_0}\right)^{\beta} \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_B T}} - 1} \tag{1.5}$$

1.3 Research Goal

The above presents DSFGs to be point-source contaminations with known SED. This work focuses on a first-principles approach to subtracting these contaminating point sources from multi-band data. The objective is to apply signal processing theory to subtract point sources from astronomical maps containing both stochastic instrumental noise and confusion noise from sources with fluxes below some detection threshold. Using this formalism, we identify both procedures to compute optimal estimators and the theoretical limit on their accuracy. We are able to compare this with results from simulations containing randomly generated noise and ensure our procedure saturates the theoretical subtraction accuracy limit. The theory is developed starting from a constrained setup that is successively relaxed to better describe real contamination sources.

While there exist studies performing fits of these data to various SED profiles (Sayers et al., 2013; Casey, 2012), studies of the optimality of such procedures and the in-

troduced errors are lacking. This work aims to begin to fill that gap with a procedure that is tested against theoretical fit accuracy limits in constrained environments that is subsequently applied to contaminating sources with known SEDs to characterise subtraction systematics. The procedure is laid out for generalizing this to SEDs of arbitrary complexity.

The code for this project is hosted at `https://github.com/yubo56/Bolocam_Source_Subtraction`.

Chapter 2

DATA SIMULATION

The entirety of the study is performed using simulated data. The parameters of the simulated data are described below.

2.1 Instrument Parameters

The modeled instrument in this study is taken roughly after the Bolocam instrument at the Caltech Submillimeter Observatory to have the following properties (Glenn et al., 1998b):

Field of View 480 arcseconds

Resolution 256 pixels

Observational Frequencies Frequency bands centered at

$$[400\text{GHz}, 353\text{GHz}, 273\text{GHz}, 231\text{GHz}, 150, 91\text{GHz}]$$

corresponding to wavelengths

$$[750\mu\text{m}, 850\mu\text{m}, 1.1\text{mm}, 1.3\text{mm}, 2.0\text{mm}, 3.3\text{mm}]$$

Point Spread Function Gaussian with width $\propto \frac{1}{\nu}$ observational frequency (diffraction limited). Taken to be 15arcsec FWHM in the 750μm band,

Confusion Noise Tabulated in Table 2.1, obtained from Béthermin simulations (Béthermin et al., 2011).

Stochastic Noise Assumed to be a sum of white noise at half the confusion limit and a small $\nu^{-8/3}$ noise component (generally taken to be on the order of the white noise at $\nu \sim \frac{1}{60}\text{arcsec}^{-1}$ the lowest spatial frequency in the map; we call this spatial frequency the *elbow* of the $\nu^{-8/3}$ noise). The $\nu^{-8/3}$ noise component corresponds to atmospheric noise (Sayers et al., 2010).

No treatment of any instrumental imperfections is handled in this study; the data are assumed to be corrected for any distortions and defects.

Wavelength	Confusion Limit (Jy)
750μm	1.81×10^{-4}
850μm	1.37×10^{-4}
1.1mm	1.12×10^{-4}
1.3mm	9.47×10^{-5}
2.0mm	4.99×10^{-5}
3.3mm	2.43×10^{-5}

Table 2.1: Confusion limit for each wavelength bands.

2.2 Data Model

Instrumental data consist of 2-dimensional discrete arrays (maps) of spectral flux density (often referred to simply as flux) values. We notate such maps v_{rs} with r, s indicies into the array, or sometimes just $v(\vec{x})$ where $\vec{x}$ is understood to be discrete indicies. The data are taken across six frequency bands ν, each of the six maps referred to as $v_\nu(\vec{x})$. The six maps are assumed to overlapping in coordinates just taken at different spectral frequencies.

Given a set of maps $v_\nu(\vec{x})$, we assume that the maps comprise some set of point sources overlaid on a stochastic background. In other words, the maps $v_\nu(\vec{x})$ can be decomposed as

$$v_\nu(\vec{x}) = n(\vec{x}) + \sum_i A_i s_\nu(\vec{x}|\lambda_i) \tag{2.1}$$

with i indexing the sources in the map, $s_\nu(\vec{x})$ the known source profile as a function of some source parameters λ_i that may differ per map, A_i the total flux from the ith source and $n_\nu(\vec{x})$ some stochastic noise obeying some known power spectral density (PSD). We will generally notate a function $f(x|\lambda)$ to be a function of x with parameters λ. With a known PSD, we can generate realisations of $n_\nu(\vec{x})$ to simulate instrumental data; see Section 3.2 for the associated procedure.

Because we only consider point sources, $s_\nu(\vec{x})$ is given solely by the point spread function of the instrument, which as discussed in Section 2.1 is taken to be Gaussian with width fixed at 15arcmin in the 750μm band. In order then for A to be the total flux from the source, we require that s be normalized to have unit integral. Due to the discretization of the map, this normalization can be mandated in two ways, either the sum of s over all pixels in the map or as the integral of a continuous Gaussian with the same width σ. These two can differ if s is not centered on one of the observational pixels. We choose the latter because it corresponds to the total flux emitted from the source rather than the total flux observed.

2.3 Point Source Distribution

In the earlier stages of the study, where the accuracy of the formalism is the primary focus, point sources with arbitrary fluxes and spatial location are used. In the later stages of the study, where accuracy on realistic data is of increased interest, we intend to use the simulated point source flux distribution from Béthermin to draw quantitative conclusions on systematic errors (Béthermin et al., 2011; Béthermin, 2014).

Chapter 3

MATHEMATICS

3.1 Conventions

Below we discuss some of the conventions used in this work.

Fourier Transforms

We begin with continuous Fourier transforms. $g(\vec{x}), \tilde{g}(\vec{k})$ for $\vec{x}, \vec{k} \in \mathbb{R}^2$ are related by the forward and reverse transforms

$$\tilde{g}(\vec{\nu}) = \iint \mathrm{d}^2\vec{x}\; g(\vec{x}) e^{-j\vec{k}\cdot\vec{x}} \tag{3.1}$$

$$g(\vec{x}) = \iint \mathrm{d}^2\vec{\nu}\; \tilde{g}(\vec{\nu}) e^{j\vec{k}\cdot\vec{x}}, \tag{3.2}$$

where $j = \sqrt{-1}, \vec{k} = 2\pi\vec{\nu}$ and the integration is performed over the entirety of $\mathbb{R}^2$ unless otherwise noted. This sign convention is used so the correspondence $\frac{\mathrm{d}}{\mathrm{d}t} \leftrightarrow j\omega$ holds. This 2π convention is used such that the units of $\tilde{g}(\vec{\nu})$ are just units of 1/Hz times those of $g(\vec{x})$. A symmetric $(2\pi)^{-1/2}$ convention unduly obfuscates the units.

Convolution is then defined under this sign convention as

$$\left[\tilde{g}(\vec{\nu})\tilde{h}(\vec{\nu})\right] \overset{\mathcal{FT}}{\longleftrightarrow} [g * h](\vec{x}) \equiv \iint \mathrm{d}^2\vec{x}_1\; g(\vec{x}_1) h(\vec{x} - \vec{x}_1). \tag{3.3}$$

Discretization

To work with the discrete, pixelated maps, we must work with the discrete forms of the above formulae. Define x_1, x_2 to be pixel indicies that range over $\left[-\frac{N}{2}, \frac{N}{2} - 1\right]$ and define $\Delta L = \frac{L}{N}$ the pixel width of the map for L the width of the map in arcminutes. We identify the below correspondences

$$\begin{aligned} \iint \mathrm{d}^2\vec{x} &\to \sum_{x_1=-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{x_2=-\frac{N}{2}}^{\frac{N}{2}-1} (\Delta L)^2 \\ &\equiv \sum_{x_1,x_2=-\frac{N}{2}}^{\frac{N}{2}-1} (\Delta L)^2 \end{aligned} \tag{3.4}$$

$$\tilde{f}(\vec{\nu}) \to \frac{\tilde{f}_{mn}}{(\Delta\nu)^2} = L^2 \tilde{f}_{mn}, \tag{3.5}$$

where the shorthand (3.4) is defined to save space. Moreover, it is necessary to insert the $(\Delta\nu)^2 = \frac{1}{L^2}$ bin width in converting between the discrete and continuous $\tilde{f}$ to ensure the units are correct. In the interest of clarity we will adopt a convention whereby all m, n incidies in $\vec{\nu}$ space and r, s indicies in $\vec{x}$.

The transform pair becomes

$$\tilde{g}_{mn} = \frac{1}{N^2} \sum_{r,s=-\frac{N}{2}}^{\frac{N}{2}-1} g_{rs} e^{-j\vec{k}_{mn}\cdot\vec{x}_{rs}} \tag{3.6}$$

$$g_{rs} = \sum_{m,n=-\frac{N}{2}}^{\frac{N}{2}-1} \tilde{g}_{mn} e^{j\vec{k}_{mn}\cdot\vec{x}_{rs}} \tag{3.7}$$

and circular convolution[1]

$$[g * h]_{rs} = \frac{1}{N^2} \sum_{r_1,s_1=-\frac{N}{2}}^{\frac{N}{2}-1} g_{r_1 s_1} h_{(r-r_1)(s-s_1)}. \tag{3.8}$$

3.2 Signal Processing Formalism

Below we establish the signal processing mathematics used to construct the source subtraction procedure.

Power Spectral Density

The physical noise $n(\vec{x})$ in a single frequency band as discussed in Section 2.2 is characterized by its PSD. The noise cannot be taken as simply Gaussian in each pixel with some fixed variance $\left\langle n(\vec{x})^2 \right\rangle$ (assuming zero mean) as such a description does not contain the correlation of $\nu(\vec{x})$ with itself in space. Physically, such correlations are meaningful: the propagation of a signal in space would encode information about $\nu(\vec{x}_0)$ in $\nu\left(\vec{x}_0 + \vec{\xi}\right)$. These correlations are instead described with the autocorrelation function

$$\begin{aligned} R(\vec{\xi}) &= \left\langle \nu(\vec{x}_1)\nu\left(\vec{x}_1 + \vec{\xi}\right) \right\rangle \\ &= \frac{1}{L^2} \iint \mathrm{d}^2\vec{x}\ \nu(\vec{x})\nu\left(\vec{x} + \vec{\xi}\right). \end{aligned} \tag{3.9}$$

We then define the noise PSD $J(\vec{\nu})$ the Fourier transform of the autocorrelation function (where $\vec{\nu}$ the spatial wavenumber is not to be confused with ν the observational

[1] i.e. we take the signal to be periodically repeating in both directions, or equivalently any indicies exceeding the length of the signal are taken modulo the signal length.

frequency band), given by

$$J(\vec{\nu}) = \iint \mathrm{d}^2\vec{x}\, R(\vec{x})\, e^{-j\vec{k}\cdot\vec{x}} \tag{3.10}$$

$$R(\vec{x}) = \iint \mathrm{d}^2\vec{\nu} J(\vec{\nu})\, e^{j\vec{k}\cdot\vec{x}}. \tag{3.11}$$

Recalling that $J(\vec{\nu}) = \dfrac{\left\langle |\tilde{n}(\vec{\nu})|^2 \right\rangle}{L^2}$ where $\tilde{n}(\vec{\nu})$ is the Fourier transform of $n(\vec{x})$, we can generate a single noise realisation $n(\vec{x})$ with PSD $J(\vec{\nu})$ by choosing a value of $\tilde{n}(\vec{\nu})$ from a normal distribution with width $L\sqrt{J(\vec{\nu})}$. In order to guarantee though that $n(\vec{x})$ is real (since $v(\vec{x})$ the observed signal is real), we require $\tilde{n}(\vec{\nu}) = \tilde{n}^*(-\vec{\nu})$.

Model Parameter Estimation

Recall from Section 2.2 that the data are modelled as $v(\vec{x}) = n(\vec{x}) + \sum_i A_i s(\vec{x}|\lambda_i)$. In order to define a goodness-of-fit χ^2 for some model $m_\nu(\vec{x}|\lambda)$ parameterized by some λ to the data, we propose

$$\chi^2 = \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{|\tilde{v}_\nu(\vec{\nu}) - \tilde{m}_\nu(\vec{\nu}|\lambda)|^2}{J_\nu(\vec{\nu})} \tag{3.12}$$

$$= \sum_\nu \sum_{\vec{\nu}} \frac{|\tilde{v}_\nu(\vec{\nu}) - \tilde{m}_\nu(\vec{\nu}|\lambda)|^2}{J_\nu(\vec{\nu})}. \tag{3.13}$$

We will generally use the integral notation because it is cleaner without numerical factors; the conversion between the two has already been laid out earlier in this section. We note that if $m_\nu(\vec{x}|\lambda) = \sum_i A_i s(\vec{x}|\lambda_i)$ exactly then

$$\chi^2 = \sum_\nu \iint \mathrm{d}^2\vec{\nu}\, \frac{|\tilde{n}_\nu(\vec{\nu})|^2}{J_\nu(\vec{\nu})}$$

$$= \sum_\nu \iint \mathrm{d}^2\vec{\nu}\, N^2,$$

and since $\mathrm{d}^2\vec{\nu}\, L^2 = 1$ we find that χ^2 is just the total number of frequency components.

Given this χ^2, the choice of parameters λ for some model m_ν is a χ^2-minimization process. For a given λ_k, the best-fit value $\hat{\lambda}_k$ satisfies

$$\left.\frac{\partial \chi^2}{\partial \lambda_k}\right|_{\lambda_k = \hat{\lambda}_k} = 0. \tag{3.14}$$

This condition can be used to compute the best-fit estimators of various λ_k in the ensuing models considered.

Estimation Uncertainty

The above expression gives the best-fit estimators for a parameter λ_k, but the uncertainties on these estimators can also be determined by the curvature of the χ^2 matrix about its minimum. We use the general statistical result $\Delta\chi^2 = 1$ along one dimension and $\Delta\chi^2 \approx 2.3$ along two dimensions correspond to 1σ deviations. Then, the χ^2 surface can be expanded quadratically about its minimum as

$$\chi^2(\vec{\lambda}) - \chi^2_{min} \approx \vec{\lambda}^T \frac{1}{2} \cdot \mathbf{H} \cdot \vec{\lambda} \tag{3.15}$$
$$\mathbf{H}_{ij} \equiv \frac{\partial^2 \chi^2}{\partial \lambda_i \partial \lambda_j}$$

where $\mathbf{H}$ is the Hessian matrix. We can identify $\mathbf{C} \equiv \left(\frac{1}{2}\mathbf{H}\right)^{-1}$ to be the *covariance matrix*, which quantifies the uncertainty on the parameterization of χ^2_{min}. The covariance matrix also enters in the multivariate probability distribution of the estimators

$$P(\hat{\lambda}_1, \ldots \hat{\lambda}_i, \ldots) = \exp\left[-\sum_{i,j} \lambda_i (C^{-1})_{ij} \lambda_j\right]$$

This is an accurate characterization so long as the χ^2 surface is well-approximated as a quadratic at its minimum, producing a normal multivariate probability distribution. For instance, for any λ_k that does not covary with any other parameters, we can marginalize over the other λ to recover the well-known result

$$\sigma^2_{\lambda_k} = [C_{kk}]^{-1} = \left[\frac{1}{2}\frac{\partial^2 \chi^2}{\partial \lambda_k^2}\bigg|_{\lambda_k = \hat{\lambda}_k}\right]^{-1}, \tag{3.16}$$

where we must evaluate λ_k at its optimal estimator $\hat{\lambda}_k$ to be at χ^2_{min}.

There exist parameters that do not obey such a description, those exhibiting non-quadratic behavior in all neighborhoods of χ^2_{min}, and these need to be handled on a case-by-case basis.

3.3 Simulation

The objective of the study is to verify that the source subtraction performs at the theoretical limit predicted by the covariance matrix elements. To this end, we perform simulations in which we generate many noise realizations $n_\nu(\vec{x})$. In each noise realization, we insert sources with known parameters λ_k and apply the

procedure to estimate $\hat{\lambda}_k$ for each parameter. The deviations of the $\hat{\lambda}_k$ from the inserted λ_k can then be plotted as a histogram. The distribution of $\hat{\lambda}_k$ is expected to be Gaussian with width σ_{λ_k} since the negative exponential of the χ^2 can be associated with a likelihood function.

We examine various parameters λ in different models and show that the distributions of their optimal estimators either agree with the prediction from the covariance matrix or deviate in well-understood ways.

Chapter 4

SOURCE SUBTRACTION: FORMALISM AND SIMULATION

In this chapter we lay out the procedure for point source subtraction and establish its agreement with analytical predictions derived from signal processing theory.

4.1 Single Frequency Band

We first examing source subtraction in a single frequency band. Only the accuracy of the formalism is of interest here, and so we place sources with unrealistic signal-to-noise ratios (SNR) to improve the approximation of the χ^2 surface as quadratic.

Single Centered Source

Consider if the spatial location of the source is known (e.g. centered on the $(0, 0)$ pixel) and we wish simply to find the flux of the source in the map. Then $s(\vec{x})$ can be chosen to be exactly the source profile including spatial offset and the data model becomes

$$v(\vec{x}) = As(\vec{x}) + n(\vec{x}) \tag{4.1}$$

with $s(\vec{x}) \propto e^{-\frac{|\vec{x}-\vec{x}_0|^2}{2\sigma^2}}$ the normalized, dimensionless beam signal fixed at some $\vec{x}_0$, A the desired flux we wish to recover and $n(\vec{x})$ some noise realization. $s(\vec{x})$ depends on no other parameters since it is a Gaussian with known width σ and offset $\vec{x}_0$. The χ^2 becomes then

$$\chi^2(A) = \iint \mathrm{d}^2\vec{\nu}\, \frac{|\tilde{v}(\vec{\nu}) - A\tilde{s}(\vec{\nu})|^2}{J(\vec{\nu})}. \tag{4.2}$$

The optimality condition for estimator $\hat{A}$ is $\left.\frac{\mathrm{d}\chi^2}{\mathrm{d}A}\right|_{A=\hat{A}} = 0$ and

$$0 = \iint \mathrm{d}^2\vec{\nu}\, \frac{2\hat{A}\,|\tilde{s}(\vec{\nu})|^2 - \tilde{s}^*(\vec{\nu})\tilde{v}(\vec{\nu}) - \tilde{s}(\vec{\nu})\tilde{v}^*(\vec{\nu})}{J(\vec{\nu})}.$$

We then recall that $s(\vec{x}), v(\vec{x})$ are real, so $\tilde{s}^*(\vec{\nu}) = \tilde{s}(-\vec{\nu}), \tilde{v}^*(\vec{\nu}) = \tilde{v}(-\vec{\nu})$. Since we are integrating over all $\mathrm{d}^2\vec{\nu}$, the two terms $\tilde{s}^*(\vec{\nu})\tilde{v}(\vec{\nu}) + \tilde{s}(\vec{\nu})\tilde{v}^*(\vec{\nu})$ make equal contributions (there is a correspondence for each contribution from the former term

at $\vec{\nu}$ and the latter term at $-\vec{\nu}$), so we can just consider twice one of these terms, and we obtain

$$\hat{A} = \frac{\iint \mathrm{d}^2\vec{\nu}\, \frac{\tilde{s}^*(\vec{\nu})\tilde{v}(\vec{\nu})}{J(\vec{\nu})}}{\iint \mathrm{d}^2\vec{\nu}\, \frac{|\tilde{s}(\vec{\nu})|^2}{J(\vec{\nu})}}. \tag{4.3}$$

The uncertainty σ_A in this estimator can be computed using the covariance-less result in (3.16) to obtain

$$\begin{aligned}\sigma_A^2 &= \left[\frac{1}{2}\frac{\mathrm{d}^2\chi^2}{\mathrm{d}A^2}\bigg|_{A=\hat{A}}\right]^{-1} \\ &= \left[\iint \mathrm{d}^2\vec{\nu}\frac{|\tilde{s}(\vec{\nu})|^2}{J(\vec{\nu})}\right]^{-1}.\end{aligned} \tag{4.4}$$

The predictions of this estimator and uncertainty can be examined by generating many noise realizations and adding a source of fixed flux A in the same location in every map. Within each map, we compute using (4.3) the best-fit estimator $\hat{A}$ and aggregate these estimators into a histogram that is expected to have width given by (4.4). The comparison of a histogram containing estimators from 10000 noise realizations to this analytic prediction can be found in Figure 4.1. Note the close agreement between the width of the Gaussian fit and the predicted width. The shape of the distribution is also very clearly Gaussian.

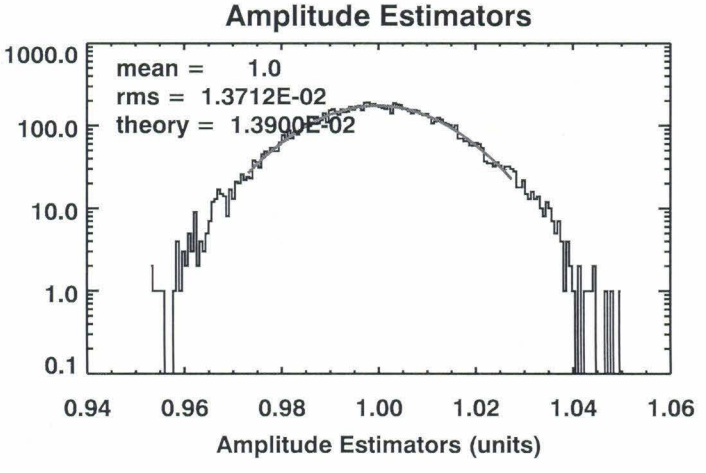


Figure 4.1: Histogram of flux estimators over 10000 noise realisations with a single source of unit flux inserted. mean/rms values from Gaussian fit and theory from (4.4).

Single off-center source

Consider now if we have a source profile with an *unknown* position offset $\vec{x}_0 = (x_0, y_0)$ such that $s(\vec{x}|\vec{x}_0) = s(\vec{x} - \vec{x}_0)$; while we restrict $\vec{x}$ to be on the pixelated grid, $\vec{x}_0$ is allowed to take on any continuous value. It is a well-known result that in the continuous Fourier transform $\mathcal{FT}[f(\vec{x} - \vec{x}_0)] = \tilde{f}(\vec{\nu})e^{-j\vec{k}\cdot\vec{x}_0}$ if $\tilde{f}(\vec{\nu})$ is the Fourier transform of $f(\vec{x})$. Sampling both sides of this result at discrete intervals demonstrates that the property holds for the discrete Fourier transform as well. The χ^2 then becomes

$$\chi^2(A, \vec{x}_0) = \iint \mathrm{d}^2\vec{\nu}\, \frac{\left|\tilde{v}(\vec{\nu}) - Ae^{-j\vec{k}\cdot\vec{x}_0}\tilde{s}(\vec{\nu})\right|^2}{J(\vec{\nu})}. \tag{4.5}$$

We can then repeat the above exercises to compute the optimality conditions on A, x_0, y_0 and their associated uncertainties. First, the optimal flux estimator $\hat{A}$ changes little

$$\hat{A} = \frac{\iint \mathrm{d}^2\vec{\nu}\, e^{j\vec{k}\cdot\vec{x}_0}\frac{\tilde{s}^*(\vec{\nu})\tilde{v}(\vec{\nu})}{J(\vec{\nu})}}{\iint \mathrm{d}^2\vec{\nu}\, \frac{\left|\tilde{s}(\vec{\nu})\right|^2}{J(\vec{\nu})}} \tag{4.6}$$

$$\sigma_A^2 = \left[\iint \mathrm{d}^2\vec{\nu}\frac{\left|\tilde{s}(\vec{\nu})\right|^2}{J(\vec{\nu})}\right]^{-1} \tag{4.7}$$

The optimal position estimators for x_0, y_0 follow the same expressions, so we will only compute the optimality condition for x_0 below

$$\left.\frac{\partial\chi^2}{\partial x_0}\right|_{x_0=\hat{x}_0} = 0 = -\frac{\partial}{\partial x_0}\iint \mathrm{d}^2\vec{\nu}\, \left.\frac{\tilde{v}(\vec{\nu})Ae^{j\vec{k}\cdot\vec{x}_0}\tilde{s}^*(\vec{\nu}) + \tilde{v}^*(\vec{\nu})Ae^{-j\vec{k}\cdot\vec{x}_0}\tilde{s}(\vec{\nu})}{J(\vec{\nu})}\right|_{x_0=\hat{x}_0} \tag{4.8}$$

$$= -2\frac{\partial}{\partial x_0}\iint \mathrm{d}^2\vec{\nu}\, \left.\frac{\tilde{v}(\vec{\nu})Ae^{j\vec{k}\cdot\vec{x}_0}\tilde{s}^*(\vec{\nu})}{J(\vec{\nu})}\right|_{x_0=\hat{x}_0}.$$

While there is no analytic form for $\hat{x}_0$, the integral is a convolution of the map $v(\vec{x})$ and the *optimal filter* $\phi(\vec{x})$ with Fourier transform

$$\tilde{\phi}(\vec{\nu}) = \frac{\tilde{s}(\vec{\nu})}{J(\vec{\nu})}. \tag{4.9}$$

The optimality condition for $\hat{x}_0$ can be cast in terms of this optimal filter as

$$0 = -2\frac{\partial}{\partial x_0}\left[v(\vec{x}) * \phi(\vec{x})\right](\vec{x}_0)\Big|_{x_0=\hat{x}_0}. \tag{4.10}$$

$\hat{x}_0$ is then the the maximum of the convolution of the map with the optimal filter. However, recall that $\vec{x}_0$ is a continuous parameter, while the convolved map is necessarily pixelated. This can be compensated by recalling that the convolution of a Gaussian with a Gaussian remains a Gaussian. Thus, the peak of the convolved map is approximately a Gaussian, completely Gaussian in the white noise limit where the optimal filter is a constant multiple of the source profile, and we can obtain the position of the maximum in the convolved map to sub-pixel accuracy by fitting the peak of the convolved map to a Gaussian.

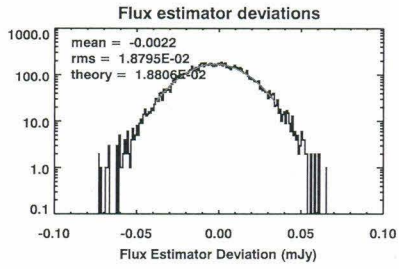
The theoretical limit on the accuracy of $\hat{x}_0$ can be determined first by noting that there is no covariance between $A, \vec{x}_0$ (i.e. $\dfrac{\partial^2 \chi^2}{\partial A \partial x_0} = 0$) and no average covariance between x_0, y_0, both of which can be verified by taking the partial derivatives. Thus, the uncertainty on $\hat{x}_0$ is given as

$$\begin{aligned}\sigma_{x_0}^2 &= \left[\frac{1}{2}\frac{\mathrm{d}^2\chi^2}{\mathrm{d}x_0^2}\bigg|_{x_0=\hat{x}_0}\right]^{-1} \\ &= \left[(2\pi)^2\hat{A}\iint \mathrm{d}^2\vec{\nu}\frac{\tilde{\nu}^*(\vec{\nu})e^{-j\vec{k}\cdot\vec{x}_0}\tilde{s}(\vec{\nu})}{J(\vec{\nu})}\nu_x^2\right]^{-1}.\end{aligned}$$

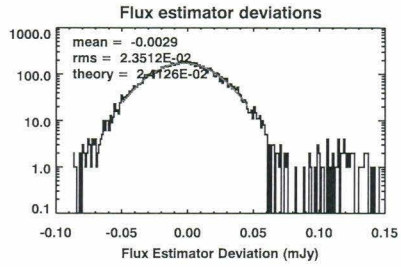
It is then a common prescription to replace $\tilde{\nu}^*(\vec{\nu})$ with its expectation $\langle\tilde{\nu}^*(\vec{\nu})\rangle = Ae^{j\vec{k}\cdot\vec{x}_0}\tilde{s}^*(\vec{\nu}) \approx \hat{A}e^{j\vec{k}\cdot\vec{x}_0}\tilde{s}^*(\vec{\nu})$. The motivation for this is that the convolution of the optimal filter with the noise is dominated by the convolution of the source with the optimal filter which is just a high-pass filter on the known source profile. Moreover, the uncertainty on an estimator should be noise realisation-independent and only depend on the noise properties. This substitution yields

$$\begin{aligned}\sigma_{x_0}^2 &= \left[(2\pi)^2\hat{A}^2\iint \mathrm{d}^2\vec{\nu}\frac{|\tilde{s}(\vec{\nu})|^2}{J(\vec{\nu})}\nu_x^2\right]^{-1} \\ &\approx \left[(2\pi)^2A^2\iint \mathrm{d}^2\vec{\nu}\frac{|\tilde{s}(\vec{\nu})|^2}{J(\vec{\nu})}\nu_x^2\right]^{-1}. \qquad (4.11)\end{aligned}$$

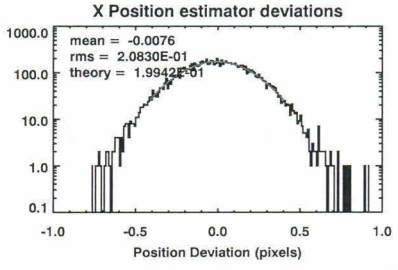
It is more sensible to compare the width of the distribution of $\hat{x}_0$ to the σ_{x_0} predicted using A instead of $\hat{A}$, since in testing the $\hat{A}$ will change from realisation to realisation while A is both constant and known. The predictions of both above analytic predictions with histogram widths is provided in Figure 4.2



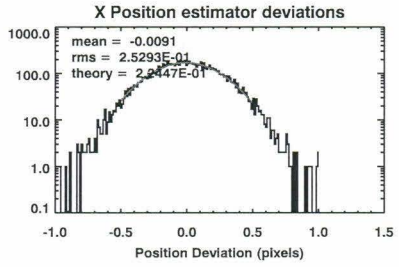
(a) Histogram of flux estimator deviations $(A - \hat{A})$ over 10000 white noise realizations with a single source of 0.2mJy flux inserted at SNR10. mean/rms from Gaussian fit and theory from (4.7).



(b) Histogram of flux estimator deviations over 10000 white + $\frac{1}{60}\text{arcsec}^{-1}$ elbow $\nu^{-8/3}$ noise realizations with a single source of 0.2mJy flux inserted at SNR10. mean/rms from Gaussian fit and theory from (4.7).



(c) Histogram of position estimator deviations over 10000 white noise realizations with a single source of 0.2mJy flux inserted at SNR10. mean/rms from Gaussian fit and theory from (4.11). Only $\hat{x}_0$ included since χ^2 is symmetric in x_0, y_0 dependence.



(d) Histogram of position estimator deviations over 10000 white + $\frac{1}{60}\text{arcsec}^{-1}$ elbow $\nu^{-8/3}$ noise realizations with a single source of 0.2mJy flux inserted at SNR10. mean/rms from Gaussian fit and theory from (4.7).

Figure 4.2: Position and flux estimator deviations over multiple trials and different PSDs using offset filtering.

Multiple Sources

The next step is to generalize the subtraction to multiple sources in a single map. The naive approach is to simply continually subtract out the brightest source in the map. However, this produces suboptimal flux estimation due to overlap between sources. Instead, we write down the χ^2 for M sources

$$\chi^2 = \iint \mathrm{d}^2\vec{\nu} \frac{\left|\tilde{\nu}(\vec{\nu}) - \sum_{i=1}^{M} e^{-j\vec{k}\cdot\vec{x}_{0,i}} A_i \tilde{s}_i(\vec{\nu})\right|^2}{J(\vec{\nu})} \tag{4.12}$$

with ith source centered at $\vec{x}_{0,i}$. Assume that the pixel locations of all have already been determined. Minimizing with respect to the A_i yields M coupled equations

indexed over i

$$\iint \mathrm{d}^2\vec{\nu}\,\frac{e^{j\vec{x}_i\cdot\vec{k}}\tilde{s}_i^*(\vec{\nu})\tilde{v}(\vec{\nu})}{J(\vec{\nu})} = \iint \mathrm{d}^2\vec{\nu}\,\frac{A_i\left|\tilde{s}_i(\vec{\nu})\right|^2 + \sum_{k\neq i}^{M} e^{-j(\vec{x}_k-\vec{x}_i)\cdot\vec{k}}A_k\left|\tilde{s}(\vec{\nu})\right|^2}{J(\vec{\nu})}$$

$$= A_i\iint \mathrm{d}^2\vec{\nu}\,\frac{\left|\tilde{s}_i(\vec{\nu})\right|^2}{J(\vec{\nu})} + \sum_{k\neq i}^{M} A_k\iint \mathrm{d}^2\vec{\nu}\,e^{-j(\vec{x}_k-\vec{x}_i)\cdot\vec{k}}\frac{\left|\tilde{s}(\vec{\nu})\right|^2}{J(\vec{\nu})}$$

$$= \sum_{k=1}^{M} A_k\iint \mathrm{d}^2\vec{\nu}\,e^{-j(\vec{x}_k-\vec{x}_i)\cdot\vec{k}}\frac{\left|\tilde{s}(\vec{\nu})\right|^2}{J(\vec{\nu})} \tag{4.13}$$

for the $\hat{A}_i$, and since the equations are all linear in $\hat{A}_i$ we can write down the optimality condition for these $\hat{A}_i$ as a matrix equation

$$\mathbf{M}\vec{A} = \vec{S}$$

$$M_{ik} = \iint \mathrm{d}^2\vec{\nu}\,e^{j(\vec{x}_i-\vec{x}_k)\cdot\vec{k}}\frac{\left|\tilde{s}(\vec{\nu})\right|^2}{J(\vec{\nu})}$$

$$S_i = \iint \mathrm{d}^2\vec{\nu}\,e^{j\vec{x}_i\cdot\vec{k}}\frac{\tilde{s}^*(\vec{\nu})\tilde{v}(\vec{\nu})}{J(\vec{\nu})} \tag{4.14}$$

that can be solved to *simultaneously* compute all flux estimators in a map.

The naive iterative algorithm described above is still used for position estimators, down to some detection limit, then the matrix solution applied to find the source fluxes at each of these position estimators. The detection limit can generally be set to be at ~ 3–5 SNR, and so we use a much higher SNR in the subtraction to guarantee all sources are identified. The resulting histogram is presented in Figure 4.3.

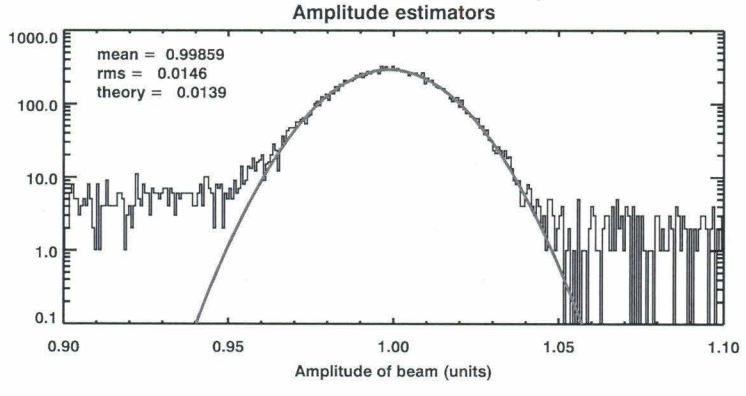


Figure 4.3: Flux Estimators with 20 sources in each of 1000 maps, each of unit flux 1, subtracted iteratively. White noise at unrealistic SNR.

The histogram in Figure 4.3 can be noted to have some systematic skew leftwards. This can be explained by the phenomenon of blended sources, whereby sources

that are too spatially close cannot be distinguished by an iterative subtraction, c.f. Figure 4.4. Blended sources are inevitable in data, but the error committed by misidentifying a blended source will generally fall under the confusion limit. A careful study of techniques used to study blended sources can be found in Appendix A.

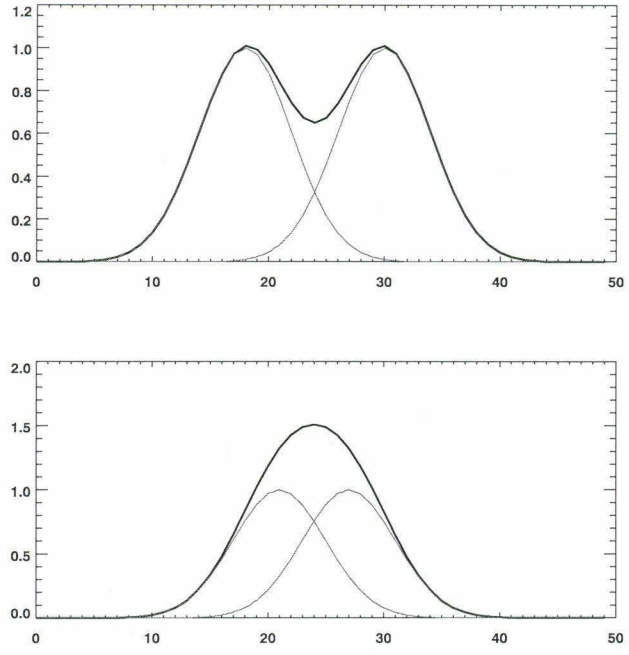


Figure 4.4: Example of blended sources. Cross section of fluxes of two point sources of unit flux (red) and their cumulative profile (black) with arbitrary units across the horizontal axis. Note that the lower plot is easily mistaken for a single source than two separate.

4.2 Multiple Frequency Band

We are now able to identify spatial location and flux of multiple point sources within a single frequency band. We next apply the formalism to sources in multiple frequency bands.

As discussed in the introduction, Section 2.1, we model our data to be collected in multiple spectral frequencies in overlapping spatial domain. Since any astronomical sources will have emissions in all spectral frequencies in the same spatial location, utilizing all maps to simultaneously identify sources results in more robust detection. For instance, detections of modest significance in individual frequencies may be highly significant if they come from the same spatial location.

In order to correlate information across multiple frequency bands, it is necessary to

assume some functional form for the SED. We use the greybody SED in the optically thin limit given in the introduction (1.5)

$$S(\nu|\beta,T) = \left(\frac{\nu}{\nu_0}\right)^{\beta} \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_B T}} - 1} \equiv A \frac{\left(\frac{\nu}{\nu_0}\right)^{3+\beta}}{e^{\frac{h\nu}{k_B T}} - 1} \tag{4.15}$$

where we absorb some numerical factors into A. Recall that typical ranges of $T_{dust} = (1+z)T$ range between 20–60K with $z \in [2.2, 2.7]$.

Moreover, as also discussed in Section 2.1, the point spread function of the instrument is assumed to be simply diffraction limited in different frequency bands, so the source profile $s_\nu(\vec{x})$ has different widths for different ν. Each $s_\nu(\vec{x})$ is assumed to be normalized to unit integral. Then the χ^2 for a single source can be written

$$\chi^2(A, \vec{x}_0, \beta, T) = \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{\left| \tilde{v}_\nu(\vec{\nu}) - A \frac{\left(\frac{\nu}{\nu_0}\right)^{3+\beta}}{e^{\frac{h\nu}{k_B T}} - 1} e^{-j\vec{k}\cdot\vec{x}_0} \tilde{s}_\nu(\vec{\nu}) \right|^2}{J_\nu(\vec{\nu})} \tag{4.16}$$

which can be verified to be normalized to the total number of frequency components across all spectral frequencies.

We will first consider various simplified schemes to verify the generalisation of our formalism to multiple frequency bands.

Fixed SED

Consider the simplification where in (1.5) both β, T of the source are known and held fixed. Then the χ^2 function in (4.29) becomes only a function of $A, \vec{x}_0$ again, as the grey body dependence can be absorbed as a normalization on the $s_\nu(\vec{x})$ (not on A which is frequency-independent). Define

$$\mathfrak{s}_\nu(\vec{x}) = \frac{\left(\frac{\nu}{\nu_0}\right)^{3+\beta}}{e^{\frac{h\nu}{k_B T}} - 1} s_\nu(\vec{x}), \tag{4.17}$$

then we obtain for a single source

$$\chi^2(A, \vec{x}_0) = \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{\left| \tilde{v}_\nu(\vec{\nu}) - A e^{-j\vec{k}\cdot\vec{x}_0} \tilde{\mathfrak{s}}_\nu(\vec{\nu}) \right|^2}{J_\nu(\vec{\nu})}. \tag{4.18}$$

The optimality condition for $\hat{A}$ follows from differentiation

$$\sum_\nu \iint \mathrm{d}^2\vec{\nu}\, \frac{e^{j\vec{x}_i\cdot\vec{k}} \tilde{\mathfrak{s}}^*_\nu(\vec{\nu}) \tilde{v}_\nu(\vec{\nu})}{J_\nu(\vec{\nu})} = A \sum_\nu \iint \mathrm{d}^2\vec{\nu}\, \frac{|\tilde{\mathfrak{s}}_\nu(\vec{\nu})|^2}{J_\nu(\vec{\nu})}. \tag{4.19}$$

To compute the uncertainty on this estimator, we once again identify that $\frac{\partial^2 \chi^2}{\partial A \partial x_0} = \frac{\partial^2 \chi^2}{\partial y_0 \partial x_0} = 0$, so the uncertainties are given simply by the second derivatives of the χ^2 with respect to A. This yields

$$\sigma_A^2 = \left[\sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{|\tilde{\mathfrak{s}}_\nu(\vec{\nu})|^2}{J_\nu(\vec{\nu})} \right]^{-1} \tag{4.20}$$

Both formulae can be observed to closely parallel their single-band counterparts.

Offset Position Estimation

The position estmator $\hat{x}_0$ satisfies similar optimality criterion to the single frequency case as well:

$$\left.\frac{\partial \chi^2}{\partial x_0}\right|_{x_0=\hat{x}_0} = 0 = \sum_\nu \left\{ -2 \frac{\partial}{\partial x_0} \iint \mathrm{d}^2\vec{\nu} \left. \frac{\tilde{\nu}_\nu(\vec{\nu}) A e^{j\vec{k}\cdot\vec{x}_0} \tilde{\mathfrak{s}}^*(\vec{\nu})}{J(\vec{\nu})} \right|_{x_0=\hat{x}_0} \right\}$$

$$\tilde{\phi}_\nu(\vec{\nu}) = \frac{\tilde{\mathfrak{s}}_\nu(\vec{\nu})}{J_\nu(\vec{\nu})} \tag{4.21}$$

$$\left.\frac{\partial \chi^2}{\partial x_0}\right|_{x_0=\hat{x}_0} = 0 = \frac{\partial}{\partial x_0} \left\{ \sum_\nu \left[v_\nu(\vec{x}) * \phi_\nu(\vec{x}) \right] (\vec{x}_0) \Big|_{x_0=\hat{x}_0} \right\} \tag{4.22}$$

where we analogously define a multi-band optimal filter $\phi_\nu(\vec{x})$. With no off-diagonal terms in the covariance matrix, this has minimum uncertainty

$$\begin{aligned} \sigma_{x_0}^2 &= \left[(2\pi)^2 A^2 \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{|\tilde{\mathfrak{s}}_\nu(\vec{\nu})|^2}{J_\nu(\vec{\nu})} \nu_x^2 \right]^{-1} \\ &= \left[\sum_\nu \frac{1}{\sigma_{x_0,\nu}^2} \right]^{-1} \end{aligned} \tag{4.23}$$

which relates the position estimator uncertainty in each individual band $\sigma_{x_0,\nu}^2$ to the overall position estimator uncertainty σ_{x_0}.

Recall that in the single-band case, we were able to compute $\hat{x}_0$ to sub-pixel resolution by fitting the convolution $v(\vec{x}) * \phi(\vec{x})$ to a Gaussian near the peak. In the present case, $\hat{x}_0$ is the argmax of the sum of the convolutions of the $v_\nu(\vec{x}) * \phi_\nu(\vec{x})$ in each frequency band ν. However, each convolution is peaked at a different position, the $\hat{x}_0$ in that band, and the sum of these six Gaussians is no longer Gaussian about its peak. Thus we cannot use a Gaussian fit to extrapolate $\hat{x}_0$ to subpixel resolution.

Instead, an estimator with optimal uncertainty can be computed as a weighted average of the optimal estimators in each individual frequency band.

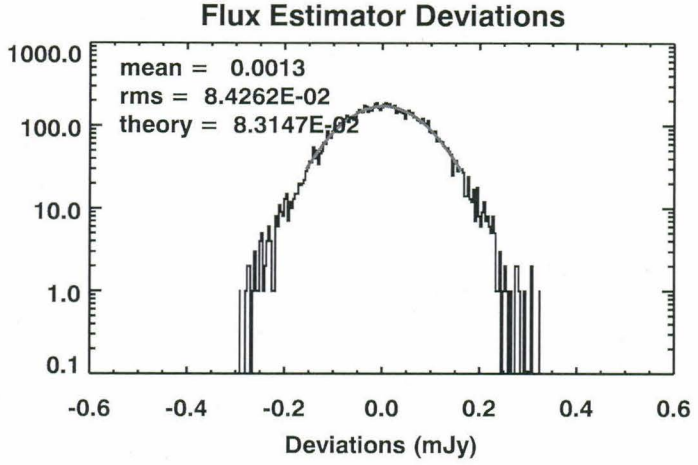
$$\frac{1}{\sigma^2_{x_0}}\hat{x}_0 \equiv \sum_\nu \frac{1}{\sigma^2_{x_0,\nu}}\hat{x}_{0,\nu} \tag{4.24}$$

We demonstrate this has the correct variance

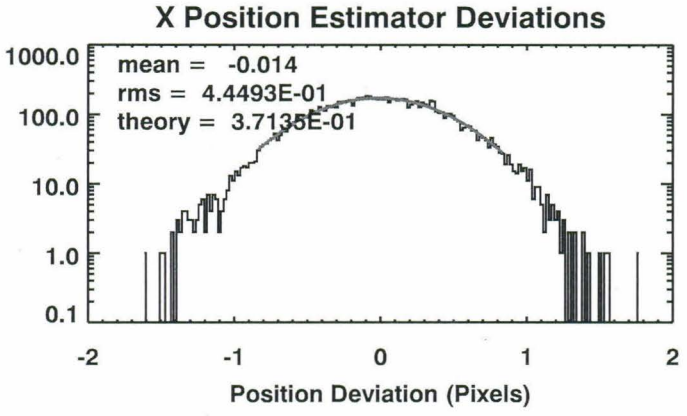
$$\begin{aligned}
\mathrm{Var}\left(\hat{x}_0\right) &= \mathrm{Var}\left(\sigma^2_{x_0}\left(\sum_\nu \frac{1}{\sigma^2_{x_0,\nu}}\hat{x}_{0,\nu}\right)\right) \\
&= \sigma^4_{x_0}\left(\sum_\nu \mathrm{Var}\left(\frac{1}{\sigma^2_{x_0,\nu}}\hat{x}_{0,\nu}\right)\right) \\
&= \sigma^4_{x_0}\sum_\nu\left(\frac{1}{\sigma^4_{x_0,\nu}}\mathrm{Var}\left(\hat{x}_{0,\nu}\right)\right) \\
&= \sigma^4_{x_0}\sum_\nu\left(\frac{1}{\sigma^2_{x_0,\nu}}\right) \\
&= \sigma^2_{x_0}
\end{aligned} \tag{4.25}$$

and observe that $\langle\hat{x}_0\rangle = \left\langle\sigma^2_{x_0}\left(\sum_\nu \frac{1}{\sigma^2_{x_0,\nu}}\hat{x}_{0,\nu}\right)\right\rangle = \sigma^2_{x_0}\left(\sum_\nu \frac{1}{\sigma^2_{x_0,\nu}}\langle\hat{x}_{0,\nu}\rangle\right) = x_0$.

The distribution of the estimators over many noise realisations can be observed in Figure 4.5 to agree well with the analytic prediction.



(a) Histogram of flux estimator deviations. mean/rms from Gaussian fit and theory from (4.20).



(b) Histogram of position estimator deviations. mean/rms from Gaussian fit and theory from (4.23).

Figure 4.5: Histograms of multi-band flux and position estimators with noise parameters described in Section 2.1 and a single source of 1.81mJy flux in 750μm band inserted corresponding to SNR 20 over 10000 trials.

This procedure is liable to fail when the SNR is sufficiently low that the optimal position estimator in certain bands identifies a statistical fluctuation rather than the source, as the weighted average will then include estimations made from non-source fluxes. One way to decrease the probability of this occuring is to seek the maximum-flux estimators in each band within a small pixel range of the maximum pixel in the convolved map (e.g. $3\sigma_\nu$ the beam width in each frequency band). Considering fewer pixels decreases the probability of finding a signal-like statistical fluctuation but does not identify when a failure occurs.

To identify the occurrence of such a failure, we associate a χ^2 statistic with the position estimators in each band. Any bands with position estimators inconsistent with the hypothesis that the same source is fitted in each map can then be ignored from the weighted average in finding the position estimator. All frequency bands can still be used to compute the flux of the source, however. We require both procedures described here for best performance as discarding frequency bands from position estimation degrades $\hat{x}_0$ in a noise realisation-dependent way, so first restricting fitting to be within a small range of the convolved maximum assures that this happens as infrequently as possible.

Multiple Sources

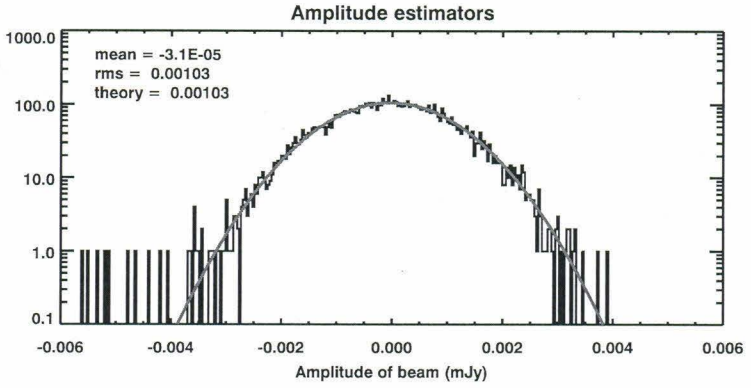
The extension to multiple sources follows a similar procedure to the single-band procedure. The χ^2 becomes

$$\chi^2(A, \vec{x}_0) = \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{\left| \tilde{\nu}_\nu(\vec{\nu}) - \sum_{i=1}^{M} A_i e^{-j\vec{k}\cdot\vec{x}_{0,i}} \tilde{\mathfrak{s}}_{\nu,i}(\vec{\nu}) \right|^2}{J_\nu(\vec{\nu})}. \tag{4.26}$$

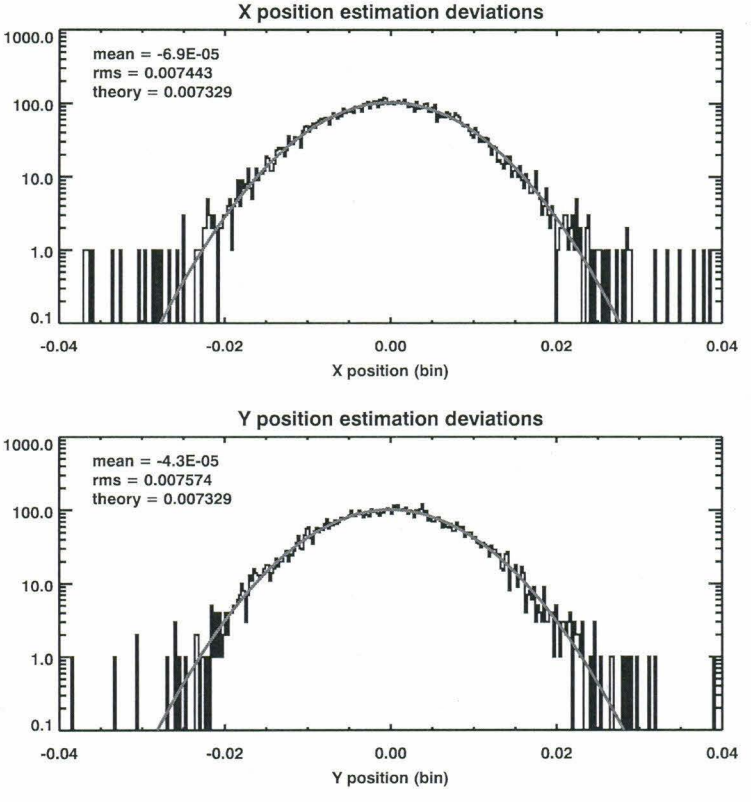
where we account for the possibility that each source has a different (but known) SED by allowing $\mathfrak{s}_{\nu,i}$ to differ between sources. Then we can apply the same procedure as in the single band case, iteratively subtracting the brightest source for source positions down to some detection threshold and then writing down M coupled equations for M source fluxes

$$\sum_\nu \left\{ \iint \mathrm{d}^2\vec{\nu} \, \frac{e^{j\vec{x}_i\cdot\vec{k}} \tilde{\mathfrak{s}}^*_{\nu,i}(\vec{\nu}) \tilde{\nu}_\nu(\vec{\nu})}{J_\nu(\vec{\nu})} \right\} = \sum_\nu \left\{ \sum_{k=1}^{M} A_k \iint \mathrm{d}^2\vec{\nu} \, e^{-2\pi j(\vec{x}_k - \vec{x}_i)\cdot\vec{\nu}} \frac{\tilde{\mathfrak{s}}_{\nu,i}(\vec{\nu}) \tilde{\mathfrak{s}}_{\nu,k}(\vec{\nu})}{J_\nu(\vec{\nu})} \right\} \tag{4.27}$$

which can be solved using the same $\mathbf{M}\vec{A} = \vec{S}$ prescription as for the single band (4.14). This subtraction is demonstrably still optimal as predicted by (4.20) and (4.23) as can be seen in Figure 4.6.



(a) Histogram of flux estimator deviations $A - \hat{A}$. mean/rms calculated from Gaussian fit and theory from (4.20).



(b) Histogram of position estimator deviations $x_0 - \hat{x}_0$, mean/rms calculated from Gaussian fit and theory from (4.23).

Figure 4.6: Estimator histograms for 1000 noise realisations with 8 randomly inserted sources with flux 0.6mJy in 750μm in each. Unrealistic SNR $\sim\sim$ 200 in leading band used to avoid cutoff by detection threshold.

4.3 Rayleigh-Jeans SED

We next examine the case where the SED satsifies the Rayleigh-Jeans approximation $k_B T \gg h\nu$. Note that a frequency of $\nu = 400\text{GHz}$ corresponds to a temperature of $T = 19.2\text{K}$.

The greybody SED from (1.5) becomes instead

$$\begin{aligned} S(\nu|\beta) &= \frac{2h\left(\frac{\nu}{\nu_0}\right)^{3+\beta}}{c^2} \frac{1}{e^{\frac{h\nu}{k_B T}} - 1} \\ &\approx \frac{2h\left(\frac{\nu}{\nu_0}\right)^{3+\beta}}{c^2} \frac{k_B T}{h\nu} \equiv A\left(\frac{\nu}{\nu_0}\right)^{2+\beta} \end{aligned} \tag{4.28}$$

where we absorb some numerical factors into A the normalization. Then the χ^2 for a single source becomes

$$\chi^2(A, \vec{x}_0, \beta) = \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{\left|\tilde{\nu}_\nu(\vec{\nu}) - A\left(\frac{\nu}{\nu_0}\right)^{2+\hat{\beta}} e^{-j\vec{k}\cdot\vec{x}_0} \tilde{s}_\nu(\vec{\nu})\right|^2}{J_\nu(\vec{\nu})} \tag{4.29}$$

The optimality conditions for x_0, A are the same as before if we simply carry the $S(\nu|\beta)$ factor with the source profile $\tilde{s}_\nu(\vec{\nu})$. The optimality condition on estimator $\hat{\beta}$ on the other hand is

$$\left.\frac{\partial \chi^2}{\partial \beta}\right|_{\beta=\hat{\beta}} = 0 = \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{2\left[A\left(\frac{\nu}{\nu_0}\right)^{2+\hat{\beta}} \tilde{s}^*_\nu(\vec{\nu}) - \tilde{\nu}^*_\nu(\vec{\nu})\right]\left[A\left(\frac{\nu}{\nu_0}\right)^{2+\hat{\beta}} \tilde{s}_\nu(\vec{\nu})\right]}{J_\nu(\vec{\nu})} \ln\left(\frac{\nu}{\nu_0}\right) \tag{4.30}$$

It is impossible to solve analytically for estimator $\hat{\beta}$ to obtain a closed form as we did for $\hat{A}$ nor to cast the condition in illuminating form as we did for $\hat{x}_0$. Instead, we must use gradient descent to numerically seek the χ^2 minimum at which (4.30) is satisfied. We use the built in Truncated-Newton minimizer `tnmin` in IDL[1] to perform the minimization. At each step of the gradient descent, we re-compute the optimal $\hat{x}_0, \hat{A}$ at the current value of β. We use $\frac{\partial \chi^2}{\partial \beta}$ to compute each time step however, which assumes fixed $\hat{x}_0, \hat{A}$; this empirically has little effect on convergence.

In order to examine the uncertainties of all these estimators, we must look at the full Hessian, which is no longer diagonal like previous applications of the formalism.

[1] Data Analysis done using IDL 8.1 (Exelis Visual Information Solutions, Boulder, Colorado).

Since χ^2 is symmetric with respect to x_0, y_0 parameters, we need only compute 6 second partial derivatives, reproduced below

$$\frac{1}{2}\frac{\partial^2\chi^2}{\partial A^2} = \sum_\nu \iint \mathrm{d}^2\vec{\nu}\frac{\left(\frac{\nu}{\nu_0}\right)^{4+2\beta}\left|\tilde{s}_\nu(\vec{\nu})\right|^2}{J_\nu(\vec{\nu})} \tag{4.31}$$

$$\frac{1}{2}\frac{\partial^2\chi^2}{\partial x_0^2} = \frac{\partial^2\chi^2}{\partial y_0^2} = (2\pi)^2A^2\sum_\nu \iint \mathrm{d}^2\vec{\nu}\frac{\left(\frac{\nu}{\nu_0}\right)^{4+2\beta}\left|\tilde{s}_\nu(\vec{\nu})\right|^2}{J_\nu(\vec{\nu})}\nu_x^2 \tag{4.32}$$

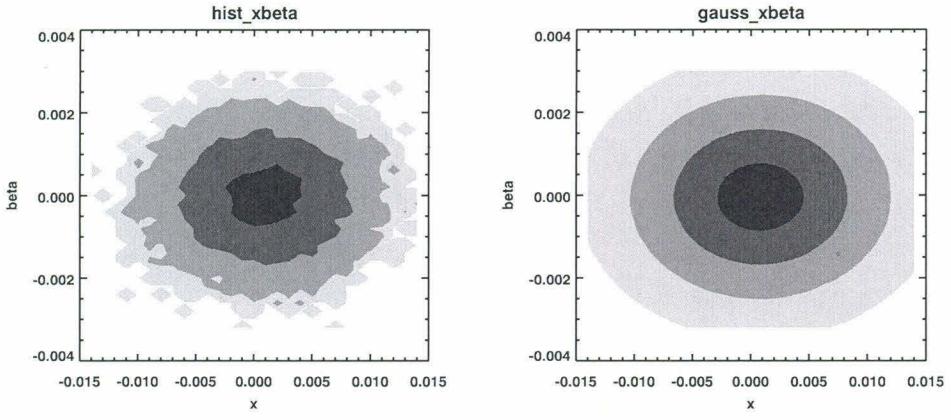
$$\frac{1}{2}\frac{\partial^2\chi^2}{\partial\beta^2} = \sum_\nu \iint \mathrm{d}^2\vec{\nu}\frac{A^2\left(\frac{\nu}{\nu_0}\right)^{4+2\beta}\left|\tilde{s}_\nu(\vec{\nu})\right|^2}{J_\nu(\vec{\nu})}\ln^2\left(\frac{\nu}{\nu_0}\right) \tag{4.33}$$

$$\frac{\partial^2\chi^2}{\partial x_0\,\partial A} = 0$$

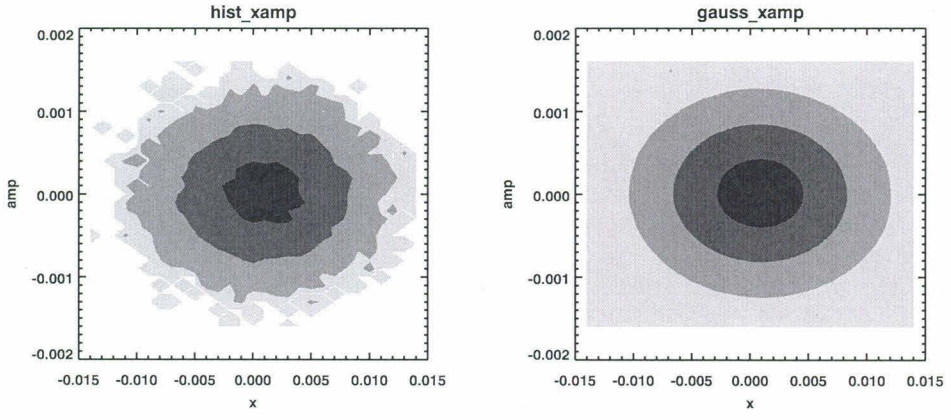
$$\frac{\partial^2\chi^2}{\partial\beta\,\partial A} = \sum_\nu \iint \mathrm{d}^2\vec{\nu}\frac{A\left(\frac{\nu}{\nu_0}\right)^{4+2\beta}\left|\tilde{s}_\nu(\vec{\nu})\right|^2}{J_\nu}\ln\left(\frac{\nu}{\nu_0}\right) \tag{4.34}$$

$$\frac{\partial^2\chi^2}{\partial x_0\,\partial\beta} = 0$$

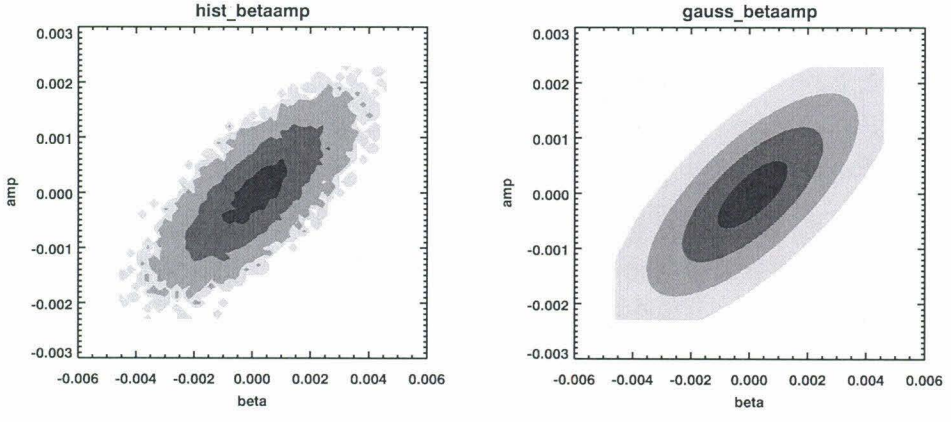
where we go ahead and substitute the expectation values for all noise-dependent terms. The observation is that there is a covariance, equivalently a degeneracy, between the β, A parameters in the χ^2 surface. With covariances the uncertainties on the $\hat{\beta}, \hat{A}$ estimators are no longer given by $\left[\frac{1}{2}\frac{\partial^2\chi^2}{\partial\beta^2}\right]^{-1}, \left[\frac{1}{2}\frac{\partial^2\chi^2}{\partial A^2}\right]^{-1}$ respectively. In order to characterize degenerate parameters, we must consider multivariate histograms to capture any covariances. If the multivariate distribution over many noise realisations conforms to the prediction made by the formalism then we can be assured our subtraction is optimal. Since the computed Hessian derivatives assume all parameters but two are held fixed (rather than being allowed to seek their optimal values freely), the proper method to generate each two-dimensional histogram is to fix all estimators but two. We generate one histogram for each of the two-dimensional subspaces of the Hessian computed above, in Figure 4.7.



(a) $\hat{x}_0, \hat{\beta}$ **(b)** Gaussian fit to $\hat{x}_0, \hat{\beta}$



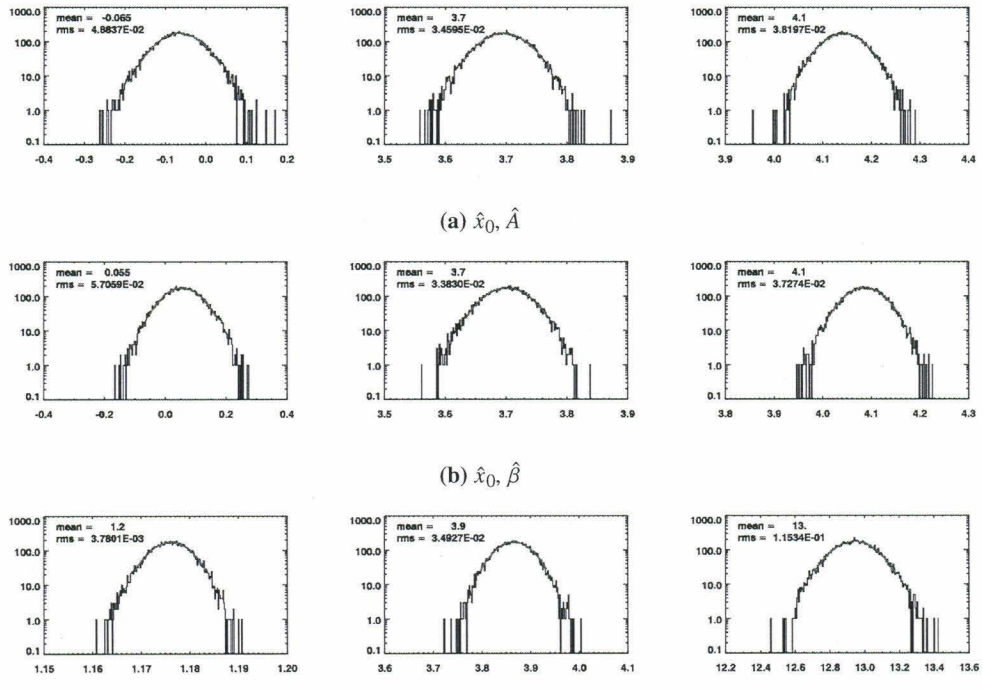
(c) $\hat{x}_0, \hat{A}$ **(d)** Gaussian fit to $\hat{x}_0, \hat{A}$



(e) $\hat{\beta}, \hat{A}$ **(f)** Gaussian fit to $\hat{\beta}, \hat{A}$

Figure 4.7: Bivariate histograms of estimator deviations $\lambda_k - \hat{\lambda}_k$ in Rayleigh-Jeans fitting with other variables held fixed. Over 20000 noise realizations. Both histograms and Gaussian fits are colored corresponding to regions of $1, 2, 3\sigma$ deviations labelled with darker colors corresponding to smaller deviations. Unrealistic SNR $\sim$ 100 used. No covariances for $\hat{\beta}, \hat{x}_0$ and $\hat{\beta}, \hat{A}$ reported since they are consistent with zero.

To compare the bivariate Gaussian histogram fits to the formalism, it is most illuminating to assign uncertainties to the fit parameters. Uncertainties for the histograms can be computed via Monte Carlo simulation by fluctuating each bin with N counts by Gaussian with standard deviation $\sqrt{N}$ (the usual Poissonian assumption and Gaussian approximation) and refitting. Such a bootstrapping technique estimates the uncertainty on the fit parameters, and we can see that the fits are nicely Gaussian as in Figure 4.8.



(a) $\hat{x}_0, \hat{A}$

(b) $\hat{x}_0, \hat{\beta}$

(c) $\hat{\beta}, \hat{A}$

Figure 4.8: Histograms of bootstrapped bivariate Gaussian fit parameters in order $\theta, \sigma_1, \sigma_2$ where σ_1, σ_2 refer to the widths along the principal axes of the respective histograms in Figure 4.7 and θ the orientation.

On the other hand, the analytical parameters can also be assigned an uncertainty by generating χ^2 surfaces for many noise realisations and performing a quadratic fit to the minimum of the surface. It should be noted that over the fitted interval, the fitted function generally differed from the computed, slightly non-quadratic χ^2 surface by $\lesssim 5\%$, so a similar systematic uncertainty must be associated with these fit parameters. The comparison of the covariance matrix terms computed via the analytical estimate, the multiple χ^2 fits and the bootstrapped histogram fits can is found in Table 4.1.

Parameters	σ_1	σ_2
$\hat{\beta}, \hat{A}$ (histogram)	$(12.10 \pm 0.10) \times 10^{-4}$	$(6.128 \pm 0.049) \times 10^{-4}$
$\hat{\beta}, \hat{A}$ (analytic)	$(11.879) \times 10^{-4}$	$(6.0875) \times 10^{-4}$
$\hat{\beta}, \hat{A}$ (χ^2)	$(11.82 \pm 0.11) \times 10^{-4}$	$(6.087 \pm 0.046) \times 10^{-4}$
$\hat{A}, \hat{x}_0$ (histogram)	$(4.111 \pm 0.036) \times 10^{-4}$	$(3.772 \pm 0.034) \times 10^{-3}$
$\hat{A}, \hat{x}_0$ (analytic)	$(4.137) \times 10^{-4}$	$(3.691) \times 10^{-3}$
$\hat{A}, \hat{x}_0$ (χ^2)	$(4.146 \pm 6 \times 10^{-5}) \times 10^{-4}$	$(3.698 \pm 0.023) \times 10^{-3}$
$\hat{\beta}, \hat{x}_0$ (histogram)	$(8.171 \pm 0.075) \times 10^{-4}$	$(3.711 \pm 0.034) \times 10^{-3}$
$\hat{\beta}, \hat{x}_0$ (analytic)	$(8.074) \times 10^{-4}$	$(3.691) \times 10^{-3}$
$\hat{\beta}, \hat{x}_0$ (χ^2)	$(8.043 \pm 0.046) \times 10^{-4}$	$(3.696 \pm 0.023) \times 10^{-3}$
Covar($\hat{\beta}, \hat{A}$) (histogram)	$(5.5 \pm 0.1) \times 10^{-7}$	
Covar($\hat{\beta}, \hat{A}$) (analytic)	$(5.31 \pm 0.002) \times 10^{-7}$	
Covar($\hat{\beta}, \hat{A}$) (χ^2)	$5.28 \pm 0.002 \times 10^{-7}$	

Table 4.1: Covariance matrix entries using bootstrapped histogram, analytic and multiple χ^2 fits. σ_1, σ_2 refer to the uncertainties of the first, second parameters in the left-most column respectively. Units on $\hat{A}$ are mJy, on $\hat{x}_0$ pixels and on $\hat{\beta}$ unitless. Unrealistic noise parameters, SNR 100 in leading band.

Note that including the uncertainty on the covariances establishes that the three estimates generally agree, particularly accounting for the χ^2 systematic misfitting, and thus we can be assured that the procedure satisfactorily attains the theoretical accuracy bound predicted by the formalism.

4.4 Greybody SED

At the beginning of Section 4.3, we found that for $T \gg 19\text{K}$ we can make the Rayleigh-Jeans approximation. In reality, the normal range of dust temperatures in DSFGs is in the range 20–60K, implying that the redshifted T for many sources lies outside the Rayleigh-Jeans limit (Casey, 2012). We thus return to the full greybody SED (1.5) reproduced below

$$S(\nu|\beta, T) = \frac{\left(\frac{\nu}{\nu_0}\right)^{3+\beta}}{e^{\frac{h\nu}{k_B T}} - 1} \tag{4.35}$$

and χ^2 also below

$$\chi^2(A, \vec{x}_0, \beta, T) = \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{\left| \tilde{v}_\nu(\vec{\nu}) - A \frac{\left(\frac{\nu}{\nu_0}\right)^{3+\beta}}{e^{\frac{h\nu}{k_B T}} - 1} e^{-j\vec{k}\cdot\vec{x}_0} \tilde{s}_\nu(\vec{\nu}) \right|^2}{J_\nu(\vec{\nu})}. \tag{4.36}$$

However, this parameterization runs into difficulties. In particular, since $e^{\frac{h\nu}{k_B T}} \sim 1$, changes in T still produce somewhat linear changes in flux, making T heavily degenerate with A the flux parameter, confounding the gradient descent. An example of the observed degeneracies can be seen in the cross section along the $A-T$ plane of the χ^2 surface about its minimum in Figure 4.9.

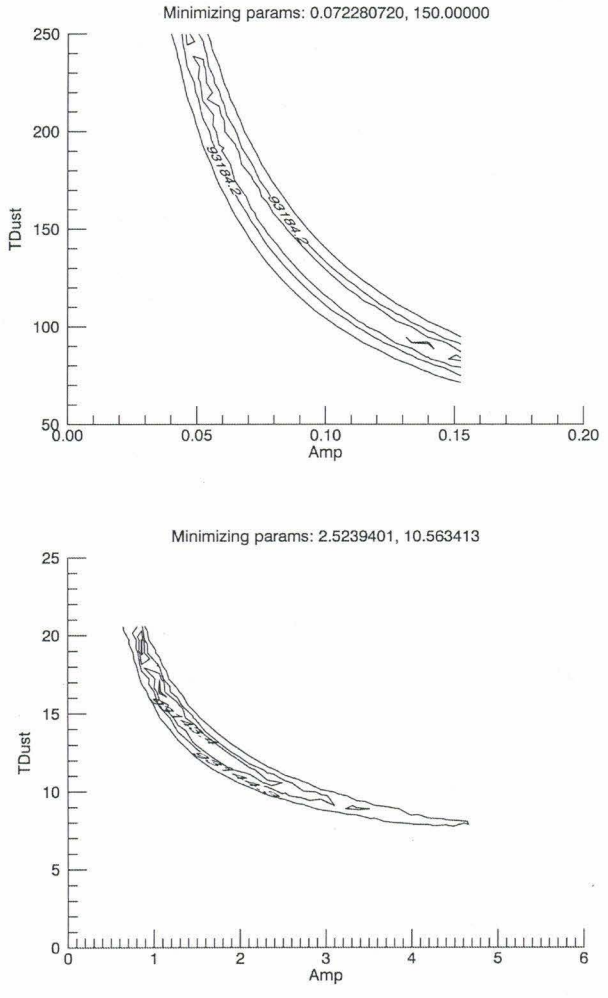


Figure 4.9: Inserted source with $A = 1.74, T = 17$ for two realisations of noise. The drastically different $\hat{A}, \hat{T}$ suggest terrible degeneracies that are consistent with the shape of the χ^2 surface observed.

To better constrain the problem, we instead define

$$S(\nu|\beta,T) \rightarrow \frac{S(\nu|\beta,T)}{S(\nu_{piv}|\beta,T)} \tag{4.37}$$

where we normalize the SED by some pivot frequency ν_{piv} that is held constant throughout the minimization. Then A takes on the interpretation of *relative* flux to the flux in frequency band ν_{piv}, and the χ^2 formula becomes

$$\chi^2(A, \vec{x}_0, \beta, T) = \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{\left| \tilde{v}_\nu(\vec{\nu}) - A\left(\frac{\nu}{\nu_{piv}}\right)^{3+\beta} \frac{e^{\frac{h\nu_{piv}}{k_B T}} - 1}{e^{\frac{h\nu}{k_B T}} - 1} e^{-j\vec{k}\cdot\vec{x}_0} \tilde{s}_\nu(\vec{\nu}) \right|^2}{J_\nu(\vec{\nu})}. \quad (4.38)$$

The optimality conditions on $\hat{A}, \hat{x}_0$ can once again be written down; we will not do so here since it follows the exact same procedure as before. The optimality conditions on β, T once again require gradient descent search via `tnmin`. Since the conditions are a bit ugly to write down, define $S(\nu|\nu_{piv}, \vec{x}_0, \beta, T) = \left(\frac{\nu}{\nu_{piv}}\right)^{3+\beta} \frac{e^{\frac{h\nu_{piv}}{k_B T}} - 1}{e^{\frac{h\nu}{k_B T}} - 1} e^{-j\vec{k}\cdot\vec{x}_0}$, then the conditions become

$$\begin{aligned} 0 &= \left.\frac{\partial \chi^2}{\partial \beta}\right|_{\beta=\hat{\beta}} \\ &= \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{2\left[AS(\nu)\tilde{s}^*_\nu(\vec{\nu}) - \tilde{v}^*_\nu(\vec{\nu})\right]\left[AS(\nu)\tilde{s}_\nu(\vec{\nu})\right]}{J_\nu(\vec{\nu})} \ln\left(\frac{\nu}{\nu_0}\right) \quad (4.39) \\ 0 &= \left.\frac{\partial \chi^2}{\partial T}\right|_{T=\hat{T}} \\ &= \sum_\nu \iint \mathrm{d}^2\vec{\nu} \frac{2\left[AS(\nu)\tilde{s}^*_\nu(\vec{\nu}) - \tilde{v}^*_\nu(\vec{\nu})\right]\left[A\left(\frac{\nu}{\nu_{piv}}\right)^{3+\beta} \frac{\left(e^{\frac{h\nu}{k_B\hat{T}}}-1\right)\left(-\frac{h\nu_{piv}}{k_B\hat{T}^2} e^{\frac{h\nu_{piv}}{k_B\hat{T}}}\right) - \left(e^{\frac{h\nu_{piv}}{k_B\hat{T}}}-1\right)\left(-\frac{h\nu}{k_B\hat{T}^2} e^{\frac{h\nu}{k_B\hat{T}}}\right)}{\left(e^{\frac{h\nu}{k_B\hat{T}}}-1\right)^2} \tilde{s}_\nu(\vec{\nu})\right]}{J_\nu(\vec{\nu})} \quad (4.40) \end{aligned}$$

It should be noted that the $(\hat{A}, \hat{x}_0, \hat{y}_0, \hat{\beta}, \hat{T})$ that `tnmin` finds is independent of ν_{piv}; for every set of parameters (A, x_0, y_0, β, T) found at one ν_{piv}, a corresponding set $\left(A\frac{S(\nu|\nu'_{piv}, \beta, T)}{S(\nu|\nu_{piv}, \beta, T)}, x_0, y_0, \beta, T\right)$ exists at a different pivot ν'_{piv} yielding the same χ^2 value. In fact, the χ^2 surfaces are exactly identical between choices of ν_{piv} up to a scaling factor on A,

The χ^2 surface even with the pivot frequency scaling exhibits intractable non-quadratic degeneracies in β, T. For instance, at SNR 20, the shape of the χ^2 surface looks like Figure 4.10. Recalling that the association of the covariance matrix with estimator uncertainties assumes a quadratic χ^2 surface, this non-quadratic surface

defies characterization using the developed formalism, and it is futile to compute the Hessian.

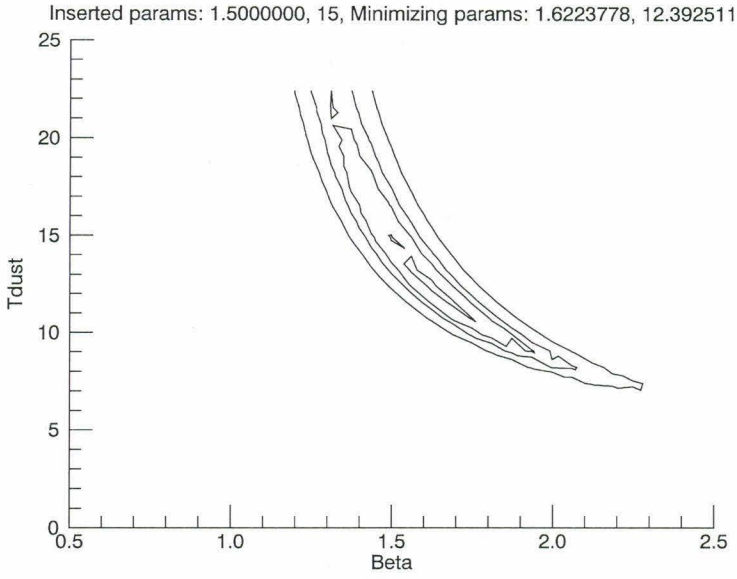


Figure 4.10: χ^2 surface for $\beta = 1.5, T = 15$K. Source inserted at SNR 20 in leading band with noise parameters from Section 2.1. Contours plotted are $\Delta\chi^2 = 0.1, 0.5, 1.0, 2.3$ the last corresponding to a 1σ deviation.

Per-band flux estimation

With little constraining power in the β–T plane of parameter space, we instead wish to quantify how well we can constrain the SED shape over the degenerate parameter space. To first get some idea of how the degeneracy degrades the SED fit, we can compute a the covariance matrix element $\sigma_A^2 = \frac{1}{2}\frac{\partial^2\chi^2}{\partial A^2}$, which describes the uncertainty on $\hat{A}$ with *fixed* β, T. The results of this computation are reported in Table 4.2 and are in good agreement with simulation.

Band	Pivot-based	Single-band
750μm	0.0531	0.0905
850μm	0.0421	0.0685
1.1μm	0.0250	0.0560
1.3μm	0.0174	0.0473
2mm	0.0064	0.0250
3mm	0.0019	0.0120

Table 4.2: Comparison of flux uncertainties in leading 750μm band with *fixed* β, T between single-band estimators and multi-band estimation, where multi-band has β, T fixed at inserted values $\beta = 1.7, T = 13$K. Units of mJy.

Of subsequent interest is the degrading of estimation once β, T are freed to their best fit values. Since β, T do not covary in Gaussian fashion, analytical prediction of the ensuing uncertainties in A are impossible and can only be obtained via simulation. Simulation results are ongoing.

Chapter 5

FUTURE WORK

5.1 Greybody Subtraction Systematics

The current stage of progress on greybody subtraction has encountered non-Gaussian degeneracies that can only be studied via simulation. Current and future work focus on characterizing these systematics. For instance, in computing the uncertainty of the flux estimator in each band using multi-band subtraction with β, T at their best-fit values (rather than fixed as studied at the end of Section 4.4), the computed uncertainties are observed to vary with SNR, and a sample of the behavior is given in Table 5.1.

Band	Fixed β–T	SNR 5	SNR 10	SNR 100
750μm	0.0531	0.0668	0.0698	0.0783
850μm	0.0421	0.0414	0.0424	0.0380
1.1μm	0.0250	0.0296	0.0327	0.0252
1.3μm	0.0174	0.0244	0.0259	0.0167
2mm	0.0064	0.0163	0.0090	0.0053
3mm	0.0019	0.0088	0.0049	0.0026

Table 5.1: Uncertainty on flux estimators in each band for fixed β–T analtical estimate and histogram widths for various SNRS over 10000 realisations while freeing β, T. SNR defined in leading frequency band 750μm.

Progress on sampling this dependence over many more values of SNR is under way, and the dependency can hopefully be cast at least qualitatively into illuminating terms. Of particular interest is the improvement in subtraction quality in the 2mm band, the peak emission frequency of the kSZ signal, and characterizing the error introduced in the 2mm band is pivotal to understanding source subtraction systematics in kSZ astronomy.

We are also in the process of examining any biases that appear when subtracting a source SED with a mis-fitted SED. For instance, if the source parameters are chosen from arbitrary Gaussian distributions $\beta = 1.7 \pm 0.3, T = 25 \pm 7$K (using arbitrary source distributions are still preferable to using the Béthermin Catalog which does not have known inserted distribution parameters for comparison) and are subtracted with the following fitting techniques:

Single Subtraction only within the single frequency band, no information from other frequencies used.

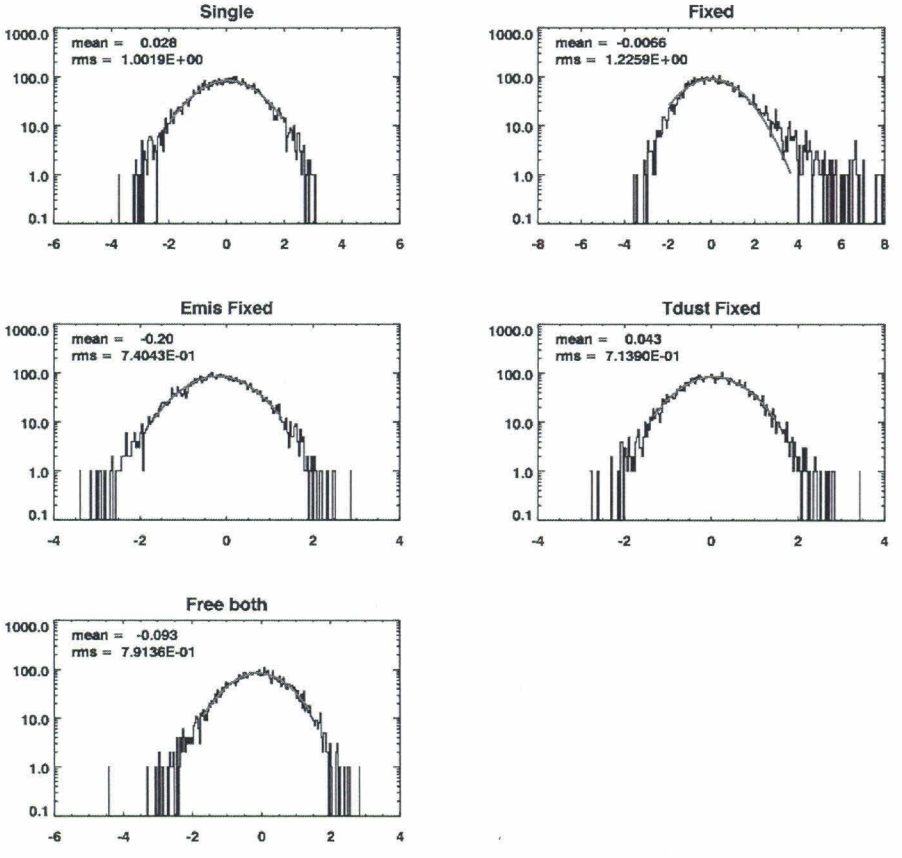
Fixed Subtraction with fixed SED $\beta = 1.7, T = 25\text{K}$.

Emis Fixed Subtraction with fixed $\beta = 1.7$ but T allowed to seek χ^2-minimizing value.

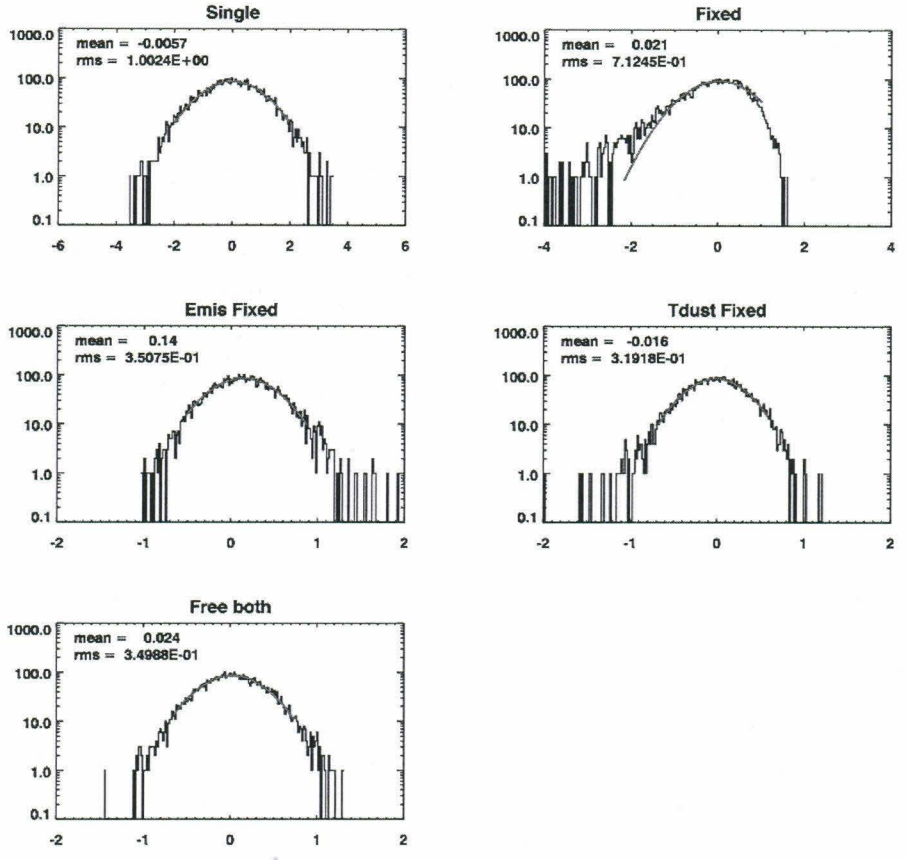
Tdust Fixed Subtraction with fixed $T = 25\text{K}$ but β allowed to seek χ^2-minimizing value.

Free Both Subtraction with both β, T allowed to seek χ^2 minimizing values.

then the resulting histograms of flux estimators in the leading 750μm band and the 2mm band of interest are in Figure 5.1. It is of interest that fixing T at some sensible temperature seems to both decrease the variance and not increase the bias of the flux estimators, but further investigations are ongoing.



(a) 750μm band



(b) 2mm band

Figure 5.1: Histograms of flux estimators using the five described fitting techniques for inserted sources with distribution of source parameters $\beta = N(1.7, 0.3), T = N(25, 7)$ where N is the normal distribution.

5.2 Application to Realistic Source Distribution

With the above results, we will have completed the characterization the behavior of the source subtraction algorithm in the absence of confusion noise, in the absence of other sources. The algorithm is believed then to perform optimally, at the accuracy limit of the formalism. The remaining work is to apply the algorithm to ensembles of point sources obtained from the updated Béthermin catalog (Béthermin, 2014). The subtraction in a map with an unknown number of sources will be performed until the variance of the convolved map doesn't change appreciably. Characterizing the SNR at which this happens, as well as the accuracy of the subtracted sources, must all be performed with the realistic source distribution provided by the catalog, and will finally provide a complete characterization of the source subtraction on a realistic catalog of point sources.

BIBLIOGRAPHY

Albrecht, A. et al. (2006). "Report of the Dark Energy Task Force". In: *ArXiv:0609591*. URL: http://adsabs.harvard.edu/abs/2006astro.ph..9591A.

Béthermin, Matthieu (2014). Personal Communication.

Béthermin, Matthieu et al. (2011). "Modeling the evolution of infrared galaxies: A Parametric backwards evolution model". In: *A&A* 529, A4.

Birkinshaw, M. (1979). "Limits to the value of the Hubble constant deduced from observations of clusters of galaxies". In: *Monthly Notices of the Royal Astronomical Society* 187.4, pp. 847–862. DOI: 10.1093/mnras/187.4.847. eprint: http://mnras.oxfordjournals.org/content/187/4/847.full.pdf+html. URL: http://mnras.oxfordjournals.org/content/187/4/847.abstract.

Birkinshaw, Mark (1999). "The Sunyaev-Zel'dovich effect". In: *Physics Reports* 310.2, pp. 97–195. URL: http://dx.doi.org/10.1016/S0370-1573(98)00080-5.

Carlstrom, John E., Gilbert P. Holder, and Erik D. Reese (2002). "Cosmology with the Sunyaev-Zel'dovich Effect". In: *Annual Review of Astronomy and Astrophysics* 40.1, pp. 643–680. DOI: 10.1146/annurev.astro.40.060401.093803.

Casey, Caitlin M. (2012). "Far-infrared spectral energy distribution fitting for galaxies near and far". In: *Monthly Notices of the Royal Astronomical Society* 425.4, pp. 3094–3103. ISSN: 1365-2966. DOI: 10.1111/j.1365-2966.2012.21455.x. URL: http://dx.doi.org/10.1111/j.1365-2966.2012.21455.x.

Casey, Caitlin M., Desika Narayanan, and Asantha Cooray (2014). "Dusty Star-Forming Galaxies at High Redshift". In: *Physics Reports* 541.2, pp. 45–161.

Cavaliere, A., L. Danese, and G. De Zotti (1979). "Cosmic Distances from X-ray and Microwave Observations of Clusters of Galaxies". In: *Astronomy and Astrophysics* 75, pp. 322–325.

Conley, A. et al. (2011). "Discovery of a Multiply Lensed Submillimeter Galaxy in Early HerMES Herschel/SPIRE Data". In: *The Astrophysical Journal Letters* 732.2, p. L35. URL: http://stacks.iop.org/2041-8205/732/i=2/a=L35.

Draine, B. T. (2006). "On the Submillimeter Opacity of Protoplanetary Disks". In: *The Astrophysical Journal* 636, pp. 1114–1120. DOI: 10.1086/498130. eprint: astro-ph/0507292.

Feldman, H. A., R. Watkins, and M. J. Hudson (2010). "Cosmic flows on 100h-1Mpc scales: standardized minimum variance bulk flow, shear and octupole moments". In: *Monthly Notices of the Royal Astronomical Society* 407, pp. 2328–2338. URL: http://adsabs.harvard.edu/abs/2010MNRAS.407.2328F.

Glenn, Jason et al. (1998a). "Bolocam: a millimeter-wave bolometric camera". In: *Annual Review of Astronomy and Astrophysics* 40, pp. 643–680. DOI: 10.1117/12.317418. URL: http://dx.doi.org/10.1117/12.317418.

Bolocam: a millimeter-wave bolometric camera (1998b). Vol. 3357, pp. 326–334. DOI: 10.1117/12.317418. URL: http://dx.doi.org/10.1117/12.317418.

Hand, N. et al. (2012). "Evidence of Galaxy Cluster Motions with the Kinematic Sunyaev-Zel'dovich Effect". In: *Physical Review Letters* 109(4).041101, pp. 1–6. URL: http://adsabs.harvard.edu/abs/2012PhRvL.109d1101H.

Itoh, Naoki, Yasuharu Kohyama, and Satoshi Nozawa (1998). "Relativistic Corrections to the Sunyaev-Zeldovich Effect for Clusters of Galaxies". In: *The Astrophysical Journal* 502.1, p. 7. URL: http://stacks.iop.org/0004-637X/502/i=1/a=7.

Kashlinsky, A. et al. (2008). "A Measurement of Large-Scale Peculiar Velocities of Clusters of Galaxies: Results and Cosmological Implications". In: *The Astrophysical Journal Letters* 686, pp. L49–L52. URL: http://adsabs.harvard.edu/abs/2008ApJ...686L..49K.

Ma, Y.-Z., E. Branchini, and D. Scott (2012). "A comparison of the galaxy peculiar velocity field with the PSCz gravity field - a Bayesian hyper-parameter method". In: *Monthly Notices of the Royal Astronomical Society* 425, pp. 2880–2891. URL: http://adsabs.harvard.edu/abs/2012MNRAS.425.2880M.

Nusser, A. and M. Davis (2011). "The Cosmological Bulk Flow: Consistency with ΛCDM and z ≃ 0 Constraints on σ_8 and γ". In: *The Astrophysical Journal* 736.93, pp. 1–9. URL: http://adsabs.harvard.edu/abs/2011ApJ...736...93N.

Osborne, S. J. et al. (2011). "Measuring the Galaxy Cluster Bulk Flow from WMAP Data". In: *The Astrophysical Journal* 737.98, pp. 1–20. URL: http://adsabs.harvard.edu/abs/2011ApJ...737...98O.

Rangwala, Naseem et al. (2011). "Observations of Arp 220 Using Herschel-SPIRE: An Unprecedented View of the Molecular Gas in an Extreme Star Formation Environment". In: *The Astrophysical Journal* 743.1, p. 94. URL: http://stacks.iop.org/0004-637X/743/i=1/a=94.

Robitaille, Thomas P. and Barbara A. Whitney (2010). "The Present-Day Star Formation Rate of the Milky Way Determined from Spitzer-Detected Young Stellar Objects". In: *The Astrophysical Journal Letters* 710.1, p. L11. URL: http://stacks.iop.org/2041-8205/710/i=1/a=L11.

Sayers, J. et al. (2010). "Studies of Millimeter-wave Atmospheric Noise above Mauna Kea". In: *The Astrophysical Journal* 708.2, p. 1674. URL: http://stacks.iop.org/0004-637X/708/i=2/a=1674.

Sayers, J. et al. (2013). "A Measurement of the Kinetic Sunyaev-Zel'dovich Signal Towards MACS J0717.5+3745". In: *The Astrophysical Journal* 778.52, pp. 1–20. URL: http://stacks.iop.org/0004-637X/778/52.

Sunyaev, R. A. and Y. B. Zeldovich (1972). "The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies". In: *Comments on Astrophysics and Space Physics* 4, pp. 173–178. URL: http://adsabs.harvard.edu/abs/1972CoASP.

Appendix A

BLENDS

As discussed in Section 4.1, we occassionally can have overlapping sources that are mischaracterized as a single source as in Figure 4.4. The effect of this phenomenon were manifestly clear when comparing Figure 4.1 and Figure 4.3: while the flux of a source when there is only one source in the map is estimated correctly with the expected uncertainty, when there are many sources the probability of source overlap increases, biasing the estimators and producing a larger variance than predicted. Because the procedure estimates two adjacent sources as one very bright source and one very dim source, the sum being slightly less than the sum of the original two, $\langle \hat{A} \rangle$ is biased low and σ_A^2 is larger than predicted.

To fully account for blends, two distinct steps are required: first, the number of sources that make up a blend must be identified (if it is not a blend, it can be considered a blend of a single source), and second, the position and flux of these sources must be estimated.

A.1 Iterative subtraction

First we attack the latter problem, estimating the position and flux of blended sourcces assuming we know how many sources make up the blend. It turns out that an iterative procedure works well to estimate the position and flux of the individual sources. Call N the number of sources in the blend. We first subtract N sources, and then we add each estimated source back in before re-subtracting it. Repeating this for all N sources separately for many iterations, the estimators converge to the inserted fluxes and positions. In Figure A.1 we see that over many iterations the source fluxes converge to the correct value. Such an approach is generally termed *iterative*.

To solve the former problem, identifying the number of sources that make up a blend, two methods were developed, one empirical and one theoretically rigorous, detailed below.

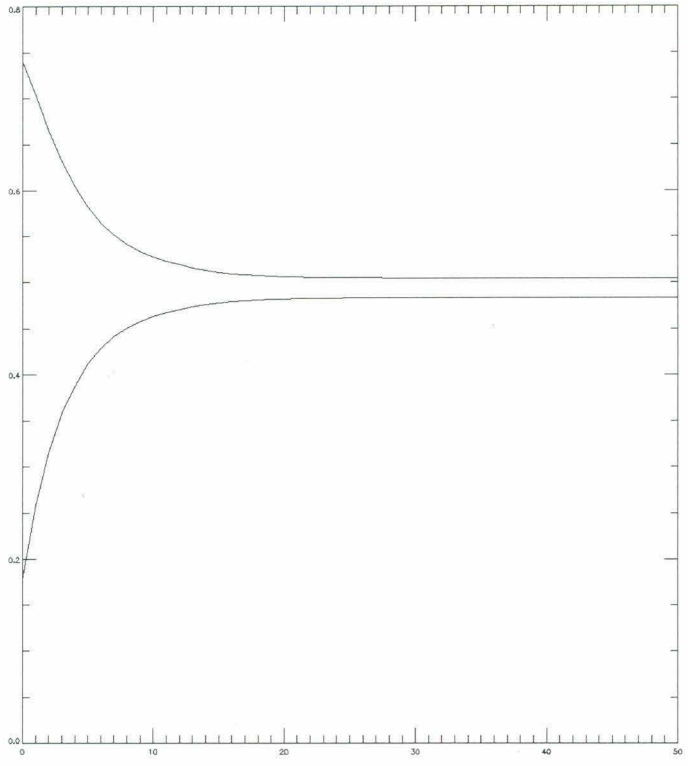


Figure A.1: Plot of flux estimators of two sources placed two FWHM apart as both position and flux estimators are iteratively improved. Note that while we initially grossly misestimate the sources, iterative estimation rapidly converges towards the correct values of two sources of flux 0.5 arbitrary units.

A.2 Mean-based Proof of Concept

One method for identifying the number of sources in a blend was devised based on the mean of the subtracted map. Recall that the generated maps are assumed to have zero-mean noise and positive source beams, so a perfectly subtracted map should on average have zero mean as well.

First, the above iterative technique is applied assuming the blend comprises some number N of sources. If the mean of the subtracted map is appreciably greater than zero, the iterative technique is reapplied assuming the blend comprises $N+1$ sources. This procedure is repeated until the mean of the subtracted map is sufficiently close to zero; the criterion for this closeness is discussed below.

The approach was motivated by the observation that the mean of the residual map would only be close to zero if the correct number of sources was subtracted, illustrated in Table A.1

Beams in map	1 source	2 source	3 source	4 source
1 source	0.000197	8.94e-05		
2 source	0.0230	0.000371	0.000571	
3 source	0.0341	0.0149	-6.70e-05	0.000200

Table A.1: Means of residual map as a function of number of sources present and number of sources subtracted, averaged over 200 trials each. While the number of trials is formally inconclusive, it is clear that the mean only falls close to zero when the correct number of sources is subtracted.

However there are a few notable problems with this approach:

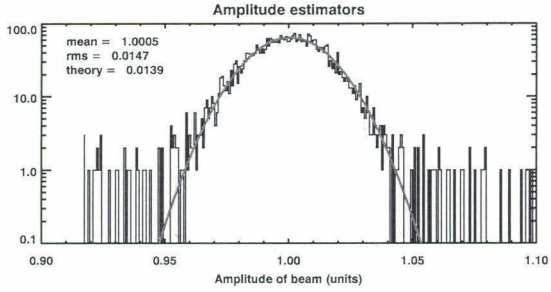
- The mean of a map is generally not identified with anything physical, and data are generally reported with zero mean including the sources. This implies that the mean of a subtracted map with the correct number of sources will not be zero but some indeterminate negative quantity.
- There is no *ab initio* prediction for the deviation of the mean of the subtracted map from zero. Without this, it is impossible to define the criterion for the number of sources required to describe a blend.
- There is a more statistically rigorous way to formulate the necessity of adding an additional pulse, in the form of an F-test.

However, it is evident that using the mean to distinguish blends does decrease subtraction bias, as can be seen in Figure A.2. Thus, while incorrect, the mean-based subtraction demonstrates that seeking an appropriate method to distinguish blends is a worthwhile pursuit.

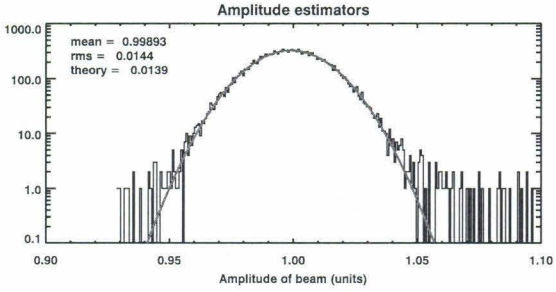
A.3 F-test

The statistically rigorous criterion for the number of sources required to describe a blend is based on the F-test. A form of the F-test can be applied to determine whether addition of more parameters to a fit is statistically justifiable, which in our case translates to determining whether an additional beam is required to describe a particular blend.

Given two hypothesis H_1, H_2 with H_2 an extension of H_1 with additional fit parameters, the test statistic $\frac{(\chi_1^2 - \chi_2^2)/(\nu_1^2 - \nu_2^2)}{\chi_2^2/\nu_2^2}$ (with χ^2 defined in the usual way) can be shown to follow an F-distribution with degrees of freedom $\nu_1 - \nu_2, \nu_2$. Then if



(a) Histogram of flux estimators using mean as subtraction threshold, with 1000 realisations of noise and 20 sources in each of unit flux. The histogram is centered on the correct value, though being slightly wider than the theoretical limit. This is understandable due to some maps containing sources in close proximity. Arbitrary values of no significance in magnitude used.



(b) Histogram of flux estimators using no subtraction threshold, with 1000 realisations of noise and 20 sources in each of unit flux. The histogram is noticably biased rightwards.

Figure A.2: Histogram of flux estimators between using and not using mean subtraction threshold. Note the less biased and more symmetric estimator histogram when using the mean subtraction threshold.

the test statistic exceeds some critical value we know with high probability that H_2 is the preferred fit to the data over H_1.

Given this F-test, the procedure is then the same as before, which is to fit each blend with an increasing number of beams until the F-test predicts that the addition of another beam produces insignificant changes in the χ^2.

However the implementation of this procedure produced an unexpected result: the test statistic changes depending on the size of the map used even when all other characteristics are held constant. If a map much larger than the beam width is used to calculate the χ^2 for a single blend, the test statistic frequently asserts that a one-source map should be subtracted as a two-source map. Only by arbitrarily truncating the map, say, 3FWHM on either side of the target source, does the test statistic correctly identify the number of sources in a map.

The numerical results of these tests can be found in Table A.2. The F-test predicts that for all these configurations the critical value is somewhere in the range $\sim 2.64 \pm 0.05$ (since all configurations have extremely large ν_2 degrees of freedom, their critical values are very similar). Note that as the size of the considered map changes, so too does the average value of $\chi_{11} - \chi_{12}$, the χ^2 difference between subtracting a single-source map as a single beam or as a two-beam blend. This is inexplicable: in a single-source map, attempting to subtract a second source should not change χ^2 much, as the second subtracted source cannot correspond to a real source and so should have very low flux regardless of the size of the map considered.

Number of pixels	χ_{11}	$\sigma_{\chi_{11}}$	χ_{12}	$\sigma_{\chi_{12}}$	$\chi_{11} - \chi_{12}$	$\sigma_{\chi_{11}-\chi_{12}}$
20	397.54	27.58	396.30	27.63	1.23	1.59
32	1021.17	45.21	1017.90	45.22	3.27	2.49
64	4089.20	92.25	4082.50	92.11	6.70	2.92
128	16378.42	181.19	16368.92	181.04	9.51	2.94
256	65526.80	365.46	65514.52	365.50	12.28	3.07
	χ_{21}	$\sigma_{\chi_{21}}$	χ_{22}	$\sigma_{\chi_{22}}$	$\chi_{21} - \chi_{22}$	$\sigma_{\chi_{21}-\chi_{22}}$
20	918.21	50.86	528.00	199.79	390.20	200.00
32	1310.87	55.10	1019.09	45.23	291.78	34.01
64	4409.64	99.87	4087.01	92.13	322.64	36.78
128	16702.46	186.40	16376.18	181.02	326.28	36.04
256	65851.73	366.16	65524.56	365.51	327.17	36.52

Table A.2: Average χ^2 values for various different configurations of map and subtraction parameters; notation is such that first subscript is number of inserted sources and second subcript is number of sources attempted to subtract (e.g. χ_{21} is a map with 2 sources sources being subtracted as a single source). When there are two sources, they are placed 1.2 FWHM apart. Averaged over 10000 realisations for each configuration. Image dimensions are 256×256. All uncertainties were computed by performing Gaussian fits to very Gaussian data, as expected as the χ^2 distribution tends to a larger number of degrees of freedom.

The test statistics indicate that only an appropriately sized subportion of the map can be considered when using the F-test to determine whether a second source exists. This is unacceptable as there is no obvious criterion for determining how large a sub-map to consider.

In conclusion, it does not seem at the present that there is a viable criterion to identify how many sources comprise a blend, without which blends cannot be treated differently than single sources. Were such a criterion to be identified, Figure A.2 suggests that the biases observed on the $\hat{A}$ estimators could be remedied. Ultimately however, a technique for treating blends may not be necessary, as with a

realistic source distribution the residual of a mis-subtracted blend is expected to be dominated by confusion noise generated by sources too dim to be subtracted.