



PCAT-DE: Reconstructing Pointlike and Diffuse Signals in Astronomical Images Using Spatial and Spectral Information

Richard M. Feder¹ , Victoria Butler² , Tansu Daylan^{3,7} , Stephen K. N. Portillo^{4,5} , Jack Sayers¹ , Benjamin J. Vaughan² , Catalina V. Zamora⁴ , and Michael Zemcov^{2,6}

¹ California Institute of Technology Division of Physics, Math, and Astronomy, 1200 East California Boulevard, Pasadena, CA 91125, USA; rfederst@caltech.edu

² Rochester Institute of Technology, 1 Lomb Memorial Drive, Rochester, NY 14623, USA

³ Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA

⁴ DIRAC Institute, Department of Astronomy, University of Washington, 3910 15th Avenue NE, Seattle, WA 98195, USA

⁵ Department of Mathematical and Physical Sciences, Concordia University of Edmonton, 7128 Ada Boulevard, Edmonton, AB, T5B 4E4 Canada

⁶ Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Received 2023 January 4; revised 2023 June 21; accepted 2023 July 10; published 2023 August 10

Abstract

Observational data from astronomical imaging surveys contain information about a variety of source populations and environments, and their complexity will increase substantially as telescopes become more sensitive. Even for existing observations, measuring the correlations between pointlike and diffuse emission can be crucial to correctly inferring the properties of any individual component. For this task, information is typically lost, because of conservative data cuts, aggressive filtering, or incomplete treatment of contaminated data. We present the code PCAT-DE, an extension of probabilistic cataloging, designed to simultaneously model pointlike and diffuse signals. This work incorporates both explicit spatial templates and a set of nonparametric Fourier component templates into a forward model of astronomical images, reducing the number of processing steps applied to the observed data. Using synthetic Herschel-SPIRE multiband observations, we demonstrate that point-source and diffuse emission can be reliably separated and measured. We present two applications of this model. For the first, we perform point-source detection/photometry in the presence of galactic cirrus and demonstrate that cosmic infrared background galaxy counts can be recovered in cases of significant contamination. In the second, we show that the spatially extended thermal Sunyaev–Zel’dovich effect signal can be reliably measured even when it is subdominant to the pointlike emission from individual galaxies.

Unified Astronomy Thesaurus concepts: [Catalogs \(205\)](#); [Diffuse radiation \(383\)](#); [Markov chain Monte Carlo \(1889\)](#); [Bayesian statistics \(1900\)](#); [Astrostatistics tools \(1887\)](#)

1. Introduction

The signal of interest in astronomical images is often contaminated by one or more other signals. These additional components can bias estimates of the desired signal when left unmodeled and lower the precision with which we can infer correlated spatio-spectral structure. Estimating the effect of such components is a challenge, and mitigation strategies are situation-dependent.

Oftentimes, the goal is to measure the emission from spatially unresolved sources (hereafter referred to as point sources) in the presence of diffuse signals; for example, radio sources in front of the cosmic microwave background (CMB) or behind galactic synchrotron (Hale et al. 2019), or stars embedded in regions of high nebulosity (de Bruijne et al. 2015). The effects of diffuse structured signals can sometimes be mitigated using the fact that many diffuse astrophysical signals have angular power spectra that decrease with wavenumber. This motivates spatial (or angular) high-pass filtering, either in real or Fourier space. However, filtering approaches necessarily attenuate and distort the signal of interest and often add uncertainties to signal estimates that can

be difficult to assess. In the other limit, there are cases where the signal of interest is some type of diffuse structured emission and point sources are the contaminants. A common approach is to mask out known or suspected point-source contaminants (e.g., Barreiro 2009), but such approaches can be problematic when the spatial density of point sources and/or beam size necessitates removing a significant fraction of the image (e.g., Zemcov et al. 2014). Crucially, such removal is always to a finite detection limit, and the remaining point sources contaminate the estimate for the diffuse emission. This effect can be characterized, again at the cost of larger uncertainties on the signal of interest (e.g., Traficante et al. 2011).

Many methods for separating pointlike and diffuse signals exist. A review of source detection strategies is presented in Masias et al. (2012); work in this field since this review includes Masias et al. (2015), Zheng et al. (2015), Portillo et al. (2017), Ofek & Zackay (2018), Robotham et al. (2018), Lukic et al. (2019), Collin et al. (2022), Liu et al. (2021), and Du et al. (2022). When spatial and/or spectral source properties are well understood, matched filtering is an effective method of source extraction (Ofek & Zackay 2018; Lang & Hogg 2020), though optimal results only hold under strict assumptions, e.g., sources are isolated in background-dominated images, with perfect knowledge of the point-spread function (PSF), noise model, etc. Multiscale methods decompose images into components with fluctuation power on different spatial scales, enabling more reliable source detection and deblending in the presence of noise and structured backgrounds (Molinari et al. 2011;

⁷ LSSTC Catalyst Fellow.



Men’shchikov et al. 2012). These approaches can involve several transformations of the data, meaning the quality of extraction of one component (typically, the point source) is emphasized at the cost of poor fidelity of the other components, although work has been done to improve signal reconstructions through more informed transformations and data representations (Dabbech et al. 2015; Ellien et al. 2021). It has been shown that point-source photometry in the presence of nebulosity can be improved considerably by learning a pixelwise, nonstationary covariance matrix for the structured signal surrounding each source (see Saydjari & Finkbeiner 2022 and the application to the DECaPS2 survey in Saydjari et al. 2023). However, the use of a fixed input catalog in the post-processing step means that errors related to biases in source detection in the presence of diffuse signals are left uncorrected.

The performance of any photometry tool is tied to fundamental constraints on the information that can be extracted from astronomical images. The underlying parameters θ that describe the sky signal, the raw image data \mathcal{I} , and the processed data or downstream summary statistics $g(\mathcal{I})$ form a Markov chain $\theta \rightarrow \mathcal{I} \rightarrow g(\mathcal{I})$. As such, the data processing inequality requires that the mutual information between θ and $g(\mathcal{I})$ is always less than or equal to that between θ and the original data, i.e., $I(\theta; g(\mathcal{I})) \leq I(\theta; \mathcal{I})$ (Beaudry & Renner 2011). While in some cases (often under strict assumptions) it is possible to construct “sufficient statistics,” which satisfy $I(\theta; g(\mathcal{I})) = I(\theta; \mathcal{I})$, methods that can directly access the mutual information between θ and \mathcal{I} will be crucial for extracting the full information content of increasingly rich data sets.

The work presented in this paper builds on *probabilistic cataloging* (hereafter, PCAT; Brewer et al. 2013; Brewer & Donovan 2015; Jones et al. 2015; Daylan et al. 2018), a framework that combines transdimensional inference (Green 1995) with Bayesian hierarchical modeling to sample from a *metamodel* (a union of models with different dimensionality) consistent with observed astronomical data. We extend the forward model to handle map data in which the observed signal can be composed as the sum of point sources, a diffuse fluctuation component modeled through a set of Fourier component templates, and surface brightness templates of unknown amplitude. This extension is implemented in the code Probabilistic CATaloging in the presence of Diffuse Emission (PCAT-DE; Feder 2023).

PCAT-DE is tested on a variety of synthetic observations from the Spectral and Photometric Imaging REceiver (SPIRE), an instrument on board the 3.5 m Herschel space observatory (Griffin et al. 2010). The different applications in this work make assumptions about the spatial and spectral behaviors of the components, but PCAT-DE is flexible and can handle the properties of different models as long as they are properly specified. Possible use cases include but are not limited to: separation of infrared sources and the CMB at submillimeter wavelengths; pointlike source cataloging in the presence of large fixed-pattern detector noise; extraction of point sources over large-scale gradients caused by Zodiacal Light or fluctuations in atmospheric transmission/brightness for ground-based data; and separation of X-ray point sources from diffuse galaxy cluster emission, among others.

The paper is structured as follows. We begin in Section 2 with an introduction to PCAT and its extension to modeling diffuse emission. The mock Herschel-SPIRE data sets are

introduced in Section 3, and we test the performance of our implementation on reconstructing blended emission components based on a range of models and data in Section 4. The first application explores how well point sources and their properties are detected/measured (Section 5), while the second models out the impacts of point sources and diffuse emission on the thermal Sunyaev–Zel’dovich (SZ) effect (Section 6). We conclude in Section 7 with a discussion of the current PCAT-DE implementation and propose a number of potential applications for this formalism.

2. PCAT

As telescopes become more sensitive, source extraction becomes increasingly limited by the ability to spatially resolve overlapping sources (Marshall et al. 2017). This is driven by the gap between flux sensitivity and angular resolution, which becomes important as one pushes to fainter depths. For current and near-future surveys, an increasingly large fraction of sources that would be reliably measured in isolation will be observed as partial or full blends with adjacent sources, complicating both the identification and measurement of bright and faint objects (Breivik et al. 2022). For some data sets, a fast mapping rate is prioritized over angular resolution, and these surveys in particular will approach sensitivities where source blending is relevant, both in the spatial and spectral domains.⁸ For example, source blending will be a major source of systematic uncertainty in a variety of Stage-IV cosmology surveys that rely on accurate galaxy photometry (Melchior et al. 2021).

These challenges motivate PCAT. By sampling the full catalog space consistent with a given data set, PCAT can be used to infer both the properties of the astronomical sources present *and* the number of sources itself, above some flux density threshold. PCAT models sources below conventional significance thresholds (i.e., $<5\sigma$), which enables the detection of faint sources and less biased constraints of bright sources with faint neighbors. As a Bayesian hierarchical modeling framework, PCAT is capable of incorporating complex information into a self-consistent model of the signal and data-generating process, assuming knowledge of the causal chain that leads to observed data. In this approach, marginalization over different parameters is performed by collecting fair draws from the posterior of the forward model, given the data. Composable models like those used in PCAT are easily interpretable by directly testing the addition or removal of components or by modifying model priors.

Applied to single-band optical images of the globular cluster M2 taken from the Sloan Digital Sky Survey, PCAT recovers sources with completeness one magnitude deeper than the crowded-field photometry tool DAOPHOT (Stetson 1987), and it has a lower false discovery rate for brighter sources (Portillo et al. 2017). Performance by these metrics is further improved by extending the hierarchical model to multiband data, in which case the maps are fit simultaneously (Daylan et al. 2017; Feder et al. 2020).

⁸ For certain large-area surveys, the conventional catalog of subthreshold point sources may not contain more information than the intensity maps themselves (Cheng et al. 2019; Schaan & White 2021).

2.1. Modeling Astronomical Images

Let λ_{ij}^b denote the surface brightness in pixel (i, j) of band b . The model used to generate images within PCAT-DE can be written as a sum over point sources and diffuse signals:

$$\lambda_{ij}^b = \mathcal{P}^b \otimes \left[\sum_{n=1}^{N_{\text{src}}} S_n^b \delta(x_i - x_n^b, y_j - y_n^b) + A_b^t \sum_{t=1}^{N_{\text{temp}}} I_t^b(x_i, y_j) + B_{ij}^b \right]. \quad (1)$$

In this equation, \mathcal{P}^b is the beam function kernel that convolves the signal measured in band b by the PSF. The signal is decomposed into a sum of point sources with flux densities $\{S_n\}_{n=1}^{N_{\text{src}}}$, N_{temp} templates for resolved components with known position/spatial structure encoded in surface brightness templates $\{I_t^b\}$ and amplitudes $\{A_b^t\}$ (see Section 6), and a generic term for additional diffuse signal B_{ij}^b . In images with negligible diffuse structure (or a small enough field of view), a simpler mean normalization in each band, B_0^b , may be sufficient. We use B_{ij}^b to specify diffuse signals without a priori spatial structure, for which a more flexible, nonparametric model is used in signal reconstruction (see Section 2.2).

For each band b , let \mathbf{d}_b define the data vector corresponding to the unraveled image with size $W_b \times H_b$ pixels and corresponding per-pixel errors given by σ_b^2 . In this work, we assume that errors are known for each pixel and independent of one another, meaning the likelihood can be written as a product over all pixels. We further assume these errors are Gaussian distributed. For the purpose of Markov Chain Monte Carlo (MCMC) sampling, we compute log-likelihoods, turning the products over pixels and bands into sums:

$$\log \mathcal{L} \approx \sum_{b=1}^B \sum_{i=1}^{W_b} \sum_{j=1}^{H_b} - \frac{(\mathbf{d}_b - \lambda_b)^2}{2\sigma_b^2}. \quad (2)$$

While the likelihood is in a space of fixed dimension set by the data, the space of the models is transdimensional, i.e., it is the union of catalogs with varying N_{src} . Because these models reside in nested subspaces of one another, priors can be placed on individual mixture components (the point sources) while defining a posterior over the full model space (Portillo et al. 2017).

Figure 1 shows a representation of the PCAT-DE generative model as a probabilistic graphical model (PGM). The different layers of the PGM correspond to levels of the Bayesian hierarchical model—at the highest level, priors on the point-source population (a power-law flux prior for sources with spectral index α and a Gaussian prior on colors with mean μ_S and width σ_S) and diffuse components (e.g., the color of the diffuse component encoded in $\beta(S)$) inform the prior distributions over mixture components. The point sources and diffuse components are then convolved by the instrument beam to produce model images that can be compared directly to the observed data. For this work, note that N_t is fixed, while N_{src} is floated as a free parameter.

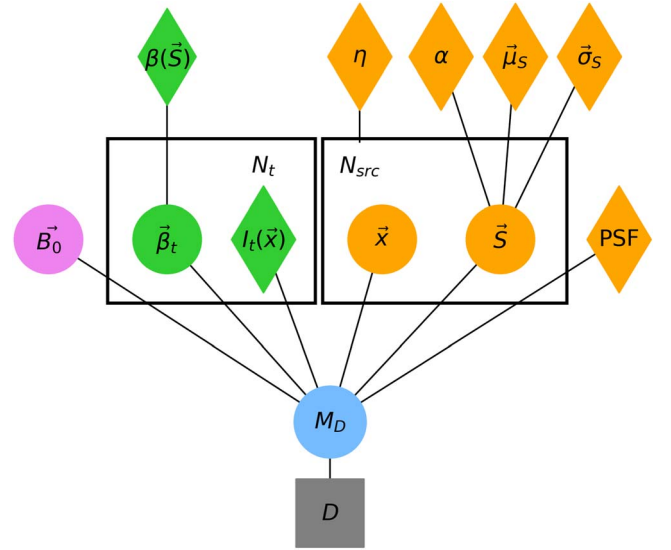


Figure 1. PGM for PCAT-DE. The top level of the PGM shows hyperparameters (α , μ_S , σ_S , η , $\beta(S)$) that characterize priors over the point-source parameters $\{x, S, N_{\text{src}}\}$ and diffuse component colors $\{\beta_i\}$. These parameters are then used to generate model images M_D that are compared with the data. The diamonds and circles indicate variables that are fixed and floated, respectively.

2.2. Modeling Diffuse Signals with Fourier Component Templates

Diffuse signals are modeled by PCAT-DE using a linear combination of Fourier component templates, where each template represents a separate Fourier mode. An arbitrary diffuse signal can be approximated by a truncated two-dimensional Fourier series:

$$B_{ij}^b = B_0^b + \sum_{n_x=1}^{N_{\text{FC}}} \sum_{n_y=1}^{N_{\text{FC}}} \beta_{n_x n_y} \cdot \mathcal{F}_{ij}^{n_x n_y}. \quad (3)$$

Here, N_{FC} refers to the order of the Fourier series and $\mathcal{F}_{ij}^{n_x n_y}$ is a vector of Fourier components corresponding to wavevector $(k_x, k_y) = (W/n_x, H/n_y)$ evaluated at a pixel with index (i, j) :

$$\mathcal{F}_{ij}^{n_x n_y} = \begin{pmatrix} \sin\left(\frac{n_x \pi x_j}{W}\right) \sin\left(\frac{n_y \pi y_j}{H}\right) \\ \sin\left(\frac{n_x \pi x_j}{W}\right) \cos\left(\frac{n_y \pi y_j}{H}\right) \\ \cos\left(\frac{n_x \pi x_j}{W}\right) \sin\left(\frac{n_y \pi y_j}{H}\right) \\ \cos\left(\frac{n_x \pi x_j}{W}\right) \cos\left(\frac{n_y \pi y_j}{H}\right) \end{pmatrix}. \quad (4)$$

The vector $\beta_{n_x n_y}$ encodes the amplitudes of the Fourier components. All four components of $\mathcal{F}_{ij}^{n_x n_y}$ are necessary in the absence of boundary conditions on the images, which in general will be arbitrary. Throughout this work, we use the parameter N_{FC} when comparing models. The minimum angular scale captured by the Fourier components can be approximated by the half-period of the highest-frequency Fourier mode along each dimension.⁹

⁹ The smallest angular scale is actually set by the norm of the wavevector, $|k|_{\text{max}} = \sqrt{k_{x,\text{max}}^2 + k_{y,\text{max}}^2}$.

Fourier component templates are well suited for the tasks at hand. Using a truncated Fourier series has the benefit of robustness against bias from small-scale, localized signal map features (e.g., unmodeled point sources). This is because each template has global extent over the image and the choice of truncation scale implies certain modes are impossible to reconstruct with the Fourier series. For astronomical images, the minimum effective scale is typically bounded by the beam rather than the chosen map pixel size. In general, the point-source model provides a good description of the fluctuations on the scale of the beam, but there is a range of intermediate scales larger than the beam and smaller than the image size where power from diffuse signals can reside. Fortunately, the falling angular power spectrum characteristic of many diffuse signals implies they may be well described by models for which $k_\theta^{\max} < k_\theta^{\text{beam}}$, where k_θ^{beam} corresponds to the angular scale of the PSF FWHM. Many methods for point-source detection, such as `SExtractor` (Bertin & Arnouts 1996) and `Starfinder` (Diolaiti et al. 2000), include empirical estimates of the local sky signal surrounding each source using a pixelwise mean or median. The Fourier component representation is flexible enough to accomplish effective sky subtraction, but the underlying feature of PCAT’s forward modeling is the capability to incorporate physically motivated priors for diffuse components within a larger Bayesian hierarchical model (including point sources and realistic observational noise).

A set of linear marginalization steps completed at the beginning of the sampling (see Appendix A) accelerates the burn-in phase of the sampling, after which the Fourier coefficients are sampled with the same Metropolis–Hastings algorithm used for the rest of the model parameters. In practice, the Fourier components converge in a similar number of iterations as the rest of the model. The proposal kernel of each template is chosen by approximating the Fisher information of a uniform background component in the presence of several point sources (see Appendix C for a derivation).

3. Mock Data

In this section, we describe the astrophysical components that are combined to generate mock observations with similar noise properties as a range of shallow and deep SPIRE observations. SPIRE included a three-band imager with bandpasses centered at 250, 350, and 500 μm and beam FWHMs of 18", 25", and 36", respectively (Griffin et al. 2010). While PCAT–DE has been applied to real SPIRE data in Butler et al. (2022), controlled sets of mocks are used here in order to characterize the performance of the implementation in different limits. SPIRE maps typically contain a combination of emission from cosmic infrared background (CIB) galaxies and diffuse galactic cirrus. Galaxy cluster observations with SPIRE also contain localized but faint and extended signals from the thermal SZ effect. These maps are typically dominated by fluctuations in the total signal from individually undetected (and spatially unresolved) CIB galaxies, known as “confusion noise” (Condon 1974; Nguyen et al. 2010), providing a difficult scenario for point-source extraction in the presence of diffuse contaminants. At the angular scale of the SPIRE beam, the underlying CIB luminosity function (LF) is extremely steep (for differential SPIRE number counts, see Figure 13 of Casey et al. 2014), resulting in a large number of sources just below the typical detection limit within a PSF-sized aperture. As a

result, the source confusion in SPIRE observations should be considered more severe than the “typical” use case, in which sources are well separated and Poisson fluctuations are larger.

3.1. CIB Galaxies

The majority of sources detected at far-infrared (FIR) wavelengths are $z \sim 2$ galaxies with an angular extent of $\sim 1''$. When convolved with the much larger SPIRE beam, most galaxies in SPIRE observations are well modeled as point sources. Mock realizations are generated using the CIB model described in Béthermin et al. (2012; referred to as **B12** throughout this work). On the scales of the images considered ($\theta \leq 10$ arcmin), the CIB power spectrum is dominated by shot noise from galaxies. More details about the construction of this CIB component can be found in Butler et al. (2022).

3.2. Galactic Cirrus Foregrounds

A significant source of diffuse emission, even at high galactic latitudes, is Galactic cirrus dust, which reprocesses the interstellar radiation field and emits thermal radiation in the FIR (Desert et al. 1990). While cirrus has a blue spectrum across the SPIRE bands, similar to that of many observed CIB sources, cirrus contains the most fluctuation power over larger angular scales. To calibrate the level of cirrus emission present in extragalactic observations for this study, we apply the Planck SZ union foreground mask¹⁰ to Planck observations and sample positions uniformly across the unmasked sky. The maps at these positions are queried, regridded to SPIRE resolution, and extrapolated to 250, 350, and 500 μm using the Planck-estimated parameters of a modified blackbody spectral energy distribution (SED),

$$S(\nu) = A \left(\frac{\nu}{\nu_0} \right)^\beta B_\nu(T_d), \quad (5)$$

where ν is the rest-frame frequency, ν_0 is the reference frequency at which the optical depth is measured, β is the spectral index, T_d is the dust temperature, and $B_\nu(T_d)$ is the spectral radiance at frequency ν , assuming thermal equilibrium for temperature T_d . They have dust temperatures ranging from 19 to 22 K over 15' patches of the sky, while $\beta \sim 1.5 \pm 0.1$ at the same smoothing scale (Planck Collaboration et al. 2014). We define a nominal “ $1 \times$ -Planck” case as a diffuse signal whose power spectrum is parameterized by a single power-law slope:

$$P(k) = P_0(|k|/k_0)^\gamma, \quad (6)$$

where $\gamma = -2.6$ (Bracco et al. 2011) and P_0 is determined by the unmasked sky-averaged power spectrum. Synthetic cirrus maps are drawn as Gaussian random realizations of the $1 \times$ -Planck power spectrum, and progressively more severe cirrus realizations are obtained by scaling the amplitude of fluctuations, $\sqrt{P(k_\theta)}$, by factors of 2, 4, and 8. The range of cirrus realizations considered in this work with increasing fluctuation power is representative of the worst 50%, 32%, 17%, and 5% of the Planck unmasked sky.

Last, mock observations are generated for a range of noise levels ranging from 1 mJy beam⁻¹ (confusion-dominated for

¹⁰ Maps can be found here: https://irsa.ipac.caltech.edu/data/Planck/release_1/ancillary-data/previews/COM_PCCS_SZ-unionMask_2048_R1.11/index.html.

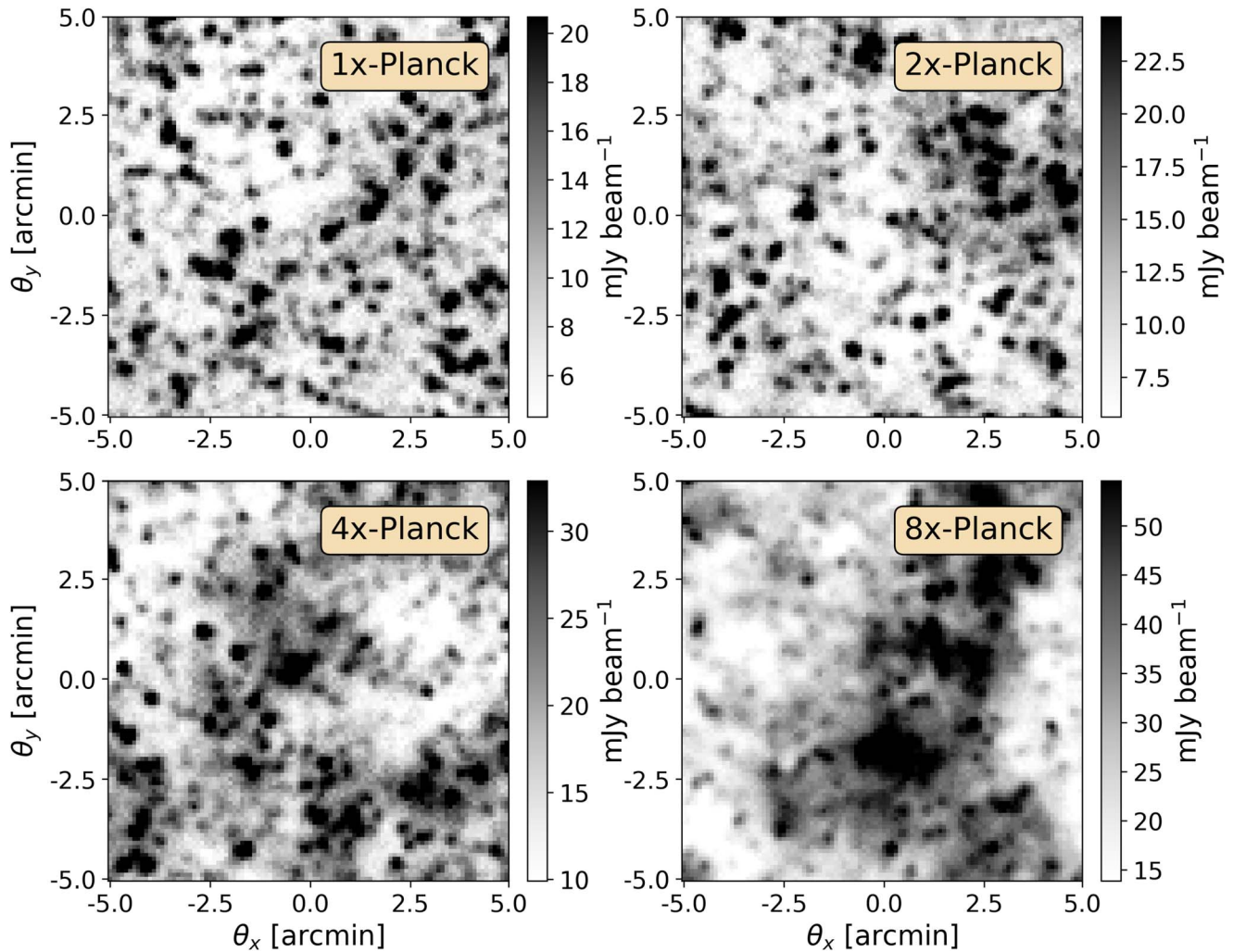


Figure 2. Mock realizations of CIB and cirrus at SPIRE resolution with 1 mJy beam^{-1} instrument noise. In this limit, confused point sources are the dominant agent reducing our modeling precision. The different panels show levels of cirrus consistent with the median “clean-sky” Planck sample (top left) and progressively more contaminated fields.

SPIRE) to 6 mJy beam^{-1} (instrument noise roughly equal to confusion noise). Instrument noise at the fiducial SPIRE map resolution is well described by a diagonal map space covariance matrix (Viero et al. 2013) and is dominated by thermal emission from the primary mirror.

Figure 2 shows a set of $10' \times 10'$ mock observations at $250 \mu\text{m}$ with varying levels of synthetic cirrus emission and instrument noise at the 1 mJy beam^{-1} level. Uncorrected diffuse signals have the effect of boosting sources spatially coincident with positive fluctuations, while suppressing sources coincident with negative fluctuations. 20 sets of multiband CIB sky realizations are combined with different levels of synthetic cirrus throughout the results. In Sections 4 and 5, only the single-band $250 \mu\text{m}$ maps are used, but the full three-band maps are used in Section 6, where color information helps to distinguish the SZ effect from other astrophysical components.

4. Separating Pointlike and Diffuse Emission in Astronomical Images

4.1. Model Priors

The priors used in this work are nearly identical to those from Butler et al. (2022) and we summarize them briefly here.

A single power-law flux distribution is assumed with slope $\alpha = -3.0$, and a uniformly distributed prior over the map is placed for source positions. A minimum flux density is imposed on the primary $250 \mu\text{m}$ band, set to $S_{\text{min}}^{250} = 3$ and $S_{\text{min}}^{250} = 5 \text{ mJy}$ for low-/high-instrument-noise configurations respectively. We find the results in this work are relatively insensitive to the specific choice of S_{min} , though for rigorous characterization of subthreshold number counts, S_{min} can be varied as a hyperparameter. The mean additive normalization of each SPIRE map, denoted as B_0 , has no physical significance (i.e., SPIRE is not absolutely calibrated), and so we place a uniform prior on this component. Likewise, we place uninformative priors on the amplitudes of the Fourier component template amplitudes, though a power spectrum prior is used in the initial set of Fourier component marginalization steps taken during burn-in (see Appendix A).

One difference made in this work is the choice of prior on the number of sources, $\pi(N_{\text{src}})$, which counteracts the effect of overfitting due to additional parameters. As explained in Section 3.1 of Daylan et al. (2017), PCAT uses two pairs of transdimensional MCMC proposals to explore the full catalog space. In the first pair, a number of sources are chosen to either add (“births”) with fluxes drawn from a flux+color prior, or to

remove sources at random (“deaths”). The second pair of proposals either split individual model sources into several components or merge pairs of sources. As the number of degrees of freedom (dofs) approaches infinity, the expected improvement in the log-likelihood approaches one-half for each additional parameter (Wilks 1938), implying an exponential prior on the number of sources:

$$\log \frac{\pi(N+1)}{\pi(N)} = -\frac{1}{2}\rho(2+n_f). \quad (7)$$

As the number of dofs approaches infinity, the average improvement in the log-likelihood approaches one-half per dof (Wilks 1938). For heavily confused observations the number of model parameters becomes non-negligible compared to the dimension of the data, which has the effect of increasing the amount of overfitting ($\langle \Delta \log \mathcal{L} \rangle > 0.5$ per dof). We use the scaling parameter ρ to modify the parsimony prior. As such, ρ may be derived a priori by computing the ratio of $\langle \Delta \log \mathcal{L} \rangle$ for some fixed source number density above S_{\min} (plus any additional model parameters), with the same expectation in the “sparse” limit. We derive an expression for $\langle \Delta \log \mathcal{L} \rangle$ in the nonasymptotic limit in Appendix B. For our runs, we use $\rho = 1.5$, which is slightly more aggressive compared to that assuming single-band source number densities from B12 and $S_{\min} = 3$ mJy, which suggests $\rho \sim 1.35$ (see Figure 13).

The chains used throughout the paper were run for 4000 thinned samples each, where within PCAT-DE one thinned sample = 10^3 samples. We confirm that the chains converge within the first 2000 thinned samples, and we use the last 1000 samples from each chain for our results.

4.2. Number of Fourier Components

The order of the Fourier component model is a hyperparameter that is chosen before running PCAT-DE. Including too few Fourier components may lead to a residual diffuse fluctuation signal, while including too many components becomes computationally inefficient and makes the model more susceptible to overfitting. In principle, a fully transdimensional approach might infer the effective order of the Fourier component model. However, constructing efficient proposals that sample across Fourier component models of varying order is nontrivial, because the number of parameters in the Fourier component model scales as k_{\max}^2 , meaning a penalization based on the number of additional parameters becomes prohibitive.

The hyperparameter N_{FC} is optimized by fitting several Fourier component models to mock data and comparing summary statistics as a function of N_{FC} . In general, we find that setting N_{FC} such that the highest-angular-frequency Fourier component has k_{θ}^{\max} that is twice the cirrus-CIB power spectrum crossover scale (i.e., $k_{\theta}^{\max} = 2k_{\theta}'$, where $P(k_{\theta}')_{\text{diffuse}} = P(k_{\theta}')_{\text{shot}}$) leads to an effective, parsimonious model, in the sense that the recovered diagnostics do not improve significantly by going to higher N_{FC} . A range of Fourier component models are tested, ranging from $N_{\text{FC}} = 2$ to $N_{\text{FC}} = 15$, which correspond to truncation scales $\theta_{\min}^{\text{FC}}$ of $5'$ and $40''$ arcseconds, respectively. For each cirrus amplitude case, the CIB realization is fixed to isolate trends due to varying N_{FC} .

Figure 3 shows the residual rms averaged over pixels for both the recovered cirrus and CIB components. This statistic speaks to the general model reconstruction performance and how it changes with Fourier order, and by inspecting componentwise residuals we can assess the point beyond which additional Fourier components do not improve model performance. Within statistical errors, the cirrus residual level converges as the order of the Fourier component model increases. While the rms is an incomplete measure for how well the data can constrain diffuse signals with arbitrary Fourier structure, the relative rms contribution from degeneracies with point sources can be estimated. In particular, the maximum a posteriori (MAP) solution from each set of Fourier component templates is computed with respect to the same cirrus realizations, including instrument noise but without injected CIB. The errors from this simplified configuration are shown with the dashed lines in the top panel of Figure 3. These results suggest the rms error due to the model fit quality and instrument noise is subdominant to confusion noise for CIB-dominated maps, while for maps with more fluctuation power from cirrus (e.g., $4\times$ -Planck and $8\times$ -Planck), the error from each component is roughly equal.

Unlike the cirrus reconstruction, which plateaus at larger N_{FC} , the CIB residual rms levels increase by 60% and 20% relative to the minima of the $1\times$ -Planck and $2\times$ -Planck cases, respectively. The goodness of fit does not change significantly across the same range, suggesting the increased component residual rms is not due to an overall lack of convergence. A large portion of the CIB is undetected, due to steeply falling number counts, so in the presence of a parsimony prior on point sources and the absence of a power spectrum prior on the diffuse model, the Fourier components can preferentially absorb fluctuations from the CIB signal.

In addition to reducing rms fluctuations in recovered signals, higher-order Fourier component models reduce skewness in the componentwise residuals. Figure 4 shows the distribution of componentwise residuals over pixels for the most contaminated case ($8\times$ -Planck). As N_{FC} is increased from 2 to 15, the skewness in both the CIB and cirrus residuals is reduced considerably. The anticorrelation between the CIB and cirrus residual one-point distributions reflects the oversubtraction of the CIB by spurious point sources, which may compensate for mismodeled diffuse signal when the order of the Fourier model is decreased.

For the remaining results, N_{FC} is fixed at each Planck cirrus level using the prescription described in Section 4.2. This corresponds to $N_{\text{FC}} = 5, 7, 11,$ and 15 for $1\times$ -, $2\times$ -, $4\times$ -, and $8\times$ -Planck, respectively.

4.3. Component Separation

Figure 5 shows the input and recovered component maps for a mock CIB observation with cirrus at the $8\times$ -Planck level, i.e., the top 5% of the most contaminated Planck clean-sky observations. While the residual of the full model is consistent with noise, inspection of the individual component residuals shows errors in the recovered CIB and cirrus components. Notably, these errors are anticorrelated. The primary failure mode occurs when spurious sources compensate for spatially coincident residual diffuse signal, rather than when diffuse signals conceal true point sources (this is supported by the results of Section 5). Likewise, we find that the CIB residual is weakly correlated with the cirrus spatial curvature (Pearson

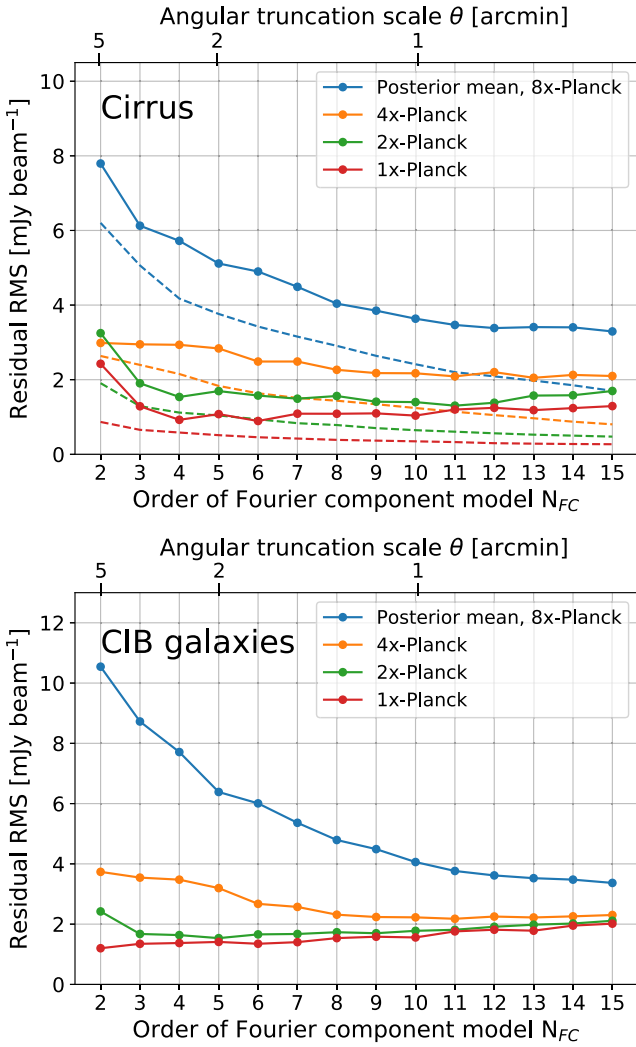


Figure 3. Reconstruction accuracy for diffuse cirrus (top) and CIB galaxies (bottom) in mock SPIRE observations, as a function of Fourier component truncation scale (ranging from $\theta_{FC}^{\min} = 5'$ to $40''$). The different colors show how the results change upon increasing the level of cirrus signal. The top axes indicate the approximate angular truncation scale corresponding to different N_{FC} . The dashed lines in the top figure show the residual rms levels obtained from fitting Fourier components to the same cirrus realizations but without CIB injected.

correlation coefficient $\rho = +0.2$) and the positions of spurious point sources. Regions of high curvature imply the presence of higher-angular-frequency modes that may be difficult to capture with the truncated Fourier series, so this correlation is expected.

4.4. Componentwise Power Spectrum Recovery

The PCAT-DE model separates signals effectively in both map space and Fourier space. Figure 6 shows the recovered component power spectra of observations using a fixed CIB realization and four Planck cirrus realizations of increasing fluctuation power. The power spectra are computed from the model images, with a Hanning window to mitigate spurious fluctuations sourced by the map boundaries. In nearly all cases, the power spectra of both components are reliably recovered, while for the $4\times$ -Planck and $8\times$ -Planck cases the recovered CIB has a slight positive bias, which can be attributed to leakage from the much brighter cirrus signals. While cirrus-

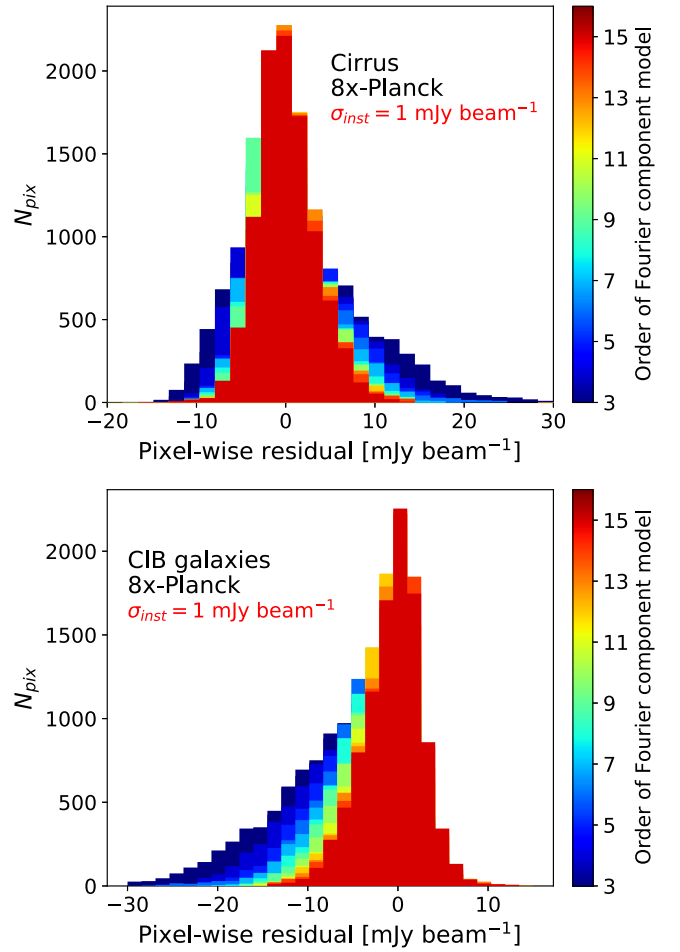


Figure 4. Histograms of pixelwise residuals between cirrus (top), CIB galaxies (bottom), and the posterior mean of each respective model component.

dominated observations have more false detections and faint-end flux boosting on average (see Section 5), the residual fluctuation power of the CIB signal remains relatively small. This is reasonable in the limit where false detections are unclustered, i.e., they contribute to the mean normalization of the component model, but not to its fluctuations. Low-level systematic biases in component separation with PCAT-DE may be more important to quantify in studies of the large-scale ($\theta > 10'$) CIB clustering power spectrum, where linear clustering and diffuse emission are spatially degenerate.

4.5. Computational Requirements

Forward modeling approaches like PCAT are computationally demanding but tractable for targeted science fields. Proposals that perturb the template and mean normalization components are the dominant computational expense for PCAT-DE, because they involve evaluating the delta log-likelihood over the full image or set of images, with an execution time that scales with the total number of pixels. For a fixed effective sample size (ESS), the time to obtain an independent sample from the chain naively scales as the square of the number of parameters when using Metropolis-Hastings proposals (Daylan et al. 2017). While this is mitigated for the point sources by evaluating the likelihood of point-source proposals in smaller image patches, this is not possible with the Fourier templates, which are defined over the full region of

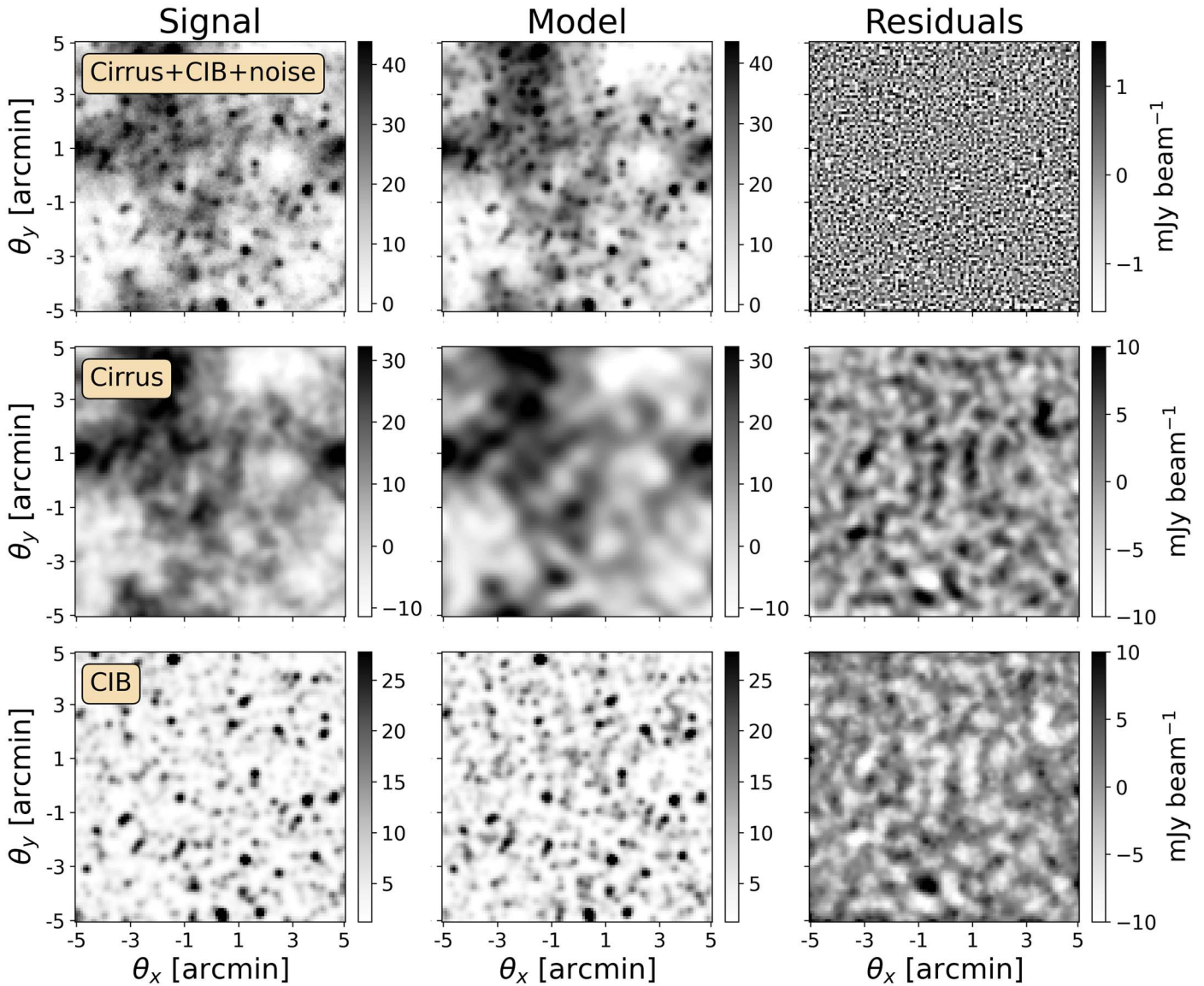


Figure 5. Component separation results for a mock CIB realization with injected cirrus dust (at the highest level, $8\times$ -Planck) and SPIRE-like noise ($\sigma_{\text{inst}} = 1 \text{ mJy beam}^{-1}$). The columns show the data signal (left), the median PCAT model (middle), and the corresponding residuals (right).

interest. In the absence of mean background and template-based proposals, PCAT takes ~ 30 minutes in wall-clock time to fit a 100×100 pixel SPIRE image on a MacBook Pro with a 2.2 GHz Intel Core i7 processor using the Intel Math Kernel Library (MKL) and ~ 1 CPU hour without the MKL library. With mean background and template-based proposals included in the Metropolis–Hastings sampling, the wall-clock time increases by a factor of ~ 2 .

Despite the computational challenge, it should be possible to make the sampling algorithm more efficient. One option to speed up the performance is by executing marginalization steps, as are currently used during burn-in within PCAT-DE to accelerate χ^2 minimization, and intermittently during sampling, making PCAT-DE similar to a collapsed Gibbs sampler (Liu 1994). The marginalization step integrates out the coefficients of the Fourier component model, while fixing the remaining parameters at a given sample, and it may be more appropriate for analyses where the diffuse component coefficients are nuisance parameters. More generally, efficient

proposals and sampling schemes can reduce the run time required for chain convergence and a sufficient ESS.

5. Point-source Detection and Population Inference in the Presence of Diffuse Emission

By incorporating a Fourier component model into source detection and deblending, PCAT can recover sources obscured by negative diffuse signal fluctuations (relative to some mean normalization of the image) and reduce the number of false detections and boosted sources. In this section, we test PCAT-DE on a set of CIB realizations with known positions/flux densities, from which we can examine the collection of detected sources and their properties as the level of injected cirrus is gradually increased.

PCAT requires precise control over systematic effects in the observed data in order to constrain point-source populations without incurring substantial errors. This is a consequence of the general fact that when a finite-mixture model is misspecified (e.g., when it does not fully describe the data),

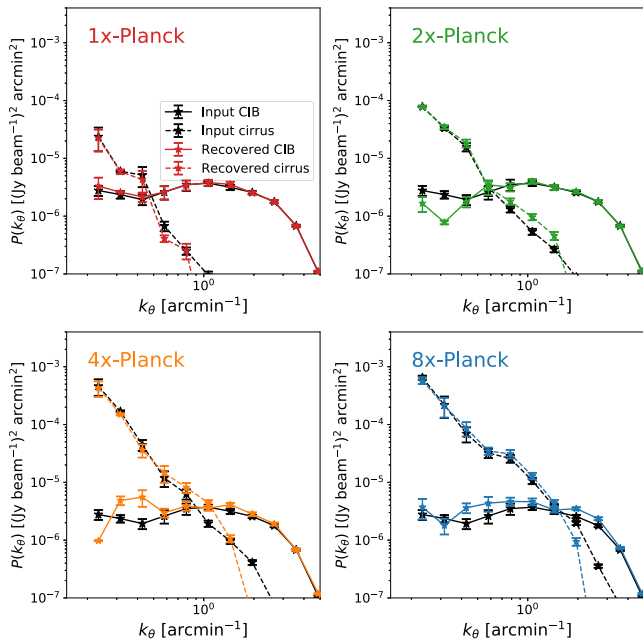


Figure 6. Comparison of input and recovered CIB (solid) and cirrus (dashed) power spectra for four cirrus realizations of increasing power. The recovered cirrus is represented by the Fourier component model image, while the CIB comes from the point-source model. The per-pixel instrument noise is 10^{-3} Jy beam $^{-1}$, corresponding to a power spectrum amplitude of 10^{-8} Jy 2 beam $^{-2}$ arcmin 2 .

the posterior on the number of components can diverge (Cai et al. 2020). Within PCAT, a minimum flux density is chosen for computational convenience, but also represents an instance of model misspecification, i.e., the true number counts extend below S_{\min} . Diffuse signals are relevant in this context as well—while they may not even be visible in an image, neglecting them when modeling observations can lead to biases on downstream measurements that rely on catalogs as starting points. These effects can be seen in Figure 7, where catalog ensembles from PCAT-DE are compared with ground-truth catalogs for three different runs. The middle panel shows how running PCAT on low-level, unmodeled cirrus leads to several spurious point sources clustered on the scale of the beam. The spurious sources are correlated with the gradient of the cirrus emission (along the θ_x direction), residing in regions where the mean normalization underestimates the diffuse component. The positions of spurious sources tend to trace faint underlying sources with flux densities that are below S_{\min} (the green points), suggesting that in this case small errors in the diffuse model primarily flux boost existing sources rather than generate completely fictitious sources. When the mean normalization overestimates the diffuse component, the modeled source flux densities will tend to bias low, which may also lead to a degradation of the catalog completeness for sources near S_{\min} . The inclusion of a simple Fourier component model ameliorates the effects of foreground contamination significantly, with the recovered catalog posterior (right panel) nearly identical to the cirrus-free case (left panel). The stacked samples shown in blue can be converted into a “condensed catalog” using an iterative crossmatching procedure (see Portillo et al. 2017 for an outlined procedure), with posteriors obtained from the collection of samples near each source. One can see visually in Figure 7 that brighter sources have more compact stacked samples, i.e., the posteriors on positions are well constrained.

On the other hand, low-significance sources and/or fictitious sources sourced by cirrus systematics are “fuzzier,” corresponding to posteriors that are much less constrained and that deviate from idealized Gaussian uncertainties¹¹

There is an intrinsic labeling degeneracy in PCAT, due to the fact that the number of sources is not fixed. As such, we compute metrics related to catalog completeness and reliability as expectations over the catalog ensemble returned by PCAT. We calculate the completeness of each catalog sample by finding the closest model source to each true source within $6''$ (one-third of the SPIRE 250 μm beam FWHM) without replacement (i.e., the same PCAT source cannot be matched to several true catalog sources). Any PCAT source that has no true counterpart above S_{\min} after this crossmatching procedure is classified as spurious. While a more stringent crossmatching procedure might include a match on flux density or log-fluxes, we are primarily interested in the trends of these statistics with varying Fourier order. The level of blending for SPIRE sources further complicates the interpretation of more detailed cross-matches (see Section 5.1 for more details).

Figure 8 shows the completeness and false discovery as a function of flux density, evaluated for our ensemble of CIB mocks at two noise levels (1 and 6 mJy beam $^{-1}$). As the level of cirrus contamination increases, fainter sources become suppressed or entirely subsumed by diffuse signal fluctuations, leading to a mild degradation in source recovery. In contrast, the 90% source reliability thresholds degrade from 8 (25) mJy for the $1\times$ -Planck low- (high-) noise configurations to 16 (35) mJy for $8\times$ -Planck. Spurious sources are included when the improvement in the log-likelihood from modeling the residual diffuse emission with a spurious point source is greater than the penalty from adding parameters to the model and any other priors. The false discovery rate is also sensitive to the minimum source flux density permitted by the model.

5.1. Predicting Source Blending

Measuring the number density of sources as a function of flux density is a core task in astronomy. To recover correct number counts, one makes corrections for all effects that lead to observed counts, including catalog completeness, survey selections, astrophysical uncertainties, etc. One consideration is flux boosting, an Eddington-like bias where symmetric flux uncertainties added to sources drawn from a steeply falling LF source an asymmetric scatter in the observed counts. In the context of submillimeter analyses, flux boosting can also be sourced by faint blended neighbors. Blending effects in single-dish submillimeter observations are often so severe that number counts are estimated from the one-point distributions of the maps (P(D) analysis; see Glenn et al. 2010; Vernstrom et al. 2014), rather than from catalogs with individual sources. As a result, the method is limited to fields that are free of contamination from other components that would otherwise contribute to the skewness of the one-point function.

The flux boosting induced by source confusion can be well approximated through Bayesian model comparison. Using mock catalogs, we predict the blended catalog by identifying potential blends and evaluating the delta log-likelihood between two- and one-source models. This is an approximation

¹¹ The departure from idealized uncertainties (assuming well-isolated point sources and perfect background subtraction) is quantified with the *degradation factor* (see Appendix C of Portillo et al. 2017). That work demonstrated that the degradation factors for positions and fluxes are highly correlated.

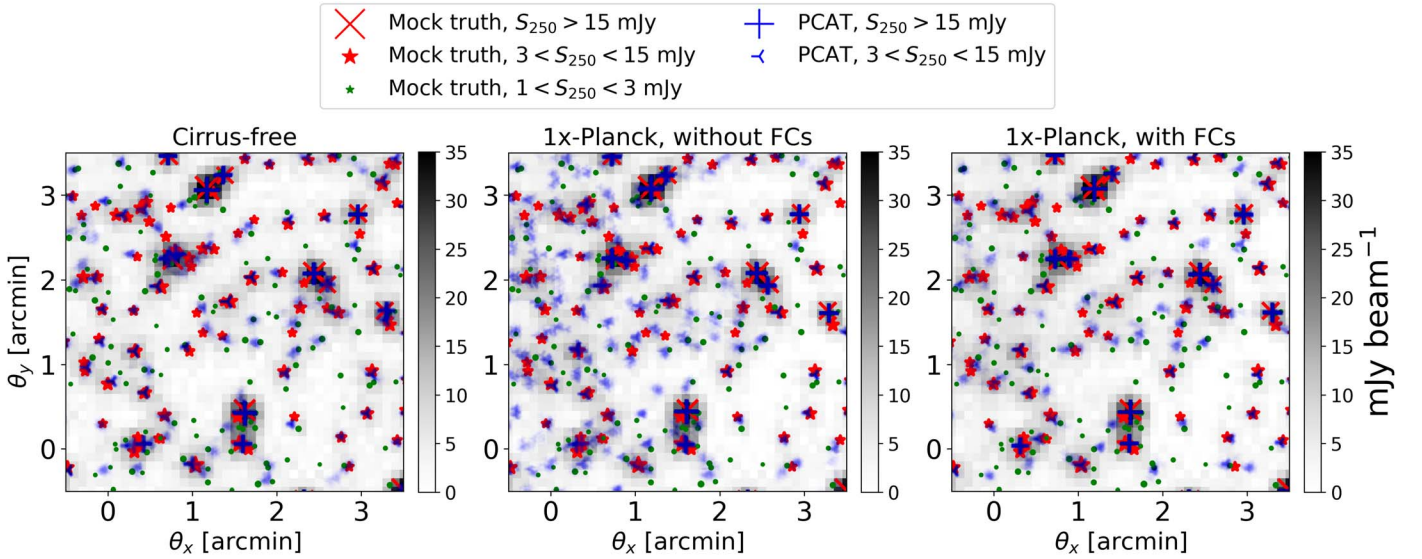


Figure 7. Comparison of recovered catalog ensembles for a fixed 250 μm SPIRE CIB observation in the cirrus-free and 1 \times -Planck cases, with and without the use of Fourier components. In each panel, the red symbols indicate the true positions of the input catalog down to $S_{250} = 3$ mJy, while the blue symbols show 200 stacked samples that are uniformly distributed across the last 1000 posterior samples. “Fuzzier” regions reflect the posterior uncertainty recovered in sources, typically on the faint end. The left panel shows PCAT run on CIB and instrument noise realizations, with no Fourier component model. The middle and right panels show results based on the same CIB realization, but with additional synthetic cirrus dust drawn at the 1 \times -Planck cirrus level. While PCAT infers several spurious model sources in the absence of a diffuse signal model (middle), the inclusion of the Fourier component model leads to a recovered catalog ensemble nearly identical to the cirrus-free case (right).

to the full transdimensional inference performed with PCAT. For a given blend of two (or more) sources, PCAT estimates the relative Bayesian evidence between models with different N_{src} . If the likelihood does not improve significantly for observations with high underlying source multiplicity, PCAT will favor a simpler model to describe the observed signal.

We predict the level of flux boosting for a given catalog in a probabilistic manner, evaluating the delta log-likelihood of two- and one-source models fit to underlying two-source configurations that might be blended by PCAT. This approach assumes prior information about the number counts of the underlying distribution, but marginalizing over uncertainties of the faint-end LF is straightforward with this method if synthetic catalogs are available.

The delta log-likelihood for each candidate blend is calculated as follows: the best-fit one-source model position is approximated to be at the position where a PSF has the maximum overlap integral with the sum of two PSFs with positions and amplitudes corresponding to the two catalog sources. This approximation is exact in the infinite signal-to-noise ratio limit. The maximum overlap integral position is on the line connecting the two sources, and its distance along this line depends only on the two sources’ flux ratio and separation. Then we approximate the best-fit two-source model positions with the catalog positions and calculate the expected delta log-likelihood between the one-source model and the two-source model. Again, the best-fit two-source model positions are equal to the true (catalog) positions in the infinite signal-to-noise ratio limit. While the maximum overlap position does not depend on the noise level, the expected delta log-likelihood does. Our validation of these two approximations using simulated images of pairs of point sources will be presented in a future manuscript.

The delta log-likelihood $\Delta \log \mathcal{L}$ for a given pair of sources can be combined with the relative parsimony prior, $\log(\pi(N=2)) - \log(\pi(N=1))$ (using Equation (7) with

$\rho = 1.5$), to obtain a delta log-posterior between models, $\Delta \log \mathcal{P} = \Delta \log \mathcal{L} + \Delta \log \pi$. For this calculation, we ignore differences in posterior volume, though these differences are used in calculating acceptance probabilities within PCAT-DE. Let us assume that $P(N=2) + P(N=1) \approx 1$, i.e., $P(N > 2) \ll 1$. Then $\Delta \log \mathcal{P}$ is related to the deblending probability by the following:

$$p(N=2) \approx \frac{1}{1 + e^{-\Delta \log \mathcal{P}}}. \quad (8)$$

The algorithm iteratively evaluates blends, starting with the brightest source and finding the brightest neighbor within one FWHM of the source position. If there is no neighbor in the vicinity, the recovered flux is assumed to be the true flux, on average. If there is a neighbor, the delta log-likelihood of the two-source configuration is used to simulate blending by making a draw on a Bernoulli distribution with parameter p ($N=2$). If the draw results in a blend, the best-fit flux/position of the one-source model is added to the catalog and both original sources are removed. Once the full catalog has been processed in this way, the number counts are recomputed. These “recovered” catalogs should more closely resemble the recovered flux distribution using a Bayesian approach like PCAT. These recovered catalogs encode an approximation of the posterior distribution. Through our simulated blending procedure, we find the predicted number counts are consistent with those recovered using PCAT-DE for a range of flux densities. Figure 9 shows the input and recovered flux distributions for different levels of cirrus contamination. All of the recovered flux distributions show overproduction of intermediate/bright sources relative to the input CIB catalog. Given the LF of submillimeter galaxies and the angular resolution of SPIRE, this behavior is explainable by source blending. For a given blend of two (or more) sources, PCAT

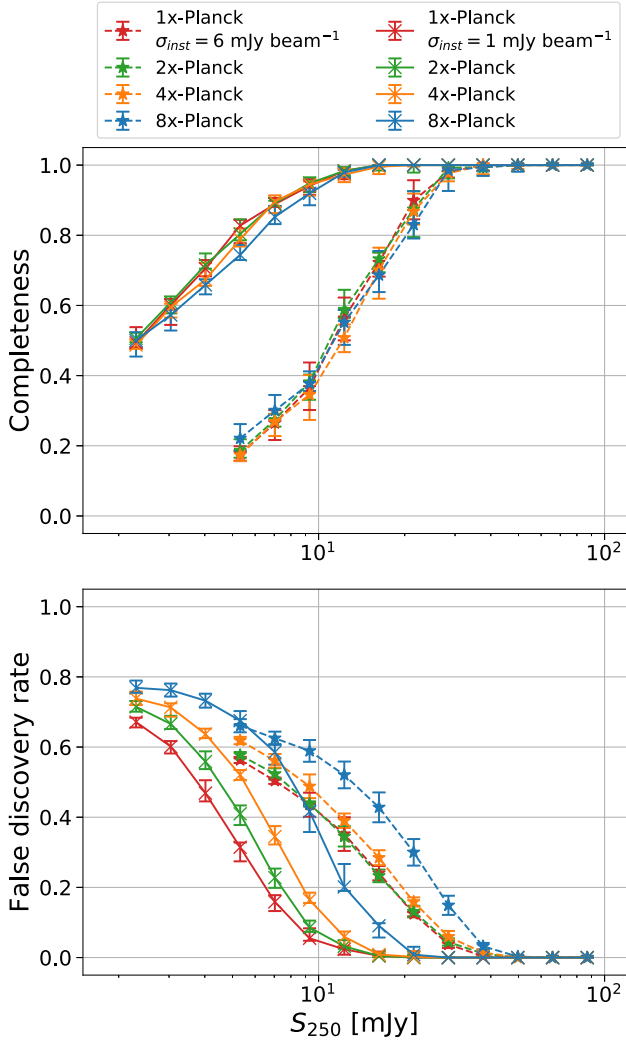


Figure 8. Completeness (top) and false discovery rate (bottom) of PCAT-detected sources for mock CIB data with $\sigma_{inst} = 1 \text{ mJy beam}^{-1}$ (solid) and $\sigma_{inst} = 6 \text{ mJy beam}^{-1}$ (dashed). Each line shows the mean and scatter from 20 mock CIB realizations.

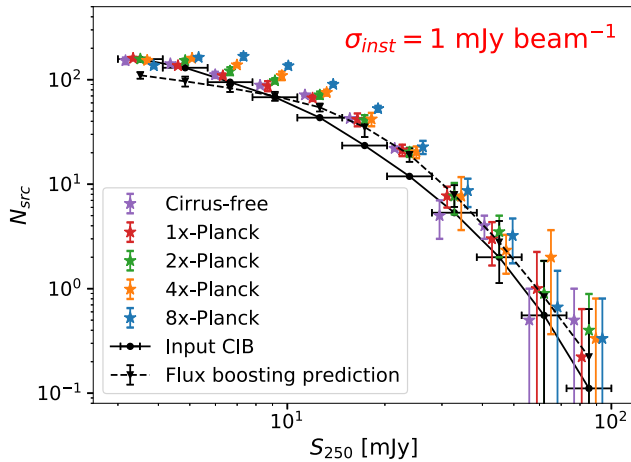


Figure 9. Input and recovered flux distributions for 20 mock CIB realizations with increasing cirrus contamination ($\sim 0.3 \text{ deg}^2$ in total area). The recovered flux distributions are shifted slightly along the x-axis to highlight the dependence on diffuse contamination for each bin. The error bars indicate the scatter across the $10' \times 10'$ CIB realizations. The same set of catalogs is used as input for each cirrus level in order to reduce additional uncertainties due to sample variance.

estimates the relative Bayesian evidence between models with different N_{src} . If the likelihood does not improve significantly for observations with high underlying source multiplicity, PCAT will favor a simpler model to describe the observed signal. This means for some range of source separations there is not enough information in the observed data to properly deblend sources. This is a well-known limitation for analyses of Herschel-SPIRE data, and there are methods in the literature to correct for this mode of flux boosting, both for individual objects and at the population level (Coppin et al. 2005; Crawford et al. 2010). The number count predictions including the effects of blending closely match those obtained with PCAT-DE. The prediction does not take into account the covariance of $N_{src} > 2$ source configurations, nor the prior volume effects associated with source parameters, and so they will be less accurate for fainter source flux densities.

Nonetheless, Figure 9 shows that on the bright end ($S_{250} > 20 \text{ mJy}$), the recovered flux distributions are insensitive to all levels of injected cirrus. For $S_{250} \leq 20 \text{ mJy}$, the recovered number counts become increasingly correlated with the injected cirrus level. While the observed flux boosting of an analysis procedure will depend on the details of its implementation, these results suggest it should be possible to empirically deboost the observed number counts as a function of foreground contamination.

5.2. Sensitivity to N_{FC}

Figure 10 shows the completeness and false discovery rate as a function of flux density for a subset of Fourier component models spanning the same range in N_{FC} as tested in Section 4.1. For the $1 \times$ -Planck mocks, there is little to no dependence on the results from varying N_{FC} , aside from some mild trends at low flux density. This validates the robustness of the Fourier component model even when N_{FC} is larger than necessary. The more severely contaminated $8 \times$ -Planck mocks show similar results for completeness, but a strong dependence of the false discovery rate on N_{FC} . As more Fourier components are fit to the data, fewer spurious sources are favored to absorb residual diffuse signal.

6. Measuring the Extended SZ Effect in Galaxy Clusters

To highlight the ability of PCAT-DE to disentangle components using spatial and spectral information, we apply our model to measure the thermal SZ effect toward massive galaxy clusters. The SZ effect describes the spectral distortion of CMB photons that are inverse-Compton-scattered by electrons comprising hot gas in the intracluster medium (ICM). By measuring the SZ effect signal toward galaxy clusters and its dependence with frequency, one can probe the thermodynamics of the cluster ICM through the thermal SZ effect and its relativistic corrections (Sunyaev & Zeldovich 1972; Wright 1979; Itoh et al. 1998; Chluba et al. 2012) and cluster peculiar velocities through the kinematic SZ effect (e.g., Sheth & Diaferio 2001; Sayers et al. 2016a, 2019). At moderate redshifts, the SZ effect signal typically has an angular extent of several arcminutes, meaning it can be spatially correlated with CIB galaxies, foreground cirrus dust, cluster member galaxies, and potentially thermal dust emission associated with the cluster itself (Erler et al. 2018).

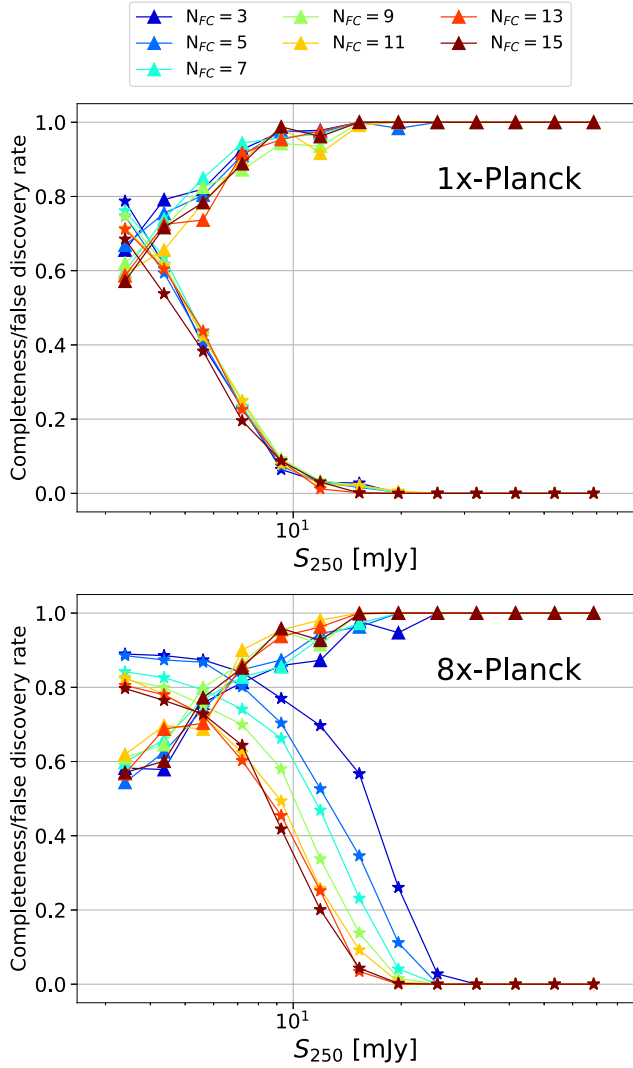


Figure 10. Completeness (triangles) and false discovery rates (stars) as a function of flux density. These are evaluated for a range of Fourier component models indicated by the different colors. The top panel shows the 1×-Planck results, while the bottom shows the same for the 8×-Planck case.

The SZ effect has a SED that rises with wavelength across the SPIRE bandpasses. However, at 500 μm , in which SZ has the largest surface brightness distortion, source blending is also more pronounced due to diffraction-limited optics. In contrast, at shorter wavelengths, where the SZ effect signal is smaller in amplitude, CIB sources are more effectively detected and deblended. This is because the corresponding SPIRE maps have higher angular resolution and because the majority of observed CIB sources have blue spectra. A joint fit across all bands can incorporate these various properties in a consistent manner and can help to reliably extract the SZ component. Color priors can help enhance catalog inferences, with the caveat that unique sources in color space may become more difficult to recover (Feder et al. 2020).

Identifying and separating these components can be challenging, even when instrument noise is low, especially for single-dish submillimeter measurements for which individual point sources are difficult to separate from truly diffuse emission. While ancillary catalogs provide information about the potential positions of submillimeter-detected sources (e.g., deep optical or mid-infrared catalogs), extrapolations of source

SEDs over a large wavelength range are required to predict submillimeter flux densities. The number of external counterparts per SPIRE beam can be as high as 30 per SPIRE beam (Roseboom et al. 2010), meaning some reduction of the external catalog is necessary if the SPIRE data are to be used to constrain the submillimeter flux densities.

Measuring the SZ effect signal from SPIRE observations is a transdimensional task, because the field of faint, confused CIB sources is spatially correlated with the diffuse components. Within the formalism of PCAT, samples are drawn from the marginalized posterior on surface brightness template amplitudes,

$$P(A_{SZ}|D) \propto \int P(C)P(A_{SZ}|D, C)dC, \quad (9)$$

where C denotes the full catalog space.

6.1. Tests on Mock Galaxy Clusters

PCAT-DE is tested in this section on mock data based on a set of clusters previously observed by the Chandra and Herschel observatories. Herschel-SPIRE observed 56 galaxy clusters as part of the HerMES and Herschel Lensing Survey programs, with map depths of $\simeq 1\text{--}2$ mJy beam⁻¹ noise rms (Egami et al. 2010; Oliver et al. 2012), which are subdominant to the SPIRE confusion noise, which is ~ 6 mJy beam⁻¹ at 250 μm (Nguyen et al. 2010). Three clusters from this sample are chosen, with the properties listed in Table 1. For each cluster, we compute an effective angular FWHM, θ_{FWHM} , as the geometric mean of the cluster profile principal axes following image convolution of the cluster gas pressure profile with the SPIRE PSF. Our three clusters vary between 2.1 and 3.6 arcminutes and have temperatures spanning 8.3 and 17.3 keV, allowing us to probe a range of sizes and SZ effect amplitudes.

The procedure for generating mock cluster observations is detailed in Butler et al. (2022). In brief, we use the same B12 CIB model, combined with SPIRE noise realizations unique to the cluster observation with a mean noise rms of ~ 2 mJy beam⁻¹. The SZ signal component is modeled with a set of fixed templates I_b^{SZ} with amplitudes A_b^{SZ} , convolved with the beam

$$\lambda_{ij}^{b,SZ} = \mathcal{P}^b \otimes [A_b^{SZ} I_b^{SZ}(x_i, y_j)]. \quad (10)$$

The same templates are then included in the forward model, i.e., $\lambda_{ij}^{b,SZ}$ is added to λ_{ij}^b in Equation (1). The SZ effect signal is negligible at 250 μm , so we only fit SZ template amplitudes for SPIRE’s 350 and 500 μm bands, denoted A_{PMW}^{SZ} and A_{PLW}^{SZ} , respectively. The morphology of the SZ profile is assumed to follow an elliptical generalized Navarro–Frenk–White profile (Evans & An 2006). This model is fit to ancillary Bolocam 140 GHz data, after which the best-fit profile is extrapolated and regridded to match the SPIRE observations, following the method from Sayers et al. (2019). We place no priors on the SZ template amplitudes nor on their colors, for the purpose of obtaining more data-driven constraints on the SZ effect signal. When one has a complete model for the signal considered (which is not the case for SZ spectral measurements), PCAT-DE is able to incorporate priors across diffuse components across bands. Cirrus realizations are not added to this set of mocks and no Fourier component model is included, though in

Table 1
Input and Recovered SZ Effect Surface Brightnesses for the Three Test Clusters A1835, RXJ 1347, and MACS J0025

Cluster Name	Redshift	T_{PW} (keV)	θ_{FWHM} (arcmin)	Input $A_{\text{PMW}}^{\text{SZ}}$ (MJy sr $^{-1}$)	$\hat{A}_{\text{PMW}}^{\text{SZ}}$ (MAP, 68% C.I.)	Input $A_{\text{PLW}}^{\text{SZ}}$	$\hat{A}_{\text{PLW}}^{\text{SZ}}$
A1835	0.25	8.9	3.6	0.012	$0.007_{-0.014}^{+0.016}$ (0.5 σ)	0.068	$0.063_{-0.026}^{+0.014}$ (2.4 σ)
RX J1347.5-1145	0.45	17.3	2.5	0.033	$0.028_{-0.013}^{+0.016}$ (2.2 σ)	0.130	$0.125_{-0.020}^{+0.014}$ (6.3 σ)
MACS J0025	0.59	8.3	2.1	0.010	$0.004_{-0.018}^{+0.033}$ (0.2 σ)	0.055	$0.050_{-0.018}^{+0.031}$ (2.2 σ)

Note. These clusters have been observed by Chandra, Bolocam, and SPIRE, with pressure-weighted temperatures derived from Chandra observations. Surface brightnesses are in units of MJy sr $^{-1}$ and uncertainties are reported using the highest posterior density intervals. The significance of each SZ detection is also computed assuming Gaussian uncertainties.

reality a small fraction of clusters are observed through lines of sight where significant cirrus contamination exists. Fortunately, cirrus is well constrained by the high-resolution 250 μm data, for which the Fourier components can be fit simultaneously across bands assuming some color prior.

Galaxy clusters gravitationally lens background emission, which has the effect of deflecting and magnifying light from CIB sources. While surface brightness is conserved by lensing, the net effect after removing bright detected sources is a surface brightness deficit near the center of the cluster. This was first measured in Zemcov et al. (2013) in four clusters, and can bias measurement of the SZ signal because the two can be highly spatially degenerate. Bias due to lensing is estimated and corrected in Sayers et al. (2019) and Butler et al. (2022), but for simplicity, the results shown in this work use unlensed mock CIB realizations. In the absence of lensing in the observed data, our forward model is fully specified.

6.2. SZ Results

We test PCAT-DE on mock observations toward galaxy cluster RXJ 1347.5-1145, which has been the subject of numerous SZ studies (Pointecouteau et al. 1999; Komatsu et al. 2001; Kitayama et al. 2004, 2016; Zemcov et al. 2012; Sayers et al. 2016b), including one that uses PCAT-DE (Butler et al. 2022). To derive constraints on cluster properties like the temperature of gas comprising the ICM or the cluster peculiar velocity, multiwavelength data from several instruments (e.g., Bolocam, Planck, Chandra, and the Hubble Space Telescope) are commonly employed, but in this work we focus on surface brightness measurements from SPIRE data alone.

6.2.1. Convergence of SZ Parameters

To validate that the SZ template amplitude parameters are converged, we compute the Gelman–Rubin statistic, or the potential scale reduction factor (also known as \hat{R}), from several Markov chains run on the same data. 20 MCMC walkers are independently initialized and run on a single mock cluster realization of RXJ 1347.5-1145 for 4000 thinned samples, with the second half of each chain used to compute \hat{R} . We estimate $\hat{R} = 1.07$ and $\hat{R} = 1.08$ for $\hat{A}_{\text{PMW}}^{\text{SZ}}$ and $\hat{A}_{\text{PLW}}^{\text{SZ}}$, respectively, suggesting the chains are well mixed.

6.2.2. Component Separation

An advantage of PCAT-DE for this application is that, by modeling all components simultaneously, one mitigates parts of the SZ signal being apportioned to point sources and vice versa. This can be understood upon visual inspection in Figure 11, where the observed cluster field, the best-fit CIB

model, and the residual between the two are plotted. Even when the input SZ signal has a small signal-to-noise ratio (for example, at 350 μm) or is heavily confused, as seen at 500 μm , PCAT-DE is able to reliably separate the underlying signal from contaminants. Unmodeled point-source emission can be seen as well in the residual maps, meaning confusion noise remains a significant systematic in the surface brightness measurement.

6.2.3. Sensitivity to Cluster Properties

For each of the three clusters listed in Table 1, the same set of 20 CIB + instrument noise realizations are used to generate mock cluster observations, after which samples from each set of 20 chains run on the data are aggregated and plotted in Figure 12. Computing this full distribution allows us to quantify the systematic uncertainty associated with the CIB, identify any consistent biases, and assess the sensitivity of our results to the details of the cluster itself.

While source confusion can have large effects on the recovered SZ effect signal, the recovered surface brightness estimates are fairly unbiased over several CIB realizations. The estimated MAP values and 68% credible intervals for $A_{\text{PMW}}^{\text{SZ}}$ and $A_{\text{PLW}}^{\text{SZ}}$ are compared with the input surface brightnesses in Table 1. The mean bias is $\lesssim -0.3\sigma$ for both $A_{\text{PMW}}^{\text{SZ}}$ and $A_{\text{PLW}}^{\text{SZ}}$, and this bias is consistent across our three clusters. This implies it is primarily correlated with the common CIB mocks used to make each set of cluster observations. In general, the derived uncertainties do not vary significantly from cluster to cluster, but a more thorough investigation of the uncertainties from a larger sample of clusters may reveal trends with respect to gas temperature, angular extent, etc. We find that MACS J0025, the cluster with lowest gas temperature (i.e., the smallest SZ distortion) and smallest angular extent, has larger uncertainties by $\sim 40\%$ and 20% for PMW and PLW, respectively.

7. Summary and Outlook

In this work, we have considered a variety of inference tasks that rely on the effective separation of pointlike and diffuse signals. This is done by extending the forward modeling framework of PCAT with a flexible, template-based model for diffuse signals, which results in the efficient separation of CIB emission from diffuse Galactic cirrus. Our point-source completeness limits (Section 5) are robust to additional diffuse emission, degrading by less than $\sim 20\%$ at nearly all flux densities and both noise levels. The purity degrades by a larger amount in the presence of prominent diffuse signals, and we find this degradation correlates with the local signal curvature. At the population level, PCAT-DE enables the robust recovery of number counts across a range of cirrus foreground levels.

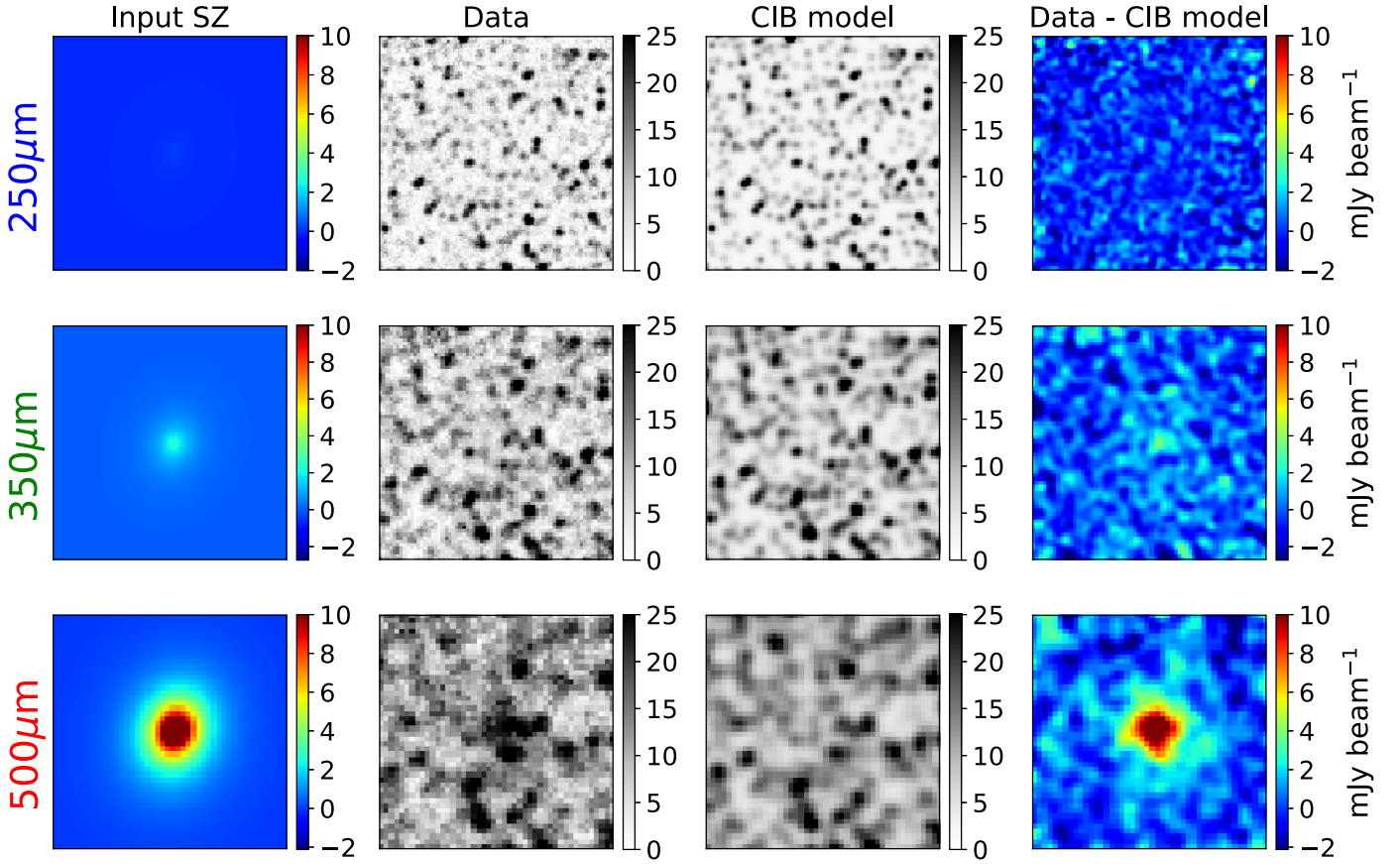


Figure 11. $10' \times 10'$ cirrus-free mock SPIRE observations toward galaxy cluster RX J1347.5-1145. The columns show (going from left to right) the input SZ effect signal, observed maps, median CIB model, and difference between the observed maps and CIB model. By fitting both SZ and CIB simultaneously, we can account for the presence of submillimeter point sources without overfitting the underlying SZ signal. The maps in the rightmost column are shown for visualization purposes only.

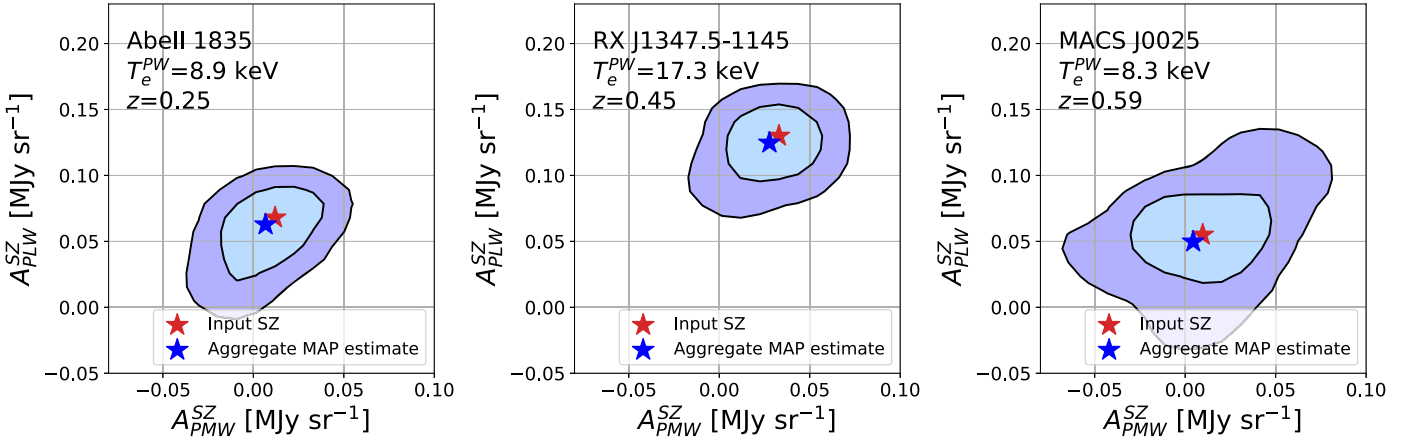


Figure 12. Recovered posteriors on A_b^{SZ} surface brightness parameters for clusters of varying redshift/angular extent. The cluster realizations corresponding to these results are cirrus-free, and the input SZ surface values for each cluster (marked by the red stars) are determined from a combined analysis of Chandra and Bolocam data. 1σ and 2σ density contours are plotted for each cluster.

For our low-noise ($\sigma_{\text{inst}} = 1 \text{ mJy beam}^{-1}$) case, we obtain consistent flux density number counts across all levels of injected cirrus for $S_{250} > 25 \text{ mJy}$, and the counts in cases up to $4\times$ -Planck are consistent for $S_{250} > 12 \text{ mJy}$.

In our second application of PCAT-DE, we demonstrate that the faint, spatially extended SZ effect signal can be recovered in the presence of CIB galaxy contamination along the same line of sight. Averaged over several mock CIB realizations, the

mean recovered SZ surface brightnesses are unbiased at the $\lesssim 0.3\sigma$ level, with uncertainties dominated by confusion noise. PCAT-DE was recently used to measure the gas temperature of the galaxy cluster RX J1347.5-1145, to which the SZ spectrum is sensitive through relativistic effects (Butler et al. 2022). That work found a temperature of $\langle T_{\text{sz}} \rangle_{2500} = 22.4_{-12.0}^{+10.6} \text{ keV}$, consistent with the X-ray-measured $\langle T_{x,\text{pw}} \rangle_{2500} = 17.3 \text{ keV}$. The results from this work further demonstrate that similar

measurements of the SZ effect should be robust for a collection of different cluster profiles and gas temperatures. When a spatial template for the diffuse signal is available, it can be easily incorporated into the forward model, as shown in Section 6, but more detailed signal parameterizations may be used when appropriate.

Surveys with strict requirements on photometric accuracy may place stringent cuts on sources based on the estimated signal-to-noise ratio or excise regions with pronounced diffuse contamination. In addition to a loss of information, the situation can be especially problematic when uncertainties due to diffuse signals are underestimated. We anticipate that the tools from this work can expand the sample size of “usable” sources in astronomical catalogs when both source confusion and diffuse foregrounds are prominent.

While the Fourier component model performs well for the examples considered in this work, there are limitations on the types of signals it can reconstruct effectively. In particular, structures comparable to or smaller than the PSF FWHM may be more difficult to model with Fourier components, as suggested by results on cirrus-dominated maps (see Figure 3). More flexible generative models may be able to capture nonlinear structures such as filaments, as demonstrated on CMB data to model foregrounds (Régaldo-Saint Blancard et al. 2023; Thorne et al. 2021). Under the assumption that the color of the diffuse signal component is constant over a given field of view, it is effective to model the Fourier components in several bands with simple linear scaling factors, incorporating color prior information when appropriate. Position-dependent color variations may be non-negligible for some observations, which can be addressed either by processing smaller regions with fixed color or by incorporating a model for color variations. The formalism of Bayesian hierarchical modeling permits for more detailed extensions of the forward model, in a way that reflects an appropriate level of knowledge. By the same token, one should always be careful in characterizing the effect of priors, both explicitly specified and those implicit to the method (Schmidt et al. 2020), on a given inference task.

PCAT-DE is publicly available on GitHub at <https://github.com/RichardFeder/pcat-de>, with corresponding documentation¹² and examples demonstrating applications from this work.

Acknowledgments

This work was supported by NASA’s Astrophysical Data Analysis Program, Award 80NSSC19K1018. This work was also supported by an LSSTC Catalyst Fellowship awarded by LSST Corporation to T.D., with funding from the John Templeton Foundation grant ID 62192.

The scientific results reported in this article were based on data obtained from the HerMES project (<http://hermes.sussex.ac.uk/>), a Herschel Key Programme utilizing Guaranteed Time from the SPIRE instrument team, ESAC scientists, and a mission scientist. The HerMES data were accessed through the Herschel Database in Marseille (<http://hedam.lam.fr>), operated by CeSAM and hosted by the Laboratoire d’Astrophysique de Marseille. The Bolocam data are publicly available at https://irsa.ipac.caltech.edu/data/Planck/release_2/ancillary-data/bolocam/bolocam.html.

The authors acknowledge Research Computing at the Rochester Institute of Technology for providing computational

resources and support that have contributed to the research results reported in this publication.

Facilities: Herschel, Caltech Submillimeter Observatory (CSO), NASA Chandra X-ray Observatory (CXO), Planck.

Software: corner, matplotlib, numpy, Lion.

Appendix A Fourier Component Marginalization

To accelerate the burn-in phase of sampling, we apply a series of linear marginalization steps for the set of Fourier component templates, the number of which are fixed a priori. Let \mathcal{F}_{ij} be the i th pixel of the j th Fourier component template, and let β_j be the j th Fourier component’s amplitude. Then $\mathcal{F}\beta$ is a column vector with the flux in each pixel from each Fourier component. Let Σ be a diagonal matrix, where Σ_{ii} is the variance in pixel i and K the corresponding data vector. Then the β that minimizes the chi-squared statistic $\chi^2(\beta)$ is given by the Moore–Penrose inverse:

$$\hat{\beta} = (\mathcal{F}^T \Sigma^{-1} \mathcal{F})^{-1} \mathcal{F}^T \Sigma^{-1} K. \quad (\text{A1})$$

While the linear inversion from Equation (A1) minimizes χ^2 with respect to the data, in practice the Fourier component parameters are driven to local minima that are difficult to leave in the sampling phase. In addition, because the marginalization procedure is only applied to Fourier component templates, the MAP estimates obtained at the beginning of burn-in are conditioned on an unconverged point-source model. When used at the beginning of the sampler, to obtain an initial guess of the diffuse model, the MAP estimate is biased due to unmodeled point sources. To prevent divergence of the Moore–Penrose inverse, the solution is regularized by imposing a prior on the Fourier component coefficients. This is done through ridge regression, which penalizes the loss function with the ℓ_2 norm of the component amplitudes:

$$\hat{\beta}_{\text{ridge}} = (\mathcal{F}^T \Sigma^{-1} \mathcal{F} + \sigma \mathbf{I})^{-1} \mathcal{F}^T \Sigma^{-1} K. \quad (\text{A2})$$

The vector σ acts as a Gaussian prior on the fluctuation amplitude of each component. In PCAT-DE, σ is inversely proportional to the power spectrum of the underlying signal evaluated at the scale of each Fourier component. An iterative scheme is implemented in PCAT-DE, in which the Fourier components are repeatedly fit to the residual of the data and point-source model. The terms in Equation (A2) only need to be computed once up front and then stored for fast evaluation with the residual data vector K . For a 100×100 pixel image fit using a 15th-order Fourier component model, computing $\hat{\beta}$ takes 8 ms per evaluation (with precomputed quantities), accelerating the burn-in MCMC phase of the fitting routine. Extending the marginalization procedure to include source fluxes should be possible, but is left to future work.

Appendix B Parsimony Prior in the Weakly Nonasymptotic Limit

In PCAT, goodness of fit is enforced with a delta log-prior proportional to the number of point-source parameters being added or removed from the model. This reflects the fact that adding parameters to a model will, on average, improve the log-likelihood of a reconstructed signal. When the number of parameters is much smaller than the dimension of the data, the expected improvement in the log-likelihood is one-half per

¹² <https://pcat-de.readthedocs.io>

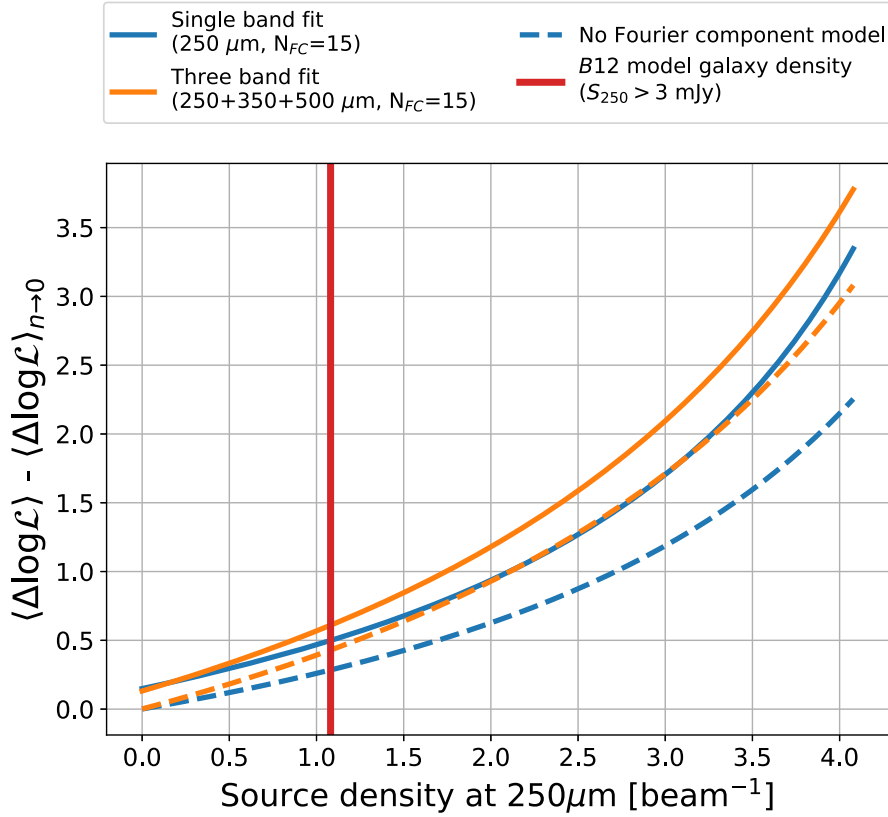


Figure 13. Expected improvement in log-likelihood from adding a point source to the model relative to the sparse limit, $\langle \Delta \log \mathcal{L} \rangle_{n \rightarrow 0}$, as a function of source density. Three parameters are added to the model per source in the single-band case, while each additional band included in a joint fit increases the number of parameters per source by one. The solid and dashed curves are for results with and without Fourier components, respectively.

additional dof. However, this asymptotic result may be ill-suited in the limit of severely confused observations.

We consider the F -statistic (Fisher 1922), which models the significance of a model’s improved fit to data using Snedecor’s F -distribution. Let \mathcal{M}_1 be an initial model with p_1 parameters and $\mathcal{M}_2 = \mathcal{M}_1 \cup \delta$, the union of the initial model, and δ with $p_2 = p_1 + p_\delta$ parameters. The F -statistic is expressed in terms of the chi-squared statistic, the number of parameters associated with each model, and the length of the data vector:

$$F = \frac{\chi_1^2 - \chi_2^2}{p_\delta} \frac{N_{\text{pix}} - p_2}{\chi_2^2}. \quad (\text{B1})$$

Then one can rearrange the terms to relate the delta chi-squared to the F -distribution with p_δ and $N_{\text{pix}} - p_2$ dofs:

$$\Delta \chi^2 \sim \frac{\chi_2^2 p_\delta}{N_{\text{pix}} - p_2} F(p_\delta, N_{\text{pix}} - p_2). \quad (\text{B2})$$

The expectation of this F -distribution is $\frac{N_{\text{pix}} - p_2}{N_{\text{pix}} - p_2 - 2}$, and so the expected delta log-likelihood is

$$\langle \Delta \log \mathcal{L} \rangle = \frac{1}{2} \frac{\chi_2^2 p_\delta}{N_{\text{pix}} - p_2 - 2}. \quad (\text{B3})$$

In Figure 13, we plot Equation (B3) for both single- and three-band cases, assuming good fits to the data (i.e., $\chi_2^2 \approx N_{\text{pix}}$). This is done as a function of source density, ranging from zero (the “sparse limit”) to four sources per beam. For both curves, the larger improvement relative to the asymptotic limit is expected—correlations between model

parameters become non-negligible, meaning the model components can conspire to produce a better reconstruction of the data. These results further suggest that a three-band fit to Herschel-SPIRE data becomes more susceptible to overfitting as the source density increases. This is explained by the poorer angular resolution of the 350 and 500 μm SPIRE maps—while there are additional pixels to constrain the model, these are outnumbered by the additional parameters required to model source fluxes in these bands, leading to more overfitting. This would not be the case if the resolution were the same across maps; for example, in a joint fit of several 250 μm observations of the same field. A simpler point-source parameterization—for example, colors modeled by a single temperature—would reduce the number of additional parameters and thus the amount of overfitting, but care would be needed to manage the transition between sources whose observed spectra trace the blackbody peak (i.e., sources with temperature-driven colors) and sources with more complicated SEDs.

We use the expected number density of SPIRE sources with $S_{250} > 3$ mJy based on the B12 model ($\sim 1 \text{ beam}^{-1}$) to predict the expected delta log-likelihood relative to the sparse limit, which informs our choice of scaling parameter ρ defined in Equation (7). Additional parameters for the Fourier component model modify the parsimony prior as well, as can be seen by comparing the solid and dashed curves in Figure 13.

Appendix C Source/Background Covariance

Quantifying the covariance between point sources and a generic foreground/background helps to inform the proposal

kernel widths for model components within PCAT-DE. As detailed in Portillo et al. (2020), the uncertainty of a uniform background b in the presence of a single point source with flux density f can be calculated by computing the 2×2 Fisher information matrix and its inverse:

$$\begin{aligned} \sigma_b^2(\theta_{\text{ML}}) &= \frac{-\partial_f^2 \ln \mathcal{L}(\theta_{\text{ML}})}{\partial_b^2 \ln \mathcal{L}(\theta_{\text{ML}}) \partial_f^2 \ln \mathcal{L}(\theta_{\text{ML}}) - (\partial_b \partial_f \ln \mathcal{L}(\theta_{\text{ML}}))^2} \\ &= \frac{\sigma^2}{A - A_{\text{psf}}}, \end{aligned} \quad (\text{C1})$$

where $A = nm$ and $A_{\text{psf}} = (\sum_i^{N_{\text{pix}}} p_i(x, y)^2)^{-1}$ is the effective PSF area in pixels. This means that a larger effective PSF increases the corresponding uncertainty on b .

We can generalize the previous result to include n sources in the covariance matrix. Let us assume that our source fluxes are independent of one another, i.e., mixed-derivative terms $\partial_{f_i} \partial_{f_j} \ln \mathcal{L}^* = 0$, where the maximum log-likelihood is abbreviated $\ln \mathcal{L}^* = \ln \mathcal{L}(\theta_{\text{ML}})$. Our Fisher matrix can then be written as the following:

$$\mathcal{F}_{(\{f_i\}_{i=1}^n, b)}(x, y) = \begin{bmatrix} f_1 & f_2 & \cdots & f_n & b \\ \partial_{f_1}^2 \ln \mathcal{L}^* & & & & \partial_{f_1} \partial_b \ln \mathcal{L}^* \\ \partial_{f_2}^2 \ln \mathcal{L}^* & & & & \partial_{f_2} \partial_b \ln \mathcal{L}^* \\ \vdots & & \ddots & & \vdots \\ f_n & & & \partial_{f_n}^2 \ln \mathcal{L}^* & \partial_{f_n} \partial_b \ln \mathcal{L}^* \\ b & \partial_b \partial_{f_1} \ln \mathcal{L}^* & \partial_b \partial_{f_2} \ln \mathcal{L}^* & \dots & \partial_b \partial_{f_n} \ln \mathcal{L}^* & \partial_b^2 \ln \mathcal{L}^* \end{bmatrix}.$$

Assuming the diagonal elements in the above arrowhead matrix are nonzero, the inverse is a rank-one modification of a diagonal matrix:

$$C_{(\{f_i\}_{i=1}^n, b)}(\theta_{\text{ML}}) = [\mathcal{F}_{(\{f_i\}_{i=1}^n, b)}(\theta_{\text{ML}})]^{-1} = \begin{bmatrix} D^{-1} & \\ & 0 \end{bmatrix} + \rho uu^T, \quad (\text{C2})$$

where

$$\begin{aligned} D &= \text{diag}(\partial_{f_1}^2 \ln \mathcal{L}^*, \dots, \partial_{f_n}^2 \ln \mathcal{L}^*); \\ u &= [D^{-1}z \quad -1]^T; \\ z &= [\partial_{f_1} \partial_b \ln \mathcal{L}^* \quad \partial_{f_2} \partial_b \ln \mathcal{L}^* \quad \dots \quad \partial_{f_n} \partial_b \ln \mathcal{L}^*]^T; \\ \rho &= \frac{1}{\alpha - z^T D^{-1}z}; \\ \alpha &= \partial_b^2 \ln \mathcal{L}^*. \end{aligned}$$









Computing the covariance matrix from Equation (C2), and evaluating the diagonal background element, one obtains

$$\sigma_b^2(\theta_{\text{ML}}) = \frac{\sigma^2}{A - nA_{\text{psf}}}. \quad (\text{C3})$$

This follows the intuition that as the number of sources in a given image increases, the effective number of pixels that contribute to constraining the background normalization decreases. In more realistic situations, where the correlations between sources are taken into account, the uncertainties on b

will in general be larger. While we consider uncertainties on the mean normalization, a discussion of the uncertainties on a generic sky model is presented in Appendix C of Portillo et al. (2020).

ORCID iDs

Richard M. Feder  <https://orcid.org/0000-0002-9330-8738>
 Victoria Butler  <https://orcid.org/0000-0002-0941-0407>
 Tansu Daylan  <https://orcid.org/0000-0002-6939-9211>
 Stephen K. N. Portillo  <https://orcid.org/0000-0001-8132-8056>
 Jack Sayers  <https://orcid.org/0000-0002-8213-3784>
 Benjamin J. Vaughan  <https://orcid.org/0000-0002-9813-0270>
 Catalina V. Zamora  <https://orcid.org/0000-0001-9765-0197>
 Michael Zemcov  <https://orcid.org/0000-0001-8253-1451>

References

- Barreiro, R. B. 2009, in *Data Analysis in Cosmology*, ed. V. J. Martínez et al., Vol. 665 (Berlin: Springer), 207
- Baudry, N. J., & Renner, R. 2011, arXiv:1107.0740
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Béthermin, M., Daddi, E., Magdis, G., et al. 2012, *ApJL*, **757**, L23
- Bracco, A., Cooray, A., Veneziani, M., et al. 2011, *MNRAS*, **412**, 1151
- Breivik, K., Connolly, A. J., Ford, K. E. S., et al. 2022, arXiv:2208.02781
- Brewer, B. J., & Donovan, C. P. 2015, *MNRAS*, **448**, 3206
- Brewer, B. J., Foreman-Mackey, D., & Hogg, D. W. 2013, *AJ*, **146**, 7
- Butler, V., Feder, R. M., Daylan, T., et al. 2022, *ApJ*, **932**, 55
- Cai, D., Campbell, T., & Broderick, T. 2020, arXiv:2007.04470
- Casey, C. M., Narayanan, D., & Cooray, A. 2014, *PhR*, **541**, 45
- Cheng, Y.-T., de Putter, R., Chang, T.-C., & Doré, O. 2019, *ApJ*, **877**, 86
- Chluba, J., Nagai, D., Sazonov, S., & Nelson, K. 2012, *MNRAS*, **426**, 510
- Collin, G. H., Rodd, N. L., Erjavec, T., & Perez, K. 2022, *ApJ*, **260**, 29
- Condon, J. J. 1974, *ApJ*, **188**, 279
- Coppin, K., Halpern, M., Scott, D., Borys, C., & Chapman, S. 2005, *MNRAS*, **357**, 1022
- Crawford, T. M., Switzer, E. R., Holzapfel, W. L., et al. 2010, *ApJ*, **718**, 513
- Dabbech, A., Ferrari, C., Mary, D., et al. 2015, *A&A*, **576**, A7
- Daylan, T., Cyr-Racine, F.-Y., Diaz Rivero, A., Dvorkin, C., & Finkbeiner, D. P. 2018, *ApJ*, **854**, 141
- Daylan, T., Portillo, S. K. N., & Finkbeiner, D. 2017, *ApJ*, **839**, 4
- de Bruijne, J. H. J., Allen, M., Azaz, S., et al. 2015, *A&A*, **576**, A74
- Desert, F. X., Boulanger, F., & Puget, J. L. 1990, *A&A*, **237**, 215
- Diolaiti, E., Bendifelli, O., Bonaccini, D., et al. 2000, in *ASP Conf. Ser.* 216, *Astronomical Data Analysis Software and Systems IX*, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco, CA: ASP), 623
- Du, J., Hu, S., Chen, X., & Cao, H. 2022, *MNRAS*, **511**, 3377
- Egami, E., Rex, M., Rawle, T. D., et al. 2010, *A&A*, **518**, L12
- Ellien, A., Slezak, E., Martinet, N., et al. 2021, *A&A*, **649**, A38
- Erlar, J., Basu, K., Chluba, J., & Bertoldi, F. 2018, *MNRAS*, **476**, 3360
- Evans, N. W., & An, J. H. 2006, *PhRvD*, **73**, 023524
- Feder, R. 2023, PCAT-DE: Probabilistic Cataloger in the presence of Diffuse Emission, v0.0.2, Zenodo, doi:10.5281/zenodo.8067131
- Feder, R. M., Portillo, S. K. N., Daylan, T., & Finkbeiner, D. 2020, *AJ*, **159**, 163
- Fisher, R. A. 1922, *RSPTA*, **222**, 309
- Glenn, J., Conley, A., Béthermin, M., et al. 2010, *MNRAS*, **409**, 109
- Green, P. 1995, *Biometrika*, **82**, 711
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, **518**, L3
- Hale, C. L., Rbotham, A. S. G., Davies, L. J. M., et al. 2019, *MNRAS*, **487**, 3971
- Itoh, N., Kohyama, Y., & Nozawa, S. 1998, *ApJ*, **502**, 7
- Jones, D. E., Kashyap, V. L., & van Dyk, D. A. 2015, *ApJ*, **808**, 137
- Kitayama, T., Komatsu, E., Ota, N., et al. 2004, *PASJ*, **56**, 17
- Kitayama, T., Ueda, S., Takakuwa, S., et al. 2016, *PASJ*, **68**, 88
- Komatsu, E., Matsuo, H., Kitayama, T., et al. 2001, *PASJ*, **53**, 57
- Lang, D., & Hogg, D. W. 2020, arXiv:2012.15836
- Liu, J. 1994, *JASA*, **89**, 958
- Liu, R., McAuliffe, J. D., & Regier, J. 2021, arXiv:2102.02409
- Lukic, V., de Gasperin, F., & Brüggén, M. 2019, *Galax*, **8**, 3

- Marshall, P., Anguita, T., Bianco, F. B., et al. 2017, arXiv:1708.04058
- Masias, M., Freixenet, J., Lladó, X., & Peracaula, M. 2012, *MNRAS*, 422, 1674
- Masias, M., Lladó, X., Peracaula, M., & Freixenet, J. 2015, *A&C*, 9, 10
- Melchior, P., Joseph, R., Sanchez, J., MacCrann, N., & Gruen, D. 2021, *NatRP*, 3, 712
- Men'shchikov, A., André, P., Didelon, P., et al. 2012, *A&A*, 542, A81
- Molinari, S., Schisano, E., Faustini, F., et al. 2011, *A&A*, 530, A133
- Nguyen, H. T., Schulz, B., Levenson, L., et al. 2010, *A&A*, 518, L5
- Ofek, E. O., & Zackay, B. 2018, *AJ*, 155, 169
- Oliver, S. J., Bock, J., Altieri, B., et al. 2012, *MNRAS*, 424, 1614
- Planck Collaboration, Abergel, A., Ade, P. A. R., et al. 2014, *A&A*, 571, A11
- Pointecouteau, E., Giard, M., Benoit, A., et al. 1999, *ApJL*, 519, L115
- Portillo, S. K. N., Speagle, J. S., & Finkbeiner, D. P. 2020, *AJ*, 159, 165
- Portillo, S., Lee, B. C. G., Daylan, T., & Finkbeiner, D. P. 2017, *AJ*, 154, 4
- Régalo-Saint Blancard, B., Allys, E., Auclair, C., et al. 2023, *ApJ*, 943, 9
- Robotham, A. S. G., Davies, L. J. M., Driver, S. P., et al. 2018, *MNRAS*, 476, 3137
- Roseboom, I. G., Oliver, S. J., Kunz, M., et al. 2010, *MNRAS*, 409, 48
- Saydjari, A. K., & Finkbeiner, D. P. 2022, *ApJ*, 933, 155
- Saydjari, A. K., Schlafly, E. F., Lang, D., et al. 2023, *ApJ*, 264, 28
- Sayers, J., Golwala, S. R., Mantz, A. B., et al. 2016a, *ApJ*, 832, 26
- Sayers, J., Montaña, A., Mroczkowski, T., et al. 2019, *ApJ*, 880, 45
- Sayers, J., Zemcov, M., Glenn, J., et al. 2016b, *ApJ*, 820, 101
- Schaan, E., & White, M. 2021, *JCAP*, 2021, 067
- Schmidt, S. J., Malz, A. I., Soo, J. Y. H., et al. 2020, *MNRAS*, 499, 1587
- Sheth, R. K., & Diaferio, A. 2001, *MNRAS*, 322, 901
- Stetson, P. 1987, *The Astronomical Society of the Pacific*, 99, 191
- Sunyaev, R. A., & Zeldovich, Y. B. 1972, *CoASP*, 4, 173
- Thorne, B., Knox, L., & Prabhu, K. 2021, *MNRAS*, 504, 2603
- Traficante, A., Calzoletti, L., Veneziani, M., et al. 2011, *MNRAS*, 416, 2932
- Vernstrom, T., Scott, D., Wall, J. V., et al. 2014, *MNRAS*, 440, 2791
- Viero, M. P., Wang, L., Zemcov, M., et al. 2013, *ApJ*, 772, 77
- Wilks, S. S. 1938, *Ann. Math. Stat.*, 9, 60
- Wright, E. L. 1979, *ApJ*, 232, 348
- Zemcov, M., Aguirre, J., Bock, J., et al. 2012, *ApJ*, 749, 114
- Zemcov, M., Blain, A., Cooray, A., et al. 2013, *ApJL*, 769, L31
- Zemcov, M., Smidt, J., Arai, T., et al. 2014, *Sci*, 346, 732
- Zheng, C., Pulido, J., Thorman, P., & Hamann, B. 2015, *MNRAS*, 451, 4445