Dissecting and Reconstructing the Cosmic Infrared Background with Spaceborne Experiments

Thesis by
Richard M. Feder

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
Defended May 24th, 2024

Richard M. Feder
ORCID: 0000-0002-9330-8738

ACKNOWLEDGEMENTS

In physics, we are often warned about what are called “unknown unknowns”. These encapsulate all of the things we do not anticipate that can affect some outcome or result. These often emerge in research and are a sobering reminder that the reality of a situation may be more complicated or different than we know ahead of time, despite our best efforts. I could not be happier, however, for the unknown unknowns that have shaped my time at Caltech the past six years. I never anticipated the breadth of science I would have the opportunity to think about, the variety of collaborators I would get to know, the places I would travel. Graduate school opened my eyes to the world of research in new, impactful ways. For this I owe thanks to many people.

My first acknowledgement goes to my advisor, Jamie Bock. I am grateful for his guidance over the years, for helping me work incrementally through difficult problems, and for his willingness to engage on both familiar and unfamiliar topics. I am inspired by his ability to grapple with broad scientific questions while maintaining an acute, consistent focus on the steps needed to find answers. Thank you for making me more of an experimentalist and more of a scientist.

I’ve been blessed to have several mentors throughout graduate school. I owe a debt of gratitude to Olivier Doré for his thoughtful mentorship, both academically and professionally, for fostering communication and collaboration within a stimulating research group, and for his steadfast kindness. Many thanks to Tzu-Ching Chang — I cherish all of the conversations we’ve had about science, and I’m grateful for her mentorship that goes back to my early days as a bright-eyed, first-year graduate student. I’m grateful to work with Phil Korngut, whose sharp knowledge of instrumentation and experiment is unparalleled and whose enthusiasm for science is palpable. Thanks for exposing me to my first crab boil on Chincoteague Island. Thanks to Jack Sayers for introducing me to the world of galaxy clusters, for his attention to detail, and for his mentorship and collaboration these past years. Thank you to Michael Zemcov for guiding my experience in the world of hardware and instrumentation, for working with me to improve my scientific writing, for tolerating my bad jokes, for entertaining my home town chronicles, and for continuing to associate after witnessing me spend ten minutes unsuccessfully drill a screw into a shipping box.

There are many others who have had a positive influence on my journey throughout

graduate school. Thank you to Yun-Ting Cheng, who was always ready to help when I started out in our group and has been a true collaborator. We will always be *CIBER* data siblings! Many thanks to Daniel Masters, Olivier Ilbert, Yi-Kuan Chiang, Bomee Lee, Ami Choi, Shuji Matsuura, Kohji Takimoto, Victoria Butler, Benjamin Vaughan and Ryan Willis for their collaboration and many insightful discussions. Thank you to Tansu Daylan and Stephen Portillo, who introduced me to computational research in college and have become wonderful collaborators and friends. Thanks as well to Doug Finkbeiner for his continued mentorship and for inspiring me to think deeply about hard problems.

My friends in the ObsCos group made graduate school a special time, filled with countless pizza lunches and happy hours. I'd like to thank my office-mates Albert Wandui, Lorenzo Minutolo, Silvia Zhang and Howard Hui. We've gone through a lot together, and I will miss Bagel Tuesdays. Thank you to Ritoban Thakur, Bryan Steinbach, Hien Nguyen, Jordan Mirocha, Chen Heinrich, Henry Gebhardt, Robin Wen, Spencer Everett, Ari Cukierman, Sofia Fatigoni and Chi Nguyen for many great conversations.

While I've always loved science, my path in physics and astronomy began in earnest during high school. I want to thank my high school teacher Alan Schorn for igniting my passion for physics and science research. Thank you to Francesca Civano, who advised my first astronomy project at the CfA. I am grateful to Eric Schles for believing in me and for many wonderful conversations.

My friends, old and new, have enriched my life with so much joy. They have also kept me sane. Thanks you Alex Ogren and Faye Golden for lots of great hikes, bike rides and forays into nature. We were a great germ circle. Paul Stainier, Akul Sharma, Julien Marlatt, Dylan Vartikar — we make a great VT crew, thanks for looking at all of the memes I send and affirming how they reflect my great sense of humor. Thanks to Michael Daugherty (aka Wolf Man) for many fun adventures, I'm looking forward to many more ahead at Berkeley. Thanks to Catherine Goohs, Dominic Yurk, Nitika Yadlapalli, Jacque Tawney, Gianfranco Canales, Karen Orta, Daniel Echeverri, Tyler Amick, Josh Lassman and Prithvi Akella. Here's to many more cabin trips. I am grateful to Kelly Kadlec and Jose Gonzalez for their friendship, for their consistent kindness and support, and for always knowing the best spot to grab tasty food and cocktails in Los Angeles (or any city, really). Thank you to those from Great Neck; Conor O'Byrne, Paula Tartell, Kamla Kumar, Alex Schneidman, Tyler Bloom, Nicole Biton, Dani Bar-Lavi, Emi Schaufeld, Russell Katz, Larry Li

and Ariel Kadouri. I feel lucky to have been friends so long and grown close despite living far away from each other.

Music has always been a restoring force in my life, and it's been a highlight to continue playing throughout graduate school. Thank you to friends Rounak Maiti, Michael Stevenson, Campbell Scott, Ian Green, Gil Young and Gina Segall for lots of adventures recording music and playing shows with Small Forward, Blu Wilson and Ian the Idiot. Thanks to Katherine, Adrian, Andre and Michelle for fun times with Women Tied to Railroad Tracks. Thank you to everyone from the Caltech Jazz Band and Barb Catlin in particular, for the opportunity to keep improving and for years of great music.

Lastly, I am blessed to have the support of my amazing family. You have been my unabated cheerleaders through some difficult times. I'm so happy to share what I do, and my excitement for it, with you. Thank you to my mom, my grandma, my uncle David and the rest of my family. I love you so much!

ABSTRACT

The utilization of several tracers of large-scale structure has led to important advancements in our understanding of the history of the Universe, in both characterizing cosmological initial conditions and late-time astrophysics. With the onset of dramatic changes in data volume and quality through existing and near-future experiments, methodologies that harness the information content in imaging and spectroscopic datasets while mitigating systematic effects will have larger impacts than ever before. In this thesis, we present a variety of analysis techniques for galaxy surveys of discrete objects and diffuse light measurements that are demonstrated on both synthetic and real datasets.

In Chapter 2, we develop techniques for measurement of near-infrared extragalactic background light (EBL) anisotropies, focusing on imager data from the Cosmic Infrared Background ExpeRiment (*CIBER*). Through improvements in methodology and data quality, we present fluctuation measurements in Chapter 3 that are five to ten times more sensitive on several arcminute to degree scales than existing studies, with clear detection of diffuse anisotropies exceeding those from the Poisson noise of individual stars and galaxies. In Chapter 4, we present a new suite of empirically-based galaxy simulations which we use to examine the diversity of galaxies that will be observed with *SPHEREx*, NASA's upcoming MIDEX mission. We then develop and apply redshift estimation techniques to synthetic *SPHEREx* observations generated from these simulations, demonstrating the ability to measure the distances to several hundred million galaxies over the full sky. In Chapter 5, we describe a formalism for modeling point-like and diffuse signals in astronomical images, which can be used for robust photometry in the presence of diffuse contaminants, extraction of diffuse signals in the presence of point source contaminants, and more general component separation. In Chapter 6 we apply this modeling framework to *Herschel*-SPIRE observations of galaxy cluster RX J1347.5-1145, measuring the diffuse thermal Sunyaev-Zel'dovich (tSZ) effect at high significance and using relativistic corrections of the tSZ spectrum to constrain the intra-cluster medium temperature, for which we find consistent estimates with independent X-ray measurements.

PUBLISHED CONTENT AND CONTRIBUTIONS

References

[1] Richard M. Feder et al. In: *arXiv e-prints*, arXiv:2312.04636 (Dec. 2023), arXiv:2312.04636. DOI: 10.48550/arXiv.2312.04636. arXiv: 2312.04636 [astro-ph.CO].

[2] Richard M. Feder et al. In: *AJ* 166.3, 98 (Sept. 2023), p. 98. DOI: 10.3847/1538-3881/ace69b. arXiv: 2307.10385 [astro-ph.IM].

[3] Victoria L. Butler, Richard M. Feder, et al. In: *ApJ* 932.1, 55 (June 2022), p. 55. DOI: 10.3847/1538-4357/ac6c04.

In all of the papers above, R.M.F. participated in the conception of the project, conducted the science analysis and wrote the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 The ΛCDM Universe

The foundations of modern cosmology were established over one hundred years ago with the advent of general relativity [1]. The Einstein field equations relate the curvature of four-dimensional space-time, encoded by the Einstein tensor $G_{\mu\nu}$, to the distribution of matter and energy within that space-time, described by the energy-momentum tensor $T_{\mu\nu}$,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}\mathcal{R}g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{1.1}$$

Here, $R_{\mu\nu}$ and $\mathcal{R}$ are the Ricci tensor and scalar that describe departures from the geometry of a Euclidean space and G is the gravitational constant. Solving the Einstein field equations yields the metric tensor $g_{\mu\nu}$, whose elements describe the evolution of a given spacetime. Under the assumptions of homogeneity and isotropy, the field equations can be solved exactly and lead to the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric [2],

$$g_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & a^2(t) & & \\ & & a^2(t) & \\ & & & a^2(t) \end{pmatrix} \tag{1.2}$$

where $a(t)$ is the dimensionless "scale factor" that parametrizes cosmic expansion. For a generic energy density ρ the time-time component of the field equations under a FLRW metric yields

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho \rightarrow \frac{H^2(t)}{H_0^2} = \frac{\rho}{\rho_{cr}}, \tag{1.3}$$

in which $\dot{a}$ denotes the time derivative of the scale factor, i.e., $da(t)/dt$. This equation provides a means to calculate the cosmic expansion rate through the Hubble parameter $H(t) = \dot{a}/a$, given the matter and energy which contribute to the stress-energy tensor $T_{\mu\nu}$. When the cosmic density ρ is equal to the critical density ρ_{cr}, the Friedmann equation leads to a flat universe with expansion in perpetuity. A universe with $\rho > \rho_{cr}$ is "closed" (positive curvature) and eventually collapses on

itself, while a universe with $\rho < \rho_{cr}$ is "open" (negative curvature) and expands forever.

While not explicitly written above, a cosmological constant Λ naturally arises in deriving Einstein's field equations. Its significance eluded physical interpretation for most of the twentieth century (and for some time it was discarded entirely), however the cosmological constant is now understood to describe the vacuum energy, or the energy density of space itself. With a cosmological constant the Friedmann equation takes the form

$$H^2(t) = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}, \tag{1.4}$$

where k is the space-time curvature (equal to zero for a flat spacetime) and c is the speed of light. The cosmological constant is a limit of a more generic source often referred to as "dark energy," though in this discussion we will use them both interchangeably.

To predict the cosmic expansion history $H(a)$, one decomposes the density ρ into its constituents and their relative abundances (matter Ω_m, radiation Ω_{rad}, dark energy Ω_Λ), expresses how each component of the density scales with a and evaluates the Friedmann equation,

$$H(a) = H_0\sqrt{\Omega_m a^{-3} + \Omega_{rad}a^{-4} + \Omega_\Lambda a^{-3(1+w)}}, \tag{1.5}$$

with the constraint that $\Omega_{tot} = \Omega_m + \Omega_{rad} + \Omega_\Lambda = 1$. Both baryonic matter and dark matter constitute the total matter density $\Omega_m = \Omega_b + \Omega_{CDM}$. The parameter H_0 is commonly known as the Hubble constant for the present day expansion rate. Through this equation one can characterize cosmic history in terms of epochs when individual components drive the expansion $H(a)$. Barring the epoch that generates initial conditions, expansion in the early universe is radiation-dominated, followed by matter-dominated expansion and then dark energy-driven expansion up through the present. While in the simplest model for the dark energy Ω_Λ is taken to be independent of the scale factor (i.e., $w = -1$), the equation-of-state parameter w is used to capture deviations from a true cosmological constant and admits more detailed parameterizations $w(a)$.

The Λ Cold Dark Matter (ΛCDM) model for cosmology has been successful in predicting a wide range of observations, including: the existence of the cosmic microwave background [CMB; 3]; the large-scale structure (LSS) clustering of galaxies (discussed in §1.4), the abundances of light elements such as hydrogen,

deuterium, helium and lithium, as well as the late-type accelerating expansion observed with supernovae and LSS [4].

1.2 Initial conditions

So far we have introduced how a spacetime metric defined by some matter and energy components expands and evolves in a homogeneous, isotropic universe. However, what are the physics that generate such cosmological initial conditions (ICs)? In other words, what happened exactly during the "Big Bang"? Answering this question is a major target for modern cosmology research.

Inflation

Cosmic inflation, initially conceived by Alan Guth in 1980 [5], is a widely recognized paradigm that seeks to explain the origin of ICs for LSS by invoking quantum mechanical fluctuations and a brief period of exponential expansion. Consider a simple toy model for inflation in which we have a scalar field $\phi(t)$ called the *inflaton*, for which the background value is only time-dependent. This field has an associated potential denoted $V(\phi)$. If we assume that the stress-energy of the inflaton is the dominant term during the inflation epoch, we can write down its stress-energy tensor,

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi - V(\phi)\right). \tag{1.6}$$

This can then be inserted into (1.1) to solve the Friedmann equation, leading to expressions for the field's energy density, pressure and equations of motion

$$\rho_\phi = \frac{\dot{\phi}^2}{2} + V(\phi); \tag{1.7}$$

$$P_\phi = \frac{\dot{\phi}^2}{2} - V(\phi); \tag{1.8}$$

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV}{d\phi} = 0. \tag{1.9}$$

If the potential $V(\phi)$ dominates over the field's kinetic term (i.e., $\dot{\phi}^2 \ll V(\phi)$, known as the "slow-roll" approximation), the pressure becomes negative and leads to positive acceleration in the scale factor $a(t)$,

$$a(t) \propto \exp Ht. \tag{1.10}$$

As the field rolls down steeper portions of the potential its kinetic term eventually dominates, leading to $P_\phi > 0$ and ending the inflationary epoch. Note that the

slow-roll approximation is a sufficient but not necessary condition for generating inflation.

A period of inflation in the earliest moments of the universe solves a number of problems surrounding the properties of ICs. Exponential expansion has the effect of driving any curvature to very small values, smoothing out inhomogeneities. This explains the observed flatness of the Universe ($k \approx 0$), which otherwise requires fine-tuning. Existing constraints on Ω_k require $N = 50 - 60$ e-folds of expansion during inflation, corresponding to a change in scale factor of $\sim 10^{26}$ over a period of $\lesssim 10^{-32}$ seconds. Inflation also addresses what is known as the "causality problem": the temperature of the cosmic microwave background (CMB) is observed to be extremely uniform in large patches of sky, however these patches are separated by such large distances that in principle they have no causal contact with another. Inflation solves this by positing that these regions were in fact in causal contact at early times, after which they are driven away to super-horizon distances. Lastly, inflation provides an explanation for the low observed density of *magnetic monopoles*, which are expected to be produced in large quantities when the early universe was very hot and would otherwise be the primary cosmic constituent.

As the inflaton field is a quantum mechanical object, it has quantum fluctuations $\delta\phi(\vec{x}, t) = \phi(\vec{x}, t) - \overline{\phi}(t)$. These translate to curvature perturbations $\delta\mathcal{R}$ which then evolve over time, setting the seeds for LSS to form. This paradigm is profound; it relates fluctuations on the smallest scales in the earlist moments of the universe to cosmic structure on extremely large scales. By relating the physical conditions of inflation to LSS, it becomes possible to probe the energy scale of inflation (which depends on the tensor component of the primordial fluctuations) and the duration of inflation (bounded by Ω_k), among other things. Such energy scales are extremely high $E \sim 10^{16}$ GeV, orders of magnitude beyond the reach of terrestrial experiments such as the Large Hadron Collider.

Primordial non-Gaussianities and multi-field inflation

Given our ignorance of the particle sector at such high energies, it is natural to consider the presence of multiple fields during inflation. In single-field models, the inflaton acts as a quantum harmonic oscillator and its fluctuation modes are independent, leading to Gaussian ICs. In this limit, the statistics of primordial curvature fluctuations are fully captured by the *power spectrum* $P(k)$, defined as the Fourier transform of the two-point correlation function. The initial curvature

perturbations follow

$$P_{\mathcal{R}}(k) = A_s(k/k_{piv})^{n_s-1}. \tag{1.11}$$

Here, the power spectrum amplitude A_s corresponds to the energy scale of inflation. The spectral index n_s, is sensitive to the duration of inflation and sources a small deviation from the scale-invariant Harrison-Zel'dovich spectrum [6, 7, 8].

Under the assumption of translation invariance, i.e., $\phi(\vec{x}, t) = \phi(t)$, it has been shown that single-field models obey consistency relations [9, 10] which constrain the three-point correlation function $\langle\phi(k_1)\phi(k_2)\phi(k_3)\rangle$ in the "squeezed" limit ($k_3 \gg k_2, k_1$),

$$\langle\phi(k_1)\phi(k_2)\phi(k_3 \to 0)\rangle \approx (n_s - 1)(2\pi)^3\delta(k_1 + k_2)P_\phi(k_1)P_\phi(k_2). \tag{1.12}$$

The local primordial non-Gaussianity (pNG) parameter f_{NL}^{loc} is used to parameterize the amplitude of the primordial squeezed bispectrum, such that in the equation above $f_{NL}^{loc} \approx |n_s - 1|$. Measurements from *Planck* constrain $n_s = 0.9649 \pm 0.00042$ (68% CL) [11], meaning any non-Gaussianity from single fields is small ($f_{NL}^{loc} \sim O(10^{-2})$).

However, additional fields which are local in real space can source larger non-Gaussianities to the primordial gravitational potential, commonly expressed in terms of the Gaussian potential Φ_G,

$$\Phi(\vec{x}) = \Phi_G(\vec{x}) + f_{NL}^{loc}(\Phi_G^2(\vec{x}) - \langle\Phi_G^2\rangle). \tag{1.13}$$

In this equation, f_{NL}^{loc} captures leading-order departures from Gaussianity, however higher-order departures are possible as well (captured by parameters τ_{NL}, g_{NL} at third and fourth order in Φ_G, respectively). Many scenarios have been proposed that can generate larger non-Gaussianities. These include, but are not limited to: the *curvaton* scenario, in which an additional light scalar χ does not drive inflation but generates curvature perturbations after the inflaton field has decayed [12, 13]; inflaton self-interactions and variations in the inflaton decay rate [14], and spectator fields that decay during inflation [15]. Many (but not all) multiple-field scenarios predict $f_{NL}^{loc} \sim O(1)$, meaning that a detection of non-zero f_{NL}^{loc}, or lack thereof, has the potential to rule out large classes of inflationary models. Non-inflationary cosmologies can also produce large f_{NL}^{loc} [16] and can be constrained as well.

While we have discussed pNG of the local type from the squeezed bispectrum, we note that other bispectrum shapes are sensitive to inflationary physics as well. For example, equilateral-shape pNG ($k_1 = k_2 = k_3$) can constrain inflaton models with

non-trivial kinetic terms [17], and orthogonal-shape pNG, a linear combination of equilateral- and folded-shape ($k_1 = k_2 = k_3/2$) triangles, can arise in Dirac-Born-Infield Galileon inflation [18, 19].

1.3 Evolution of the potential through radiation- and matter-domination

Having introduced the ΛCDM model and its underpinnings, we now turn to the evolution of structure in the radiation-dominated (RD, $z > z_{eq} \sim 3600$) and matter-dominated (MD, $z < z_{eq}$) epochs. Here, z_{eq} indicates the redshift of matter-radiation equality. In Fourier space, the late-time gravitational potential is related to the primordial potential through the following expression,

$$\Phi(\vec{k}, z) = \Phi_p(\vec{k}) \times T(k) \times D(z), \tag{1.14}$$

where $\Phi_p(\vec{k})$ is the primordial potential. The matter transfer function $T(k)$ describes the suppression of fluctuations during the radiation-dominated (RD) epoch, while the growth function $D(z)$ captures how density perturbations in Φ are enhanced during the MD epoch.

During the RD epoch, the potential is driven by perturbations to the radiation and associated pressure, which in turn influences the sub-dominant matter distribution. As fluctuation modes enter the causal horizon, their potential begin to decay with an evolution that can be derived using the Boltzmann equations, given the monopole Θ_0 and dipole Θ_1 of the photon distribution. The resulting suppression depends on the duration in which each mode is within the horizon, thereby impacting small-scale modes of $\Phi_p(k \gg k_{eq})$ most severely. However, super-horizon modes during the RD epoch are left relatively unaffected, since the time for light to travel across each mode is too large. Solving the Boltzmann and Einstein field equations in the super-horizon limit, one finds that the transfer function reduces the initial fluctuations by only 10%,

$$\Phi_{MD}(k < k_{eq}) = \frac{9}{10}\Phi_{RD}(k < k_{eq}). \tag{1.15}$$

This has two important implications. The first is that the RD epoch leads to a turnover in the matter power spectrum, which peaks near $k \sim k_{eq}$. The second is that, by probing large enough scales ($k < k_{eq}$), one can measure modes of the potential from ICs that are relatively pristine. These modes are also unaffected by small-scale non-linearities introduced by gravitational evolution, making them easier to model.

To relate matter perturbations at z_{eq} to the present day, we need to model the effects of gravitational evolution. In the CDM model, dark matter is assumed to be weakly

interacting and behaving as a collisionless medium on large scales. In this limit, evolution of the dark matter potential during the MD epoch can be modeled using the Vlasov-Poisson equations, which describe the evolution of gravitational instabilities in a pressureless fluid. The dark matter 6D phase distribution $f(\vec{x}, \vec{p}, t)$ is related to the gravitational potential by

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \vec{v} \cdot \vec{\nabla}_x f - \vec{\nabla}\Phi \cdot \vec{\nabla}_v f = 0 \tag{1.16}$$

and

$$\nabla^2\Phi = 4\pi G \rho_m(\vec{x}, t). \tag{1.17}$$

The second equality in (1.16) comes from invoking phase space conservation. For predictions with reasonable fidelity, these equations are integrated numerically, though approximations valid at higher redshifts exist (e.g., the Zel'dovich approximation [20]). Another approach, which avoids issues related to shell-crossings, is to approximate the DM distribution with an initial set of test particles and solve their equations of motion under gravitational forces. These Monte Carlo simulations, commonly dubbed "N-body" simulations, are computationally demanding but form the bedrock for most modern day cosmological simulations. Through these simulations one finds that the matter distribution at late times is quite complex, with DM filaments and sheets that comprise what is known colloquially as the "cosmic web" [21].

While we will not discuss dark energy at length, it should be noted that its existence affects predictions for the structure of matter fluctuations at late times $z \leq 0.7$. By changing $H(z)$ relative to the matter-only case, dark energy modifies the growth factor $D(z)$ in a way that depends on both Ω_Λ and its equation of state w. It also indirectly affects the turnover scale in the matter power spectrum, which depends on Ω_m.

With a sufficiently accurate model that relates LSS at early times to late times, it becomes possible to use present day structure (or structure along a given lightcone) to infer properties of the early universe. The question then becomes, how do we relate this late-time structure to observations?

1.4 Tracers of large-scale structure

As coined, dark matter and dark energy lack known couplings to light, precluding direct observation with existing telescopes. Nonetheless, their outsized influence on baryonic matter means that luminous sources can be considered as tracers of

the underlying density field (dominated by DM). This is the assumption that allows cosmologists to study LSS through light production across cosmic history. Cosmological information is encoded the spatial anisotropies of these tracers, which can be characterized as a function of scale using a range of statistical estimators.

The matter density $\rho_m(\vec{x}, t)$ can be expressed in terms of the mean density $\overline{\rho}_m$ and the matter density contrast $\delta_m(\vec{x}, t)$ as

$$\rho_m(\vec{x}, t) = \overline{\rho}_m \left[1 + \delta_m(\vec{x}, t)\right] . \tag{1.18}$$

The matter density contrast $\delta_m(\vec{x}, t)$ encodes anisotropies in the matter density field.

The Press-Schechter formalism (P-S) [22] is widely used to relate $\delta_m(\vec{x}, t)$ to the number of objects with some mass within a given volume. The co-moving number density per unit mass is given by

$$\frac{dn}{dM} = \sqrt{\frac{2}{\pi}} \frac{\overline{\rho}}{M} \frac{\delta_c}{\sigma^2} \frac{d\sigma}{dM} \exp\left[-\frac{\delta_c^2}{2\sigma^2}\right], \tag{1.19}$$

where σ denotes the dispersion in δ_m. This assumes that matter perturbations δ_m follow a Gaussian distribution with some power spectrum $P(k)$, that these perturbations grow linearly according to the growth function $D_+(z)$, and that halo objects are spherical, virialized overdensities δ surpassing the critical density contrast $\delta_c \approx 1.686$. This formulation is "universal" in the sense that the derived mass function solely depends on the rms of the matter density contrast $\sigma(M)$, smoothed with a top hat filter on scale $R = \left(\frac{3M}{4\pi\overline{\rho}}\right)^{1/3}$. This has been generalized beyond the spherical halo approximation with the Sheth-Tormen mass function, which only assumes that the halos are ellipsoidal [23]. The halo mass function (HMF) predicted by P-S has been validated against N-body simulations, however we note there are limits in which the universal HMF approximation breaks down, notably in the presence of halo mergers [24, 25].

Galaxy Surveys

As galaxies form preferentially within DM halos, they are biased tracers of the halo density field and thus trace the underlying matter density field. By measuring the distances of individual galaxies through their redshifted spectral energy distributions (SEDs) along the Hubble flow, it is possible to construct three-dimensional maps of the galaxy distribution and measure its spatial anisotropies. On large scales, one can express the power spectrum of the galaxy density contrast $\delta_g(\vec{x})$ as

$$P_g(k) = b_g b_h P_m(k), \tag{1.20}$$

where b_g and b_h are the linear galaxy and halo biases, respectively. While the first galaxy survey dates back to the late 1970s/ early 1980s [the CfA Redshift Survey; 26], it was only until the development of multi-slit spectrographs that large ($> 10^5$) redshift samples were obtainable. The first cosmological constraints from galaxy surveys were obtained in the early 2000s with the 2dF Galaxy Redshift Survey [27] and the Sloan Digital Sky Survey [SDSS; 28], constraining Ω_m, the baryon fraction Ω_b/Ω_m and making first detections of baryon acoustic oscillations imprinted on the matter distribution before CMB recombination [29, 30]. Importantly, the results from these surveys were in agreement with independent, contemporary CMB measurements from the Wilkinson Microwave Anisotropy Probe [WMAP; 31, 32], providing a strong confirmation of the ΛCDM model. These pioneering experiments laid the groundwork for modern galaxy surveys, the landscape of which is now much broader. Existing surveys conducted with the Dark Energy Spectroscopic Instrument [DESI; 33], the Dark Energy Camera [DES; 34], Hyper Suprime-Cam [HSC; 35], along with near-future surveys from *Euclid* [36], *Rubin* [37] and *Roman* [38] seek to place percent and even sub-percent level constraints on ΛCDM parameters.

Beyond the standard ΛCDM cosmological parameters, galaxy surveys also offer a unique opportunity to probe the physics of inflation through measurement of f_{NL}^{loc} using the *scale-dependent bias* [39, 24, 40]. Under the assumption of universality, the scale-dependent bias takes the form

$$\Delta b(k, f_{NL}) = 3 f_{NL}[b(M) - 1]\delta_c \frac{\Omega_m H_0^2}{k^2 T(k) D(z)}, \tag{1.21}$$

which diverges on large scales. This is a primary science target for *SPHEREx* [41], NASA's upcoming MIDEX mission and the topic of Chapter 4 of this dissertation.

Galaxy surveys face a number of challenges as they continue to push clustering measurements to larger samples and higher sensitivities. One challenge is observational in nature, namely obtaining reliable distances for a large sample of galaxies. While the observed wavelengths of properly identified emission lines such as $H\alpha$ and [OII] yield precise estimates of galaxy redshift, such measurements with high-resolution spectroscopic instruments are expensive. This has led to the development of photometric redshift estimation (commonly called *photo-z* estimation), which utilizes measured broadband colors instead of a high-resolution spectrum [42]. Photometric redshift surveys sacrifice redshift accuracy for coverage of a much larger (typically by a factor of $O(10)$) sample and are more sensitive to assumptions about the rest frame SEDs of the galaxies being probed. Even for high-resolution spectroscopic

surveys, redshift measurements can be complicated by line mis-identifications and errors in star-galaxy separation. Furthermore, the mapping of *any* redshift to distance is spoiled by the peculiar velocities of galaxies within their local gravitational potential and the relative motion of large-scale structures. These two effects manifest in what are known as *redshift space distortions* [43, 44, 45].

Relating theory to galaxy clustering measurements also comes with challenges. One example is the determination of galaxy bias b_g, which is a function of galaxy formation physics and can depend on properties such as galaxy luminosity, morphology, color and redshift. On small scales, where the constraining power naively scales with the number of Fourier modes ($N_{modes} \propto k^3_{max}$), non-linear gravitational evolution enhances galaxy clustering beyond that predicted from linear perturbation theory, requiring more detailed galaxy bias prescriptions with a larger number of parameters, along with cosmological simulations with thoroughly vetted systematic uncertainties. Such challenges in practice place limits on the physical scales used in cosmological analyses.

Intensity Mapping

Galaxy surveys continue to improve in terms of depth and angular coverage, however by construction they constitute only part of the total extragalactic background light (EBL). Intensity mapping (IM) is an alternative approach for studying LSS that promises a consensus view of *all* astrophysical populations that contribute to cosmic light production, accomplished through measurement of small- and large-angle intensity fluctuations. At a given wavelength λ, the observed intensity λI_λ (with units of nW m^{-2} sr^{-1}) is the integrated light from all sources along a given line of sight,

$$\lambda I_\lambda = \int dz \frac{d(\lambda I_\lambda)}{dz}(z, \lambda). \tag{1.22}$$

By relaxing the requirement of measuring the properties of individual discrete sources, IM measurements are often able to survey larger portions of sky. At the risk of being pedantic, we note that the fundamental observable for *any* imaging survey is intensity, from which one can derive object brightnesses, distances and other downstream astrophysical information. In this sense, IM has the benefit of utilizing observables whose information content is close to that of the raw data. In practice the situation is, of course, much more complicated, and the choice of an "optimal observable" depends on the measurements being pursued, the filtering required to remove data contaminants, etc.

For the purpose of this introduction, let us restrict ourselves to contributions from galaxies and ignore redshift space distortions. Following conventions from [46], the clustering angular power spectrum from the galaxy density contrast can be expressed as a bias-weighted integral of the matter power spectrum P_m,

$$C^g_{\ell,clus} = \int dz \frac{H(z)}{c} \left(\frac{f^g(z)}{\chi(z)}\right)^2 b_g^2(z) P_m \left(k = \frac{\ell + \frac{1}{2}}{\chi(z)}, z\right), \tag{1.23}$$

in which b_g denotes the galaxy bias and $\chi(z)$ is the co-moving distance. The quantity $f^g(z)$ denotes efficiency kernel of a given galaxy sample, which includes the redshift kernel of the sample and any redshift-dependent selection effects. We then express the clustering component of the *intensity* angular power spectrum $C^I_{\ell,clus}$ as the bias- and intensity-weighted integral of the matter power spectrum P_m, given some filter spectral response $R(\lambda)$,

$$C^I_{\ell,clus}(\lambda) = \int dz \frac{H(z)}{c} \left(\frac{f^I(z)}{\chi(z)}\right)^2 b_I^2(z,\lambda) \left[\frac{d(\lambda I_\lambda)(z, R(\lambda))}{dz}\right]^2 P_m \left(k = \frac{\ell + \frac{1}{2}}{\chi(z)}, z\right). \tag{1.24}$$

The quantity $b_I(z,\lambda)$ is the *intensity bias*. This can be modeled within the same halo occupation distribution (HOD) framework [47] used in treatments of galaxy clustering that assume a fiducial halo mass function dn/dM, but instead weighted by luminosity $L(M,z)$,

$$b_I(z) = \frac{\int dM \frac{dn}{dM}(M,z) b_h(M,z) L(M,z)}{\int dM \frac{dn}{dM}(M,z) L(M,z)}. \tag{1.25}$$

The small-scale intensity power spectrum is sensitive to the integrated galactic light through the Poisson noise fluctuations of the sources,

$$C^I_{\ell,Pois.}(\lambda) = \int dz \int_{m_{min}}^{\infty} dm \Phi \left(m, z, \frac{\lambda}{1+z}\right) [\lambda F_\lambda(m)]^2 \frac{d\chi}{dz}(z) D_A^2(z), \tag{1.26}$$

where $\Phi(m, z, \frac{\lambda}{1+z})$ is the galaxy luminosity function that depends on magnitude and the rest-frame wavelength corresponding to λ_{obs} at fixed redshift.

While intensity maps (at fixed wavelength) are inherently two-dimensional quantities, spectral information can help probe the three-dimensional spatial light distribution through cross-correlations. We focus on broadband IM in this dissertation, however line intensity mapping (LIM) is a rapidly growing line of research that exploits the knowledge of one or several spectral lines to trace LSS as a function of redshift. The landscape of LIM experiments spans several decades in wavelength

(see [48] for a recent comprehensive review). In the rest-frame optical and NIR, the Balmer lines Hα and Hβ, singly- and doubly-ionized oxygen ([OII] and [OIII], respectively) and Paschen-α will be prime LIM targets for SPHEREx, in addition to redshifted Lyman-α in the rest-UV [41, 49]. At far-infrared wavelengths, the [CII] fine-structure line and rotational carbon-monoxide (CO) transitions are the target of several experiments such as COMAP [50], TIME [51], CONCERTO [52], TIM [53] and EXCLAIM [54]. At much lower frequencies, statistical measurements of the rest-frame 21 centimeter line, which arises from spin-flip transitions between the hyperfine ground states of hydrogen, are actively pursued by several experiments including: CHIME [55]; HERA [56]; MeerKAT [57]; GBT [58]; MWA [59] and LOFAR [60].

While galaxy surveys and IM differ in terms of experimental design and overall science themes, there is a natural complementarity between the two in studies of LSS. Detected galaxies can be considered as tracers of the EBL intensity distribution, just as they are tracers of the dark matter distribution. Similar in form to (1.23) and (1.24), the clustering component of the galaxy-intensity cross-power spectrum can be expressed as

$$C^{\times}_{\ell,clus} = \int dz \frac{H(z)}{c} \frac{f^g(z) f^I(z)}{\chi^2(z)} r_{\times}(z) b_g(z) b_I(z) \frac{d(\lambda I_\lambda)}{dz} P_m\left(k = \frac{\ell + 1/2}{\chi(z)}, z\right), \tag{1.27}$$

where we have introduced $r_{\times}(z)$, the cross-correlation coefficient between galaxy density fluctuations and those of an intensity map with overlapping coverage. By combining the set of auto- and cross-spectra $\{C^g_{\ell,clus}, C^I_{\ell,clus}, C^{\times}_{\ell,clus}\}$ with knowledge of $f^g(z)$ and b_g, one can extract the bias-weighted correlated intensity as a function of redshift, $b_I \frac{d(\lambda I_\lambda)}{dz}(z)$. This is referred to as *redshift tomography* [61, 62].

A challenge for IM is that, by measuring all incoming light, IM measurements are sensitive to any contaminants that source intensity and intensity fluctuations. The relevance of any given contaminant depends on the wavelengths and angular scales being probed. IM experiments are typically designed to minimize diffuse stray light from off-axis sources and reflections within the focal plane and telescope optics, however large-angle measurements are sensitive to these instrumental systematics along with image persistence, extended PSF, cosmic rays and other effects. At near-infrared (NIR) wavelengths, which we will focus on in this dissertation, atmospheric emission from atoms and molecules in the upper atmosphere ("airglow") is the largest foreground signal, nearly 1500 times brighter than the average sky intensity. Airglow places severe limitations on ground-based or balloon-borne fa-

cilities and motivates space-borne experiments. Beyond Earth's atmosphere, there are foregrounds within our solar system (zodiacal light, ZL) and within our galaxy (diffuse galactic light, DGL). Foregrounds are a concern not only for IM, but also for cross-correlations with galaxy catalogs, as they can affect the galaxy selection function $f^g(z, \theta_x, \theta_y)$. These correlated effects will in general depend on the galaxy and IM survey strategy, and whether the foregrounds are static or time variable.

Despite these challenges, IM offers an enticing new window into both cosmology and astrophysics. Given IM's intentional sensitivity to large-angle, diffuse light that would otherwise be rejected or inaccessible with conventional means, there is the discovery potential to identifying new signals that may carry broader significance or lend novel insights to our understanding of the history of the universe.

Let us take a moment to share an overly broad perspective, without which any cosmology dissertation would be incomplete. Our understanding of cosmology is in a rather awkward state. On the one hand, we have very detailed models for LSS formation and evolution that explain a multitude of observations. At the same time, nearly 95 per cent of the cosmic energy content in the ΛCDM model are in the form of dark matter and dark energy. These constituents are largely understood in terms of their phenomenology, yet their physical origins are not known. Despite heroic efforts to test particle and/or fluid descriptions of dark matter and dark energy, their direct measurement has eluded experiment. Likewise, the phenomenon of inflation has become the *de facto* paradigm describing initial conditions in ΛCDM cosmology, yet we lack a definitive detection relating its physical properties to observation. The deduction of these components follows the intuition that they are the simplest explanation for what has been observed. And yet, considered in full, there is nothing "simple" about them! What will it take to break the cosmic levee before us? Is it simply a matter of pushing existing measurements to higher precision, or will new types of observables be needed? Time will tell.

1.5 Dissertation overview

The work in this dissertation touches on a number of topics in observational cosmology. The first two chapters focus on intensity mapping in at near-infrared (NIR) wavelengths, in particular for the Cosmic Infrared Background ExpeRiment (CIBER), a sounding rocket payload designed to study the extragalactic background light (EBL). In Chapter 2 we discuss improvements in methodology required for unbiased measurement of surface brightness fluctuations. There is a strong heritage

with methods used to study CMB anisotropies, however NIR intensity mapping measurements have their own set of challenges which motivate development of new techniques. In Chapter 3, we apply this improved analysis methodology to observations from CIBER-1's fourth flight, measuring large-angle surface brightness fluctuations at 1.1 μm and 1.8 μm at high significance.

In Chapter 4, we detail galaxy simulations and redshift predictions for SPHEREx, NASA's next MIDEX explorer slated for launch in early 2025. While SPHEREx has a small aperture, large etendue telescope designed for intensity mapping, its broad spectral coverage and intermediate spectral resolution enable what will be the largest effective volume galaxy survey to date. Through measurements of galaxy clustering on ultra-large scales, SPHEREx will make large strides in confirming or rejecting the presence of local pNG in cosmological ICs by measuring f_{NL}^{loc}.

The last two chapters detail the development and application of image-level forward modeling to observations from the SPIRE photometer onboard the *Herschel* sub-millimeter space observatory. While the point-like and diffuse signals in astronomical observations are typically treated separately in measurements of LSS, the full information encoded by both can be captured by modeling directly to the data. The work in these chapters was initially motivated by a multiwavelength analysis of galaxy cluster RXJ 1347 in order to probe its ICM temperature through the Sunyaev-Zel'dovich effect (SZ effect) signal. The challenge of this analysis is to extract the surface brightness spectrum of the faint, extended SZ effect in the presence of strong contamination from cosmic infrared background (CIB) galaxies and diffuse Galactic cirrus foregrounds.

References

[1] A. E. In: *Sitzungsberichte der Koumlniglich Preussischen Akademie der Wissenschaften* (Jan. 1915), pp. 844–847.

[2] A. F. In: *Zeitschrift fur Physik* 21.1 (Dec. 1924), pp. 326–332. DOI: 10.1007/BF01328280.

[3] A. A. P. and R. W. W. In: *ApJ* 142 (July 1965), pp. 419–421. DOI: 10.1086/148307.

[4] T. M. C. A. et al. In: *ApJL* 872.2, L30 (Feb. 2019), p. L30. DOI: 10.3847/2041-8213/ab04fa. arXiv: 1811.02374 [astro-ph.CO].

[5] A. H. G. In: *PhRvD* 23.2 (Jan. 1981), pp. 347–356. DOI: 10.1103/PhysRevD.23.347.

[6] E. R. H. In: *PhRvD* 1.10 (May 1970), pp. 2726–2730. DOI: 10.1103/PhysRevD.1.2726.

[7] Y. B. Z. In: *MNRAS* 160 (Jan. 1972), 1P. DOI: 10.1093/mnras/160.1.1P.

[8] P. J. E. P. and J. T. Y. In: *ApJ* 162 (Dec. 1970), p. 815. DOI: 10.1086/150713.

[9] J. M. In: *Journal of High Energy Physics* 2003.5, 013 (May 2003), p. 013. DOI: 10.1088/1126-6708/2003/05/013. arXiv: astro-ph/0210603 [astro-ph].

[10] P. C. and M. Z. In: *JCAP* 2004.10, 006 (Oct. 2004), p. 006. DOI: 10.1088/1475-7516/2004/10/006. arXiv: astro-ph/0407059 [astro-ph].

[11] Planck Collaboration et al. In: *A&A* 641, A10 (Sept. 2020), A10. DOI: 10.1051/0004-6361/201833887. arXiv: 1807.06211 [astro-ph.CO].

[12] D. H. L., C. U., and D. W. In: *PhRvD* 67.2, 023503 (Jan. 2003), p. 023503. DOI: 10.1103/PhysRevD.67.023503. arXiv: astro-ph/0208055 [astro-ph].

[13] N. B., S. M., and A. R. In: *PhRvD* 69.4, 043503 (Feb. 2004), p. 043503. DOI: 10.1103/PhysRevD.69.043503. arXiv: hep-ph/0309033 [hep-ph].

[14] M. Z. In: *PhRvD* 69.4, 043508 (Feb. 2004), p. 043508. DOI: 10.1103/PhysRevD.69.043508. arXiv: astro-ph/0306006 [astro-ph].

[15] L. W. and A. M. In: *JCAP* 2013.5, 012 (May 2013), p. 012. DOI: 10.1088/1475-7516/2013/05/012. arXiv: 1302.2637 [astro-ph.CO].

[16] J. L. L. and S. R. P. In: *PhRvD* 80.6, 063503 (Sept. 2009), p. 063503. DOI: 10.1103/PhysRevD.80.063503. arXiv: 0906.0530 [hep-th].

[17] D. B. and D. G. In: *JCAP* 2011.9, 014 (Sept. 2011), p. 014. DOI: 10.1088/1475-7516/2011/09/014. arXiv: 1102.5343 [hep-th].

[18] S. R-P. In: *Classical and Quantum Gravity* 28.24, 249601 (July 2011), p. 249601. DOI: 10.1088/0264-9381/28/24/249601. arXiv: 1105.6366 [astro-ph.CO].

[19] E. S. and D. T. In: *PhRvD* 70.10, 103505 (Nov. 2004), p. 103505. DOI: 10.1103/PhysRevD.70.103505. arXiv: hep-th/0310221 [hep-th].

[20] Ya. B. Z. In: *A&A* 5 (Mar. 1970), pp. 84–89.

[21] J. R. B, L. K., and D. P. In: *Nature* 380.6575 (Apr. 1996), pp. 603–606. DOI: 10.1038/380603a0. arXiv: astro-ph/9512141 [astro-ph].

[22] W. P and P.S S. In: *ApJ* 187 (Feb. 1974), pp. 425–438. DOI: 10.1086/152650.

[23] R. K. S., H. J. M., and G. T. In: *MNRAS* 323.1 (May 2001), pp. 1–12. DOI: 10.1046/j.1365-8711.2001.04006.x. arXiv: astro-ph/9907024 [astro-ph].

[24] A. S. et al. In: *JCAP* 2008.8, 031 (Aug. 2008), p. 031. DOI: 10.1088/1475-7516/2008/08/031. arXiv: 0805.3580 [astro-ph].

[25] E. F. et al. In: *JCAP* 2024.2, 048 (Feb. 2024), p. 048. DOI: 10.1088/1475-7516/2024/02/048. arXiv: 2311.10088 [astro-ph.CO].

[26] J. H. et al. In: *ApJS* 52 (June 1983), pp. 89–119. DOI: 10.1086/190860.

[27] M. C. et al. In: *MNRAS* 328.4 (Dec. 2001), pp. 1039–1063. DOI: 10.1046/j.1365-8711.2001.04902.x. arXiv: astro-ph/0106498 [astro-ph].

[28] D. G. Y. et al. In: *AJ* 120.3 (Sept. 2000), pp. 1579–1587. DOI: 10.1086/301513. arXiv: astro-ph/0006396 [astro-ph].

[29] S. C. et al. In: *MNRAS* 362.2 (Sept. 2005), pp. 505–534. DOI: 10.1111/j.1365-2966.2005.09318.x. arXiv: astro-ph/0501174 [astro-ph].

[30] D. J. E. et al. In: *ApJ* 633.2 (Nov. 2005), pp. 560–574. DOI: 10.1086/466512. arXiv: astro-ph/0501171 [astro-ph].

[31] D. N. S. et al. In: *ApJS* 148.1 (Sept. 2003), pp. 175–194. DOI: 10.1086/377226. arXiv: astro-ph/0302209 [astro-ph].

[32] M. T. et al. In: *PhRvD* 69.10, 103501 (May 2004), p. 103501. DOI: 10.1103/PhysRevD.69.103501. arXiv: astro-ph/0310723 [astro-ph].

[33] M. L. et al. In: *Bulletin of the American Astronomical Society*. Vol. 51. Sept. 2019, 57, p. 57. DOI: 10.48550/arXiv.1907.10688. arXiv: 1907.10688 [astro-ph.IM].

[34] B. F. et al. In: *AJ* 150.5, 150 (Nov. 2015), p. 150. DOI: 10.1088/0004-6256/150/5/150. arXiv: 1504.02900 [astro-ph.IM].

[35] S. M. et al. In: *Ground-based and Airborne Instrumentation for Astronomy IV*. Ed. by Ian S. McLean, Suzanne K. Ramsay, and Hideki Takami. Vol. 8446. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Sept. 2012, 84460Z, 84460Z. DOI: 10.1117/12.926844.

[40] V. D., U. S., and I. T. I. In: *MNRAS* 396.1 (June 2009), pp. 85–96. DOI: 10.1111/j.1365-2966.2009.14721.x. arXiv: 0811.2748 [astro-ph].

[42] A. J. C. et al. In: *AJ* 110 (Dec. 1995), p. 2655. DOI: 10.1086/117720. arXiv: astro-ph/9508100 [astro-ph].

[43] J. C. J. In: *MNRAS* 156 (Jan. 1972), 1P. DOI: 10.1093/mnras/156.1.1P. arXiv: 0810.3908 [astro-ph].

[44] N. K. In: *MNRAS* 227 (July 1987), pp. 1–21. DOI: 10.1093/mnras/227.1.1.

[45] A. J. S. H. In: *ApJL* 385 (Jan. 1992), p. L5. DOI: 10.1086/186264.

[47] Z. Z. and D. H. W. In: *ApJ* 659.1 (Apr. 2007), pp. 1–28. DOI: 10.1086/512151. arXiv: astro-ph/0512071 [astro-ph].

[48] E. K. et al. In: 51.3, 101 (May 2019), p. 101. DOI: 10.48550/arXiv.1903.04496. arXiv: 1903.04496 [astro-ph.CO].

[49] Y. T. C. et al. In: *arXiv e-prints* (Mar. 2024). DOI: 10.48550/arXiv.2403.19740.

[50] T. Y. L. et al. In: *ApJ* 817.2, 169 (Feb. 2016), p. 169. DOI: 10.3847/0004-637X/817/2/169. arXiv: 1503.08833 [astro-ph.CO].

[51] G. S. et al. In: *ApJ* 915.1, 33 (July 2021), p. 33. DOI: 10.3847/1538-4357/abfe62. arXiv: 2012.09160 [astro-ph.GA].

[52] G. L., M. C., and M. C. In: *A&A* 609, A130 (Jan. 2018), A130. DOI: 10.1051/0004-6361/201732019. arXiv: 1711.00798 [astro-ph.GA].

[53] J. V. et al. In: *arXiv e-prints* (Sept. 2020). DOI: 10.48550/arXiv.2009.14340. arXiv: 2009.14340 [astro-ph.IM].

[54] A. R. P. et al. In: *MNRAS* 521.4 (June 2023), pp. 6124–6142. DOI: 10.1093/mnras/stad916. arXiv: 2209.02497 [astro-ph.CO].

[55] K. B. et al. In: *Ground-based and Airborne Telescopes V*. Ed. by Larry M. Stepp, Roberto Gilmozzi, and Helen J. Hall. Vol. 9145. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2014, 914522, p. 914522. DOI: 10.1117/12.2054950. arXiv: 1406.2288 [astro-ph.IM].

[56] D. R. D. et al. In: *PASP* 129.974 (Apr. 2017), p. 045001. DOI: 10.1088/1538-3873/129/974/045001. arXiv: 1606.07473 [astro-ph.IM].

[57] M. S. et al. In: *MeerKAT Science: On the Pathway to the SKA*. Jan. 2016, 32, p. 32. DOI: 10.22323/1.277.0032. arXiv: 1709.06099 [astro-ph.CO].

[58] K. W. M. et al. In: *ApJL* 763.1, L20 (Jan. 2013), p. L20. DOI: 10.1088/2041-8205/763/1/L20. arXiv: 1208.0331 [astro-ph.CO].

[59] S. J. T. et al. In: 30, e007 (Jan. 2013), e007. DOI: 10.1017/pasa.2012.007. arXiv: 1206.6945 [astro-ph.IM].

[60] M. P. v. H. et al. In: *A&A* 556, A2 (Aug. 2013), A2. DOI: 10.1051/0004-6361/201220873. arXiv: 1305.3550 [astro-ph.IM].

Chapter 2

MEASURING NIR EXTRAGALACTIC BACKGROUND LIGHT ANISOTROPIES WITH CIBER. I: PSEUDO-POWER SPECTRUM FORMALISM, IMPROVED SOURCE MASKING AND VALIDATION ON MOCKS

Abstract: Precise, unbiased measurements of extragalactic background anisotropies require careful treatment of systematic effects in fluctuation-based intensity mapping measurements. In this paper we detail improvements in methodology for the Cosmic Infrared Background ExpeRiment (*CIBER*), concentrating on flat field errors and source masking errors. In order to bypass the use of field differences, which mitigate flat field errors but reduce sensitivity, we characterize and correct for the flat field on pseudo-power spectra, which includes both additive and multiplicative biases. To more effectively mask point sources at 1.1 μm and 1.8 μm, we develop a technique for predicting masking catalogs that utilizes optical and NIR photometry through random forest regression. This allows us to mask over two Vega magnitudes deeper than the completeness limits of 2MASS alone, with errors in the shot noise power remaining below $< 10\%$ at all masking depths considered. Through detailed simulations of *CIBER* observations, we validate our formalism and demonstrate unbiased recovery of the sky fluctuations on realistic mocks. We demonstrate that residual flat field errors comprise $< 20\%$ of the final *CIBER* power spectrum uncertainty with this methodology.

2.1 Introduction

The extragalactic background light (EBL) is the integrated light from all sources outside of the Milky Way, emitted over cosmic history. The spectral and spatial characteristics of the EBL promise a wealth of information on the astrophysical processes that drive cosmic light production. However, at optical and near-infrared (NIR) wavelengths, both the measurement and interpretation of the EBL have not converged, with disagreements between various methods [63, 64, 65, 66, 67, 68, 69, 70].

Fluctuation-based measurements of the EBL bypass the conventional challenge of absolute photometric measurements, namely degeneracy with zodiacal light (ZL), taking advantage of its smoothness on large scales as measured at infrared and mid-

IR wavelengths [71, 72]. However, fluctuation measurements are sensitive to other systematics which need careful treatment. In our previous work ([73], hereafter Z14) we measured fluctuations at 1.1 μm and 1.6 μm using imaging data from the second and third flights of *CIBER-1* (the first generation of *CIBER*), revealing fluctuations on angular scales $\theta > 5'$ with an amplitude exceeding that expected from integrated galactic light (IGL), although without accounting for non-linear clustering [74].

In this work we improve upon the methodology in Z14 by focusing on two leading effects. The first involves corrections for the relative per-pixel gain within each imager, commonly known as the flat field gain (denoted FF). Contrary to Z14, which relied on field differences to mitigate FF errors, we directly estimate the FF gain from the science field observations with a stacking estimator and develop a pseudo-power spectrum formalism that quantifies and corrects for errors in the FF estimator error. The second effect involves masking bright stars and galaxies, which is required to reduce Poisson and clustering fluctuations from bright sources. Through ancillary optical and infrared photometry, we mask substantially deeper than by using 2MASS JHK$_s$ photometry alone while minimizing the fraction of masked pixels.

Monte Carlo simulations play an important role in our power spectrum pipeline. Accurate simulations of noise present in the *CIBER* maps (combined with realistic masking) allow us to estimate statistical errors and to correct for various noise biases on the power spectrum. By performing the same data processing on synthetic mocks as used for the observed data, we are able to validate our FF formalism and assess any remaining biases in the power spectrum pipeline. Our mock recovery tests on a large ensemble of synthetic *CIBER* observations allow us to estimate uncertainties and covariances in a comprehensive fashion and enable an assessment of field-to-field consistency in the observed *CIBER* data.

The paper is organized as follows. In §2 we introduce *CIBER* and describe the construction of synthetic mocks in §3, which include known astrophysical components, realizations of *CIBER* read and photon noise along with other observational effects. In §4 we describe the standard steps of the pseudo-power spectrum pipeline while in §5 we present the extended formalism that includes flat field errors. We then introduce a novel source masking procedure in §6 which includes several catalog-level validation tests. In §7 we apply our improved power spectrum pipeline to the mocks from §3, validating our ability to recover unbiased estimates of sky fluctuations and quantifying the impact of flat field errors on our power spectrum sensitivity. Lastly,

we conclude in §8 and discuss avenues for future development.

Throughout this work we assume a flat ΛCDM cosmology with $n_s = 0.97$, $\sigma_8 = 0.82$, $\Omega_m = 0.26$, $\Omega_b = 0.049$, $\Omega_\Lambda = 0.69$ and $h = 0.68$, consistent with measurements from *Planck* [75]. All fluxes are quoted in the Vega magnitude system unless otherwise specified.

2.2 Cosmic Infrared Background Experiment (CIBER)

CIBER[1] is a rocket-borne instrument [76] designed to characterize the NIR EBL through measurements of its spatial fluctuations and electromagnetic spectrum [77, 78]. In this work we focus on measurements using the *CIBER* imagers, simultaneously observing a 2×2 deg^2 field of view in two broad bands centered at 1.1 μm and 1.8 μm with 7″ pixels with wide-field refracting optics.

During the *CIBER*-1 flight integration campaigns, laboratory FF measurements were conducted using an integrating sphere for uniform illumination with a solar-type spectrum. However, these measurements were inconsistent with the FF estimates derived from flight exposures. We attribute the difference to systematic errors in the lab measurement (near field of the optics, non-uniformity of the sphere, residual spectral mismatch, etc.) `Z14` therefore analyzed field differences, and used the difference in the FFs between laboratory measurements and flight data to estimate the residual FF uncertainty.

CIBER-1 was flown four times in total, the first three from White Sands Missile Range in New Mexico, and the final non-recovered flight from NASA Wallops Flight Facility in Virginia. Unlike during the first three flights, the payload during the fourth flight achieved an altitude of 550 km (compared to ∼ 330 km), resulting in a longer total exposure time and lower levels of airglow contamination. Crucially, the higher number of science exposures from the fourth flight (five science fields compared to 2-3 for previous flights) enables the use of an improved in-flight FF stacking estimator for which per-pixel errors are sufficiently small, a condition we formalize in this work. Relaxing the requirement of field differences reduces the fraction of masked pixels in each map, which allows us mask more aggressively on individual fields, given access to sufficiently deep external catalogs.

[1] `https://ciberrocket.github.io/`

2.3 Simulations of CIBER observations

The synthetic observations described in this section serve to validate the power spectrum estimation pipeline and to estimate covariances which are then used against real datasets in Paper II. We generate mock sky realizations with a combination of point-source and diffuse clustering components. This includes random source realizations of galaxies and stars, diffuse galactic light (DGL), ZL, and EBL clustering fluctuations to match the observed power spectrum. Our synthetic observations match the characteristics of *CIBER*-1's fourth flight imaging dataset, including read and photon noise (including one read noise dominated science exposure with less than half the exposure time of the other four fields), pixel scale and point spread function (PSF). While the mock IGL and EBL clustering are taken to be statistically similar across the five *CIBER* fields, both the ZL and ISL vary due to the range of ecliptic and galactic latitudes spanned by the fields.

CIB galaxies

To generate IGL mocks we combine log-normal realizations of the matter density field with draws on empirical, redshift-dependent luminosity functions from [79] to assign galaxy positions. The log-normal mock technique has been developed for fast generation of galaxy catalogs and density fields [80, 81]. For a desired power spectrum C_ℓ, we first generate a field $G(x)$ from the log-normal power spectrum, denoted C_ℓ^G, and exponentiate the field, i.e., $\delta(x) \sim \exp[G(x)]$. To compute C_ℓ^G, we compute the angular two-point correlation function for C_ℓ using the Hankel transform

$$w(\theta) = \int \frac{\ell d\ell}{2\pi} C_\ell J_0(\ell\theta). \tag{2.1}$$

In the above equation, J_0 is a Bessel function of the zeroth kind and we have invoked the flat-sky approximation, which is valid on the scales considered. We compute the angular power spectrum C_ℓ from the projected non-linear, redshift-dependent 3D matter power spectrum using the Python version of `CAMB`. Next, we transform the angular two-point correlation function into the log-normal correlation function

$$w^G(\theta) = \log(1 + w(\theta)). \tag{2.2}$$

The log-normal correlation function is then converted back to an angular power spectrum using the inverse Hankel transform:

$$C_\ell^G = 2\pi \int \theta d\theta w^G(\theta) J_0(\ell\theta). \tag{2.3}$$

The log-normal power spectrum defines the diagonal component of the covariance matrix $\boldsymbol{C}_\ell^G$. Finally to generate the field $G(x)$, we draw a Gaussian realization from $\boldsymbol{\mathcal{N}}(0, \boldsymbol{C}_\ell^G)$ and compute its inverse Fourier transform, discarding the imaginary component. This is then exponentiated to obtain the nonlinear density field $\delta(x)$. To reliably simulate fluctuation modes on the scale of the *CIBER* fields, we generate realizations of $\delta(x)$ over a larger 4×4 deg^2 field after which we extract the central 2×2 deg^2 regions.

We then generate projected galaxy counts as Poisson realizations of each underlying density field, in which the mean number of galaxies per cell is set by the integrated number counts, i.e.,

$$N_{tot,i} = \int_{z_i}^{z_{i+1}} dz \frac{dN}{dz}. \tag{2.4}$$

We simulate independent IGL realizations in eight equally-spaced redshift bins between $0 \leq z \leq 2$. Note that these log-normal mocks do not include the effect of galaxy biasing which enhances the large-scale power from two-halo clustering; however, as the total observed *CIBER* fluctuations on large scales exceeds that from IGL by over an order of magnitude (with and without galaxy biasing), we do not incorporate a detailed biasing scheme in the IGL mocks. We refer the reader to [74] for the effects of galaxy bias and non-linear IGL clustering as predicted using MICECAT simulations.

We use the semi-empirical model from [79] to produce realizations of the IGL. The Helgason model constructs galaxy luminosity functions (LFs) assuming a Press-Schechter functional form fit to a complilation of observed LFs from existing measurements. Within each redshift bin we assign galaxy redshifts from the normalized dN/dz distribution, conditioned on the LF at each bin center:

$$\frac{dN}{dz} = \int \Phi(M|\hat{z}_i) dM. \tag{2.5}$$

Using these LFs we then draw apparent magnitudes for each source down to $m_{AB} = 28$, corresponding to different absolute magnitudes M_{abs}^{min} within each redshift bin.

The fits from [79] have uncertainties related to the faint end slopes of the LFs, from which they delineate "High Faint End" (HFE) and "Low Faint End" (LFE) model predictions. To assess the impact of these uncertainties on the predicted Poisson noise level we generate three separate sets of CIB mocks by varying α_0, the Schechter parameter that normalizes the faint end slope, i.e.,

$$\phi(M) dM \propto \phi^* \left(10^{0.4(M^*-M)}\right)^{\alpha(z)+1}, \tag{2.6}$$

where $\alpha(z) = \alpha_0(z/z_0)^r$ [79]. We use $\alpha_0 = -1.0$ (default), $\alpha_0 = -0.8$ (LFE) and $\alpha_0 = -1.2$ (HFE) and compute the corresponding power spectra with a range of masking depths. On small scales, the HFE models predict 15-25% higher power between $J = 17.0$ and $J = 18.5$ than the default model, while the LFE models predict $5 - 10\%$ less power. The models differ most on large scales, with HFE a factor of $1.5 - 2$ larger than the default model for $\ell < 1000$. We use the fiducial model prediction for the results in this work. This IGL model has the limitation that only single bands can be simulated, i.e., we cannot simulate or make predictions for cross-spectra across different wavelengths.

Zodiacal light

ZL refers to light reflected off interplanetary dust grains (IPD) within our solar system. Beyond Earth's atmosphere, ZL is the leading contribution to the intensity monopole at NIR wavelengths. [82] constrains the ZL contribution from DIRBE observations across $1.25 - 240$ μm with a precision of $\sim 1\%$, under the assumption that the IPD should be the only time-varying component on the celestial sphere. The ZL is generally known to be spatially smooth on scales $\theta \gtrsim 200''$, based on measurements from *Spitzer* [71]. We use a modified Kelsall model that accounts for the solar spectrum and ZL reddening to predict intensity over each *CIBER* bandpass [83]. The ZL intensity varies by a factor of 2.3 across the five *CIBER* fields, spanning ecliptic latitudes $11^\circ \leq \beta \leq 73^\circ$. In addition to the monopole, we inject ZL gradients into the mocks with amplitudes consistent with observations. However we give the gradients random directions, to capture the effect of the high-pass image filtering used in the pipeline.

Integrated Stellar Light

The raw *CIBER* clustering power is dominated by Poisson fluctuations from bright stars within our galaxy. We use the TRILEGAL model [84] to simulate realistic distributions of stars for each *CIBER* field. These are also used in [85] and are useful for testing the efficacy of astronomical masking in the maps and also for estimating the contribution of integrated star light below the masking threshold. On the scales and galactic latitudes probed in this work ($\theta \lesssim 2^\circ$), the angular distribution of the ISL is well approximated as uniform `Z14`. However, the ISL amplitude varies across the five *CIBER* fields, with the SWIRE field having $\sim 50\%$ higher stellar density compared to the mean of the fields.

Diffuse galactic light and EBL clustering components

We include an additional clustering signal from Gaussian realizations with power spectrum of the form $C_\ell = A\ell^{-3}$, where A is chosen to match the observed *CIBER* power spectrum in Z14. This component has the same spatial index as expected for DGL, an important foreground for intensity mapping at near-IR wavelengths. However, Z14 showed that the DGL component was small compared to the total signal. We simply model the total Z14 signal to properly estimate biases in power spectrum recovery, sample variance and covariance. The simulated Gaussian realizations neglect potential correlations between the EBL signal and simulated CIB galaxies.

Synthetic CIBER observations

To produce sky realizations consistent with *CIBER* we model the measured extended PSF from the fourth flight to populate point sources in each field. The measured PSF is a combination of the instrumental PSF and the pointing jitter and drift over each exposure. The *CIBER* PSF is estimated in [85] by stacking *CIBER* images on the positions of 2MASS sources, whose profiles are modeled empirically with a β model of the form

$$\mathrm{PSF}(r) = \left(1 + \left(\frac{r}{r_c}\right)^2\right)^{-3\beta/2}. \tag{2.7}$$

We evaluate the *CIBER* PSF at 100 sub-pixel positions and use these templates to inject sources into the maps at native resolution. Capturing the sub-pixel PSF is important for *CIBER*, which is designed to have an undersampled beam (the instrument PSF has a FWHM$\sim 9''$, slightly more than one *CIBER* pixel). Due to the relatively coarse *CIBER* angular resolution, we populate galaxies into the mocks as point sources and do not explicitly model the small number of extended extragalactic objects as these are identified and masked appropriately from catalogs. We do not include effects of PSF variation across the focal plane; however, this is an important effect in more precise measurements of small-scale ($\ell > 50000$) fluctuations.

Lastly, we add noise realizations to each sky realization. Our noise model is constructed from two components. The first is read noise from the detector and readout electronics, which we estimate from dark exposures obtained just prior to launch. Specifically, we generate Gaussian realizations with random phases from an underlying two-dimensional power spectrum describing each integration, which is necessary to capture the anisotropic structure of the read noise. The second is photon noise due to the Poisson statistics of sky signal incident on the *CIBER*

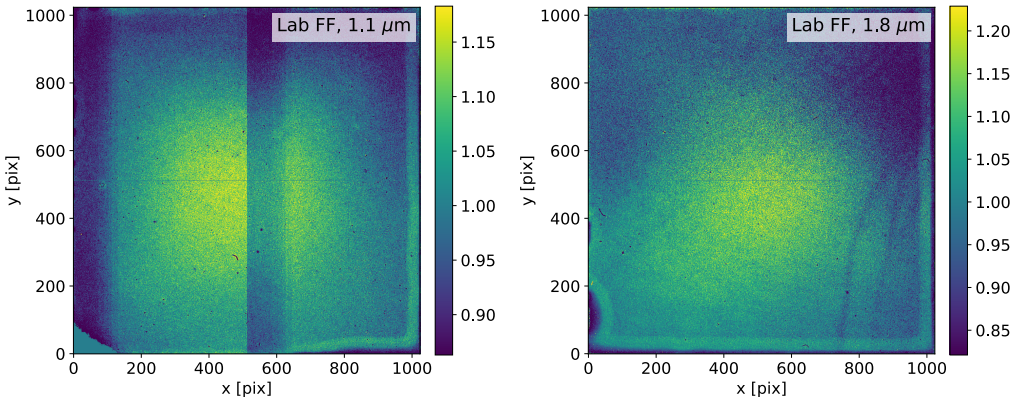

Figure 2.1: Laboratory FF measurements taken during the third *CIBER*-1 flight campaign. Structure in the *CIBER* FFs comes from a combination of optical and electrical effects. We use the laboratory data to inject a realistic FF into our mocks, which is then estimated and corrected for in our power spectrum recovery tests (see §2.7).

imagers. This is calculated per pixel for each sky realization using the following equation from [86]

$$\sigma_\gamma^2 = \frac{6}{5}\frac{i_{phot}}{T_{int}}\left(\frac{N^2+1}{N^2-1}\right), \tag{2.8}$$

where i_{phot} is the photocurrent, T_{int} is the integration time and N is the number of frames in each integration.

While our formalism for recovering sky fluctuations is insensitive to the exact shape of the true FF, we use the laboratory FF estimates from *CIBER*-1's third flight campaign to realistically modulate each sky signal (and its photon noise), after which we add read noise which does not depend on the FF. We show the laboratory flats for the 1.1 μm and 1.8 μm detectors in Fig. 2.1.

We show the signal and noise components that go into each *CIBER* mock observation in Figure 2.2 for 1.1 μm.

2.4 Power spectrum formalism

In this section we describe the steps to relate an underlying sky power spectrum to bandpower estimates of the observed pseudo power spectrum. For clarity we describe how this is calculated in the absence of FF errors, while in the following section we detail the relevant modifications given our FF stacking estimator.

The observed power spectrum can be expressed as the following:

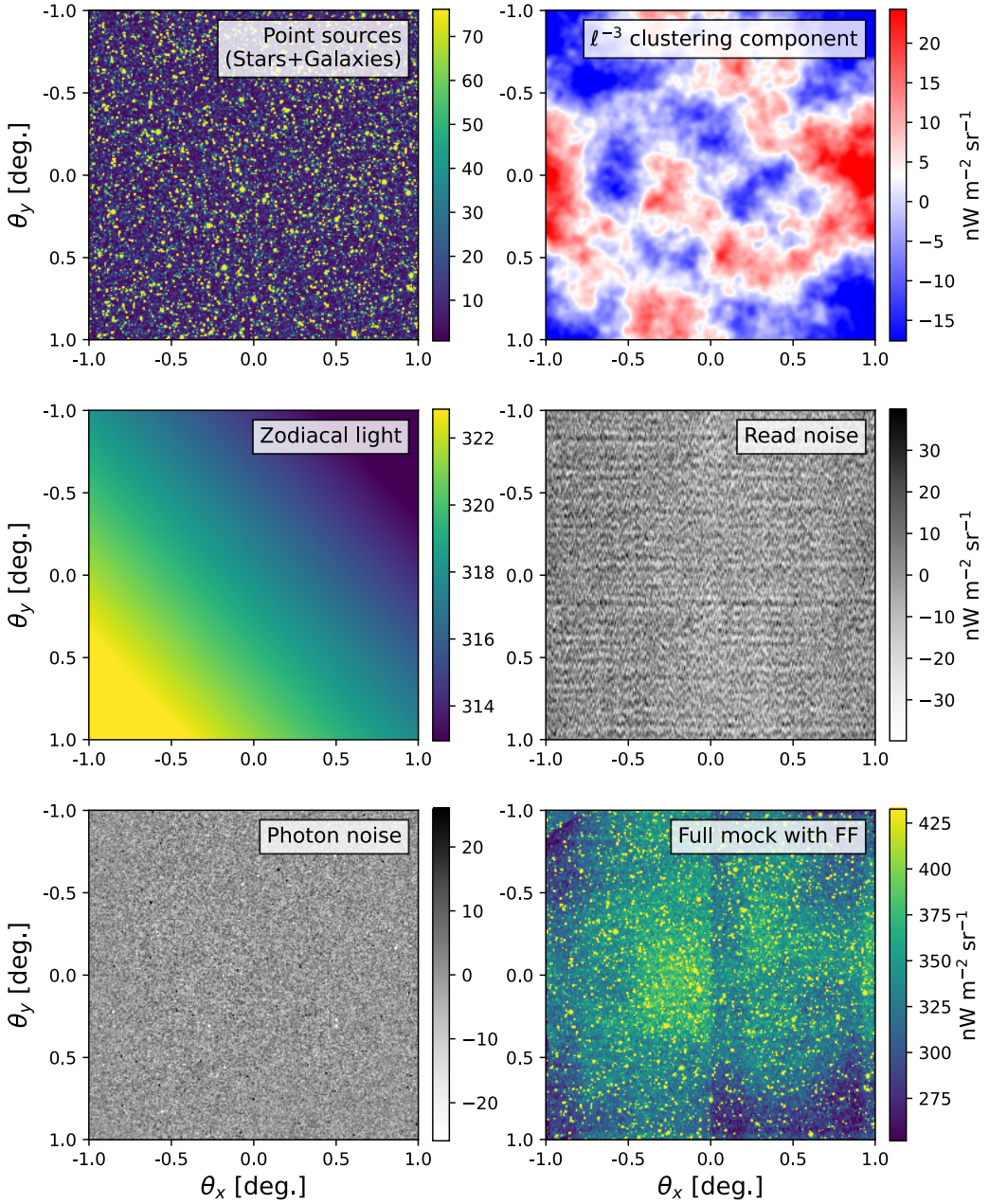

Figure 2.2: Different astrophysical signal and noise components that compose the mock *CIBER* observations used in this work, shown for 1.1 μm. These mocks are used to simulate power spectrum recovery on realistic observational data.

$$C_\ell^{obs} = \sum_{\ell'} M_{\ell\ell'}(B_{\ell'}^2 C_{\ell'}^{sky} + N_{\ell'}). \tag{2.9}$$

The sky signal with power spectrum C_ℓ^{sky} (shorthand for all astrophysical components) gets passed through the imaging process, which smooth the signal on the scale of the beam. The masking and filtering we employ couple large- and small-scale modes in the map, resulting in a transfer function that is expressed as a convolution in Fourier space through the mode mixing matrix $M_{\ell\ell'}$. In (2.9) B_ℓ is the beam transfer function, which is assumed to be diagonal in Fourier space, and N_ℓ is the noise bias (for cross-spectra the noise across instruments is in general uncorrelated and so $\langle N_\ell \rangle = 0$). To recover C_ℓ^{sky} we correct for the noise bias, apply a mode coupling correction $M_{\ell\ell'}^{-1}$ and then correct for the beam transfer function,

$$\hat{C}_\ell^{sky} = B_\ell^{-2} \sum_{\ell'} M_{\ell\ell'}^{-1}(\hat{C}_{\ell'}^{obs} - \hat{N}_{\ell'}). \tag{2.10}$$

While we use ℓ, ℓ' for presentation, in practice there is a binning operator $P_{b\ell}$ that averages the modes into bandpowers b,

$$\hat{C}_b = \sum_{\ell \in \mathcal{B}} P_{b\ell} \hat{C}_\ell, \tag{2.11}$$

where $\mathcal{B}$ denotes the set of modes ℓ' satisfying $\ell_b^{min} \leq \ell' < \ell_b^{max}$. For uniform weighting the binning operator is

$$P_{b\ell} = \begin{cases} \frac{1}{\ell_b^{max} - \ell_b^{min}}, & \ell \in \mathcal{B} \\ 0, & \ell \notin \mathcal{B} \end{cases} . \tag{2.12}$$

Noise bias subtraction

The power spectrum noise bias can be expressed as the following:

$$N_\ell = \sum_{\ell'} M_{\ell\ell'}(N_{\ell'}^{read} + N_{\ell'}^{\gamma}). \tag{2.13}$$

Note that the per-pixel photon noise $N_{\ell'}^{\gamma}$ depends on the beam-convolved sky maps. For each field we estimate N_ℓ using an ensemble of 500 independent noise realizations combined with the respective image mask. This is typically done through draws of read and photon noise alone, however as discussed in §2.5 there are additional contributions to each field's noise bias from the FF errors generated by instrumental noise in the fields used to estimate its respective FF estimate.

Fourier weighting

Detector read noise power in the maps can be mitigated by recognizing that certain 2D Fourier modes contribute significantly more noise power than others. However, masking convolution can mix concentrated power across to other modes. By calculating the per-mode variance of our Monte Carlo noise realizations from the previous sub-section, we derive inverse variance weights to the two-dimensional power spectrum which we apply before computing azimuthally-averaged bandpowers,

$$\langle C_\ell \rangle = \frac{\sum_{(\ell_x,\ell_y)} w(\ell_x,\ell_y) C(\ell_x,\ell_y)}{\sum_{(\ell_x,\ell_y)} w(\ell_x,\ell_y)}, \tag{2.14}$$

where $C(\ell_x, \ell_y)$ is the two-dimensional observed power spectrum. There is a trade off between down-weighting noisy Fourier modes and reducing the effective sample size of modes contributing to each azimuthally averaged band power. Nonetheless this scheme works well in our limit where there is a large dynamic range of noise power distributed across a relatively small set of modes.

Beam correction

The *CIBER* PSF smoothes the observed sky signal, resulting in a roll-off in power on small scales. Using the best-fit beam model for each field, we generate a 10×10 grid of sub-pixel centered PSFs, which are downsampled to the *CIBER* native pixel resolution. We then compute B_ℓ^2 as the mean power spectrum of the 100 sub-pixel PSFs.

Mode coupling correction

The application of instrument and astronomical masks on the *CIBER* maps means that Fourier modes on the underlying sky will mix with one another. Following the MASTER formalism [87], we estimate each $N_{bp} \times N_{bp}$ band power mode coupling matrix by applying the target field mask to Monte Carlo tone realizations with random phase through the target mask, and compute the corresponding pseudo-power spectra. We do not apply any apodization to the images. In this work we fix the number of logarithmically-spaced band powers to $N_{bp} = 25$. For each band power we compute $N_{sim} = 500$ phase realizations, where N_{sim} is chosen to be large enough such that statistical errors on $\hat{M}_{\ell\ell'}$ are negligible. After taking the expectation across realizations, we correct the observed, noise-debiased power spectrum by applying the inverse mode coupling matrix $M^{-1}_{\ell\ell'}$.

Image filtering

The *CIBER* maps have large-scale variations which need to be filtered out before computing power spectra. The first are array-scale gradients which come from Zodiacal light and other foreground components. The second involves a quadrant-specific detector effect we identified in both laboratory and flight data. In some quadrants, we observe a form of two-state noise, in which the ADU levels in all pixels fluctuate coherently across consecutive readout frames. This leads to a variation in the resulting slope fits across different quadrants. We correct for these effects by fitting a linear combination of per-quadrant offsets and a gradient across the full array,

$$G(x, y) = Ax + By + \sum_{k=1}^{4} Q_k O_k(x, y), \tag{2.15}$$

where A and B are the gradient parameters, Q_k is the offset parameter for quadrant k and $O_k(x, y)$ is a 2D step function equal to one for pixels in quadrant k and zero otherwise. We calculate the best-fit parameters θ from the least-squares solution to unmasked pixels $\vec{K}$ with the Moore-Penrose pseudo-inverse,

$$\vec{\theta} = (X^T X)^{-1} X^T \vec{K}. \tag{2.16}$$

A common ansatz is to treat the filtering transfer function as a separate quantity T_ℓ in (2.9) which is corrected after mask de-convolution. However, this neglects potential mode coupling induced by the image filtering. As an example, in Figure 2.3 we show the 1D transfer functions for gradient-only and gradient + offset filtering. These are derived from 1000 Monte Carlo signal realizations that include both large- and small-scale power, for which we calculate the ratio of power spectra before and after filtering. To check for potential dependence on the input clustering, we test T_ℓ with three separate spatial indices ($\gamma \in \{-2, -2.5, -3\}$ where $C_\ell \propto \ell^\gamma$ on large scales) and find estimates of T_ℓ that are consistent within statistical uncertainties. In both cases there is a suppression of power in the lowest bandpowers as well as an increase in power for intermediate bandpowers. In the next section we show that this leads to off-diagonal mode coupling, leading us to incorporate image filtering into our mode mixing matrix estimation procedure.

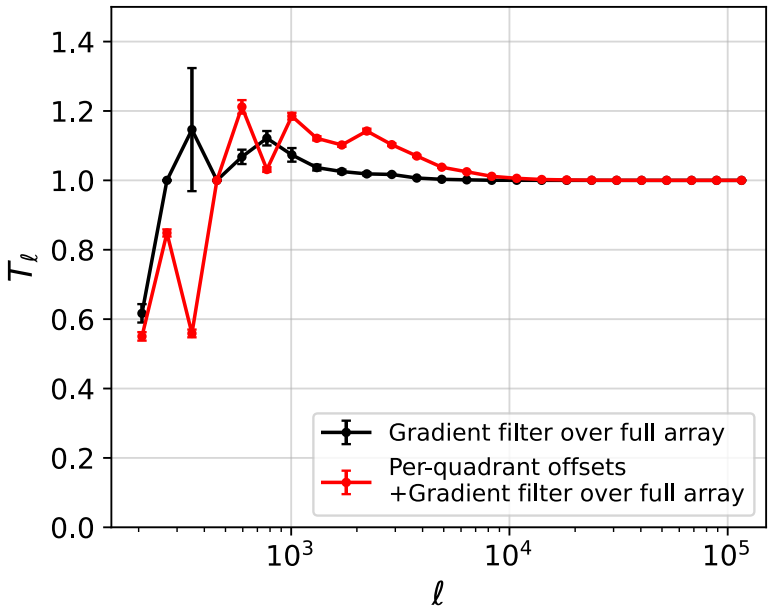

Figure 2.3: Comparison of filtering transfer functions. Despite the harsher transfer function of per-quadrant filtering (red), we use it in our fiducial results to mitigate detector effects that imprint noise on large scales.

2.5 Power spectrum biases from flat field stacking estimator

FF estimation and image filtering

The true sky signal, denoted I_{sky}, can be expressed in terms of a mean intensity $\bar{I}_{sky}$ and a general fluctuation component S,

$$I_{sky} = \bar{I}_{sky} + S. \tag{2.17}$$

The FF gain across each detector array is defined as the relative response to a uniform illuminating surface, which in practice is determined by detector effects (e.g., variations in per-pixel quantum efficiency), as well as the optical/mechanical configuration of the instrument, which can introduce effects such as vignetting. Described by a scalar field $FF(x, y)$, the flat field modulates the incident sky signal

$$I_{sky}(x, y) \rightarrow FF(x, y) I_{sky}(x, y). \tag{2.18}$$

In the absence of a well-determined FF, Z14 used field differences to mitigate errors in the FF at leading order, following the fact that FF errors primarily couple to the mean intensity of each map:

$$\delta I_{A-B} = \delta[\hat{FF}](I_A^{obs} - I_B^{obs}) \tag{2.19}$$

$$\approx \delta[\hat{FF}](\bar{I}_A - \bar{I}_B). \tag{2.20}$$

Field differences also make power spectrum de-convolution more difficult due to higher masking fractions (the effective mask is the union of individual field masks), degrading statistical sensitivity and limiting the achievable masking depth for point sources. This penalty is especially pronounced when comparing to *Spitzer* data in a single field.

The alternative approach which we pursue in this work is to estimate the FF from exposures taken during flight, stacking the FF images derived from each field, using the mean background (which is dominated by ZL) as an approximate uniform illuminator, i.e., for field i,

$$\hat{FF}_i(x,y) = \frac{I_i^{obs}(x,y)}{\bar{I}_i^{obs}}. \tag{2.21}$$

While the astrophysical signals will vary across images, the FF responsivity will not, and so in the limit of many independent exposures this estimator should converge to the true FF,

$$\lim_{N_f \to \infty} \langle \hat{FF}(x,y) \rangle = FF_{true}(x,y). \tag{2.22}$$

We model the FF error in unmasked pixels from a single field as

$$\delta\hat{FF}_i(x,y) = \frac{FF(x,y)(S_i + \epsilon_{\gamma,i}) + \epsilon_{read,i}}{\bar{I}_i^{obs}} \tag{2.23}$$

where S_i is the underlying sky signal, ϵ_γ is the photon noise and ϵ_{read} is the read noise.

The FF estimate for each target field j comes from stacking the $N_f = 4$ other science fields ("off-fields") in a round-robin approach,

$$\hat{FF}_j(x,y) = \frac{\sum_{i=1}^{N_f} w_i \hat{FF}_i}{\sum_i w_i}. \tag{2.24}$$

We apply field weights w_i that account for both photon and read noise to construct a minimum variance per-pixel FF estimate. The loss of pixels due to masking in the FF images means each field does not contribute an FF estimate for all pixels. There is a small ($\sim 1\%$) fraction of pixels that by chance have zero unmasked FF_j estimates, in which case we mask these pixels in the jth science field. To avoid non-linear effects sourced from large FF errors, we additionally mask pixels with FF estimates that deviate by $> 3\sigma$ from the mean local FF estimate, which affects $1-2\%$ of pixels depending on the field.

On large scales the FF stacking estimator is partially degenerate with large-scale ZL sky gradients across the arrays. As a result, gradients in the stacked fields will bias the FF on large scales. We implement an iterative procedure to compute the FFs and gradient+offset components in each map. For iteration n,

1. Calculate the stacked FF estimates for each field j:
$$\hat{FF}_j^n = \sum_{i \neq j} \frac{w_i I_i}{\bar{I}_i} / \sum_{i \neq j} w_i. \tag{2.25}$$
2. Fit $G(x, y)$ from (2.16) to each corrected field.
3. Subtract the gradient and relative offsets $\{Q_k - \langle Q_k \rangle\}$ from the FF-corrected image.
4. Multiply the subtracted image by $\hat{FF}_j^n$ to obtain I_j^{n+1}.
5. Go to (1) and repeat.

Convergence is achieved after a few ($\lesssim$5) iterations as evaluated by the best-fit gradient parameters. The final gradient subtracted by the end of several iterations is typically within 10% of the single pass estimate, while the per-pixel FF differs by less than one per cent. By performing the FF stacking on mocks, we find that the per-pixel error RMS is $\delta FF \sim 10\%$ for both imagers, which is driven by instrumental noise errors. Due to our round robin stacking approach, each field has a different FF error distribution, which depends on the off-field instrumental noise and the target field's mean intensity.

Power spectrum bias

We calculate the pseudo-power spectrum of each corrected image by expressing $\hat{FF}^j$ in terms of the true FF and its error, such that

$$\frac{I_j^{obs}}{\hat{FF}^j} = \frac{FF_{true}\left[I_j^{sky} + \epsilon_{\gamma,j}\right] + \epsilon_{read,j}}{FF_{true}\left[1 + \frac{\delta[\hat{FF}^j]}{FF_{true}}\right]}. \tag{2.26}$$

Taking advantage of the fact that $\delta[\hat{FF}^j]/FF_{true} \ll 1$, we Taylor expand (2.26) and in App. .1 arrive at the following approximation for the corrected map,

$$\frac{I_j^{obs}}{\hat{FF}^j} \approx \bar{I}_j^{sky} + S_j + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j} - \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}(\bar{I}_j^{sky} + S_j + \epsilon_{\gamma,j}). \tag{2.27}$$

The FF error terms in the second line above introduce additional biases on the power spectrum of sky fluctuations. The FF error coupled with the mean sky brightness of the target field, $\delta[\hat{FF}^j]\bar{I}_j^{sky}$, sources the majority of the fluctuation bias. While we simulate and correct for the point source contributions to $\delta[\hat{FF}^j](S_j + \epsilon_{\gamma,j})$, these comprise $< 1\%$ of the total noise bias in each field and have a negligible impact at *CIBER* sensitivity.

From (2.23) and (2.27) we see that the power spectrum FF bias depends on both instrument noise and sky fluctuations. While the instrument noise contribution yields an additive bias, the sky fluctuations act as a multiplicative bias. Another conclusion is that the amplitude of the PS bias depends on the relative mean sky brightnesses across fields, such that fields with higher ZL have larger biases.

Modified noise bias

The full noise bias is expressed as the following,

$$N_\ell = \sum_{\ell'} M_{\ell\ell'}(N_{\ell'}^{read} + N_{\ell'}^{\gamma} + N_{\ell'}^{\delta FF}). \tag{2.28}$$

To include $N_{\ell'}^{\delta FF}$ within the Monte Carlo procedure in §2.4, we add mean sky levels to each noise realization and apply the FF stacking estimator to each set of five maps before calculating their mean-subtracted power spectra. This Monte Carlo approach captures noise biases beyond our Taylor-approximated expression in (2.27), e.g., terms $O(\delta^2[\hat{FF}])$, however these terms are small and in the limit of large FF errors there are additional signal $\times$ noise cross-terms which are not taken into account.

Mode coupling correction with FF errors

In the absence of masking, FF errors manifest as a scalar multiplicative bias on the sky power spectrum (see .1), which ranges from 25% to $> 100\%$ and is driven by variations in ZL intensities across the five *CIBER* fields. However, an important insight of this work is that the FF stacking estimator couples modes from all masked fields with the mask of each target field. This can be seen by writing the masked version of (2.27) (see App. .1) and modifying the expression for $\delta[\hat{FF}]$. In the limit where $\delta FF \ll 1$, this additional mode coupling can be approximated with linear operators and treated within the standard pseudo-C_ℓ formalism. Through a modified Monte Carlo procedure (also detailed in Appendix .1), we capture the combined mask+FF+filtering mode mixing in a single matrix transformation. This

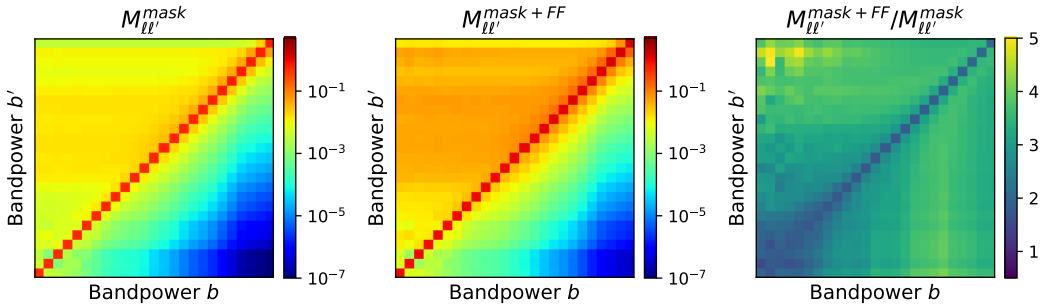

Figure 2.4: Standard mode coupling matrix derived from flight mask (left), along the hybrid mask-flat field matrix (middle) and ratio of the two (right). These highlight the impact of flat field errors, which modify both the diagonal and off-diagonal components of the mask-only mode coupling matrix.

is similar in spirit to [88], which incorporates the combined mode coupling of TOD filtering and survey masks into a single matrix transformation $J_{\ell\ell'}$.

To compare the mode couplings introduced by our data preprocessing, in Fig. 2.4 we compare three versions of the mixing matrix for a single field: one derived using the mask only, one using the mask and FF errors, and one with masking+FF errors+image filtering. The three matrices share similar structure, in the sense that power at low-ℓ tends to move to higher bandpowers. At high ℓ, $M_{\ell\ell'}$ tends to be closer to symmetric and there is very little leakage of high frequency power to large scales. Both hybrid mixing matrices exhibit stronger off-diagonal components that reflect leakage of large-angle modes to smaller scales. By taking the ratio between matrices we find that the off-diagonal elements are over three times higher than the mask only case and over five times higher for certain bandpower combinations. We find that image filtering has a large impact on the resulting mode coupling, in particular for low-ℓ bandpowers where the per-quadrant offset fitting introduces a form of ringing in Fourier space.

Cross-power spectrum measurements

A similar correction is required for *CIBER* × *CIBER* cross-power spectrum estimates. While FF errors from instrumental noise will be uncorrelated across bands (i.e., $\langle N_\ell^{\delta FF} \rangle = 0$), the common fluctuation power between bands will source coherent FF errors across bands, leading to a multiplicative bias. This can be captured by folding in the sky intensities of both bands (and the weights used to stack per-field FF estimates) into the Monte Carlo $M_{\ell\ell'}$ estimation procedure.

2.6 Masking deeper in the NIR with multi-wavelength photometry

Having laid out our power spectrum formalism in tne presence of flat field errors, we now turn to a practical analysis challenge, namely effective source masking. For clean measurements of large-angle clustering, source masking reduces the effective shot noise level associated with Poisson fluctuations. While masking deeper removes Poisson fluctuations, there is a trade off with minimizing the fraction of masked pixels. However, by bypassing the need for field differences, we avoid a large penalty in the masking fraction and resulting mode coupling. This is because the masking fraction in field differences is determined by the instrument mask and union of two independent astronomical source masks, meaning uncertainties due to reduction of effective modes and de-convolution of masked power spectra are more severe. Single-field imaging thus enables a more aggressive masking of sources.

Beyond mask de-convolution, effective source masking in Z14 was limited by external catalog completeness. In Z14 source masks were constructed using J-band catalogs from 2MASS to a depth of $J = 17.5$. However, the 2MASS completeness falls quickly beyond $J = 16.0$ and $H = 15.0$. While deeper NIR catalogs such as UKIDSS and IBIS exist, their coverage within the *CIBER* fields is highly non-uniform and unavailable in some fields. Another approach would be to over-mask, using sources identified in external catalogs at other wavelengths, e.g., from optical surveys. However the masking would be very inefficient, since it is not clear *a priori* which optically identified sources correspond to the brightest J- and H-band sources.

For our approach, we take advantage of the fact that deeper multi-band optical and infrared photometry from PanSTARRS and WISE contain sufficient information to determine cuts on J- and H-band source magnitudes. Rather than construct hand-crafted color cuts, we make direct J- and H-band magnitude predictions through random forest regression. We then use these predictions to estimate the infrared flux to set the size of masks surrounding the identified point source. In the following sub-sections we summarize the properties of the external catalogs used to train and test our random forest model.

Source catalogs

In this work we use direct NIR photometry from the 2MASS and UKIDSS catalogs along with optical and infrared photometry from PanSTARRS and WISE. With the exception of UKIDSS, these catalogs have full coverage across all five of the *CIBER*

Survey	Filters	5σ point source depth
PanSTARRS	grizy	23.4, 23.0, 22.7, 21.8, 20.7
2MASS	JHK	17.0, 16.3, 15.5
UKIDSS (LAS)[2]	JHK	18.7, 17.4, 16.3
UKIDSS (UDS)	JHK	24.7, 23.7, 23.4
unWISE	W1	17.5

Table 2.1: List of ancillary catalogs used in this work and their properties. We use 2MASS photometry to mask bright sources in our fields ($J < 16$), while for fainter sources we use a combination of PanSTARRS and unWISE photometry to predict NIR magnitudes using a model trained on the UKIDSS UDS catalog (see §2.6). All listed depths are in the Vega magnitude system.

4th flight science fields. We summarize the depths of these catalogs in Table 2.1.

2MASS

The Two Micron All Sky Survey [2MASS; 89] imaged the sky in J (1.2 μm), H (1.6 μm) and K (2.1 μm) bands using 1.3-meter telescopes at Mt. Hopkins and CTIO, Chile. The extended 2MASS catalog is 75% complete in integrated counts down to $J = 17.5$ (Vega), or 17.5 (18.4 AB) and 17.0 (18.4 AB) for *CIBER*'s 1.1 μm and 1.8 μm bands, respectively. In this work 2MASS is used to identify sources with $J < 16$. For very bright sources, the 2MASS point source catalog uses measurements from shorter integrations, either 1.3 second exposures (`rdflg` = 1) or 51 ms from the array reset for the brightest sources (`rdflg` = 3). Across the five *CIBER* fields (elat10, elat30, Boötes B, Boötes A and SWIRE), there are $\{24, 22, 24, 22, 44\}$ stars with `rdflg` = 1 while only two stars in elat30 and one in SWIRE have `rdflg` = 3.

Pan-STARRS

The Panoramic Survey Telescope and Rapid Response System [PanSTARRS; 90] is a system designed for wide-field astronomical imaging . The 1.8-meter telescope, situated on Haleakala in Maui, has a 1.4 Gigapixel camera with 7 deg^2 field of view, and has imaged the sky in five broadband filters (g, r, i, z, y). The primary 3π survey covers 3×10^4 deg^2, with full coverage over the *CIBER* fields. We query source positions and magnitudes in these bands from the DR2 `MeanObject` table, including all sources with y-band measurements and quality flags (`qualityFlag`) in the `ObjectThin` table equal to 8 or 16. PanSTARRS is a desirable catalog for our purposes given its relatively deep y-band photometry, which more strongly correlates with J- and H-band fluxes.

unWISE

The unWISE catalog consists of photometry from unblurred coadds from WISE imaging [91, 92]. We use the five-year catalog, which at 3.4 μm has a 5σ depth of $W1 = 17.5$ ($W1_{AB} = 20.8$). The performance using both $W1$ and $W2$ photometry was comparable to that from $W1$ alone, so we opt for the latter.

UKIDSS

The UKIRT Infrared Deep Sky Survey [UKIDSS; 93] consists of seven years of imaging in the near-infrared with varying depths, carried out using the UKIRT Wide Field Camera (WFCAM). The deepest coverage available covers 0.77 deg^2 in the Ultra Deep Survey (UDS) field down to $J = 24.7$, $H = 23.7$ and $K_s = 23.4$. We use the UKIDSS UDS photometry to train our random forest regression method, which is detailed in the next sub-section. The UKIDSS Large Area Survey (LAS) is considerably shallower than UDS (depth of $K \sim 18$) but has available J- and H-band photometry for two of the five *CIBER* fields (elat10 and elat30). Lastly, the UKIDSS Deep Extragalactic Survey (DXS) reached a depth of $K \sim 21$ and covers the majority of the SWIRE field for J band. The LAS and DXS photometry are used to validate our predicted science field catalogs where available.

Random forest model and training

We perform random forest regression training and validation on UKIDSS photometry in the UDS field, which probes significantly deeper than our desired masking depths. We cross-match the UDS catalog with overlapping unWISE and PanSTARRS catalogs with a matching radius of 1″. When a source lacks a PanSTARRS or unWISE detection, the missing magnitudes are replaced with $m = 30$, i.e., they are labeled as non-detections.

We split 70% and 30% of the cross-matched catalog to form our training and validation samples, respectively. We restrict our training and validation samples to sources with $J < 21$ to prevent training set imbalance relative to our target *CIBER* masking depths. We use the publicly available package `sklearn` to train separate random forest models for J- and H-band predictions. We set the maximum depth of the decision trees to eight, beyond which the regression performance plateaus.

In Figure 2.5 we compare our predicted J- and H-band magnitudes with those from our UKIDSS training and validation sets. Our results are unbiased on average, with increasing dispersion toward fainter magnitudes. There is a small number

of catastrophic outliers, which typically correspond to sources with incomplete multi-band coverage, e.g., unWISE only (red points) or PanSTARRS only (blue points). The error RMS for the sub-samples, indicated in Fig 2.5, are smallest when both optical and infrared photometry are available and largest when only infrared photometry is available.

We summarize the completeness and purity of the derived UDS masking catalogs in Table 2.2 for a range of masking thresholds. The total number of predicted sources below each magnitude threshold agrees well with those from our validation samples and the completeness and purity of our predicted catalog is $> 90\%$ in all cases, with some degradation toward fainter fluxes. As expected, masking predictions for sources with both optical and infrared counterparts have the highest completeness and purity. In comparison, the completeness of 2MASS catalog in the same field falls off quickly, starting at 89% and 82% for $J < 17.5$ and $H < 17.0$, respectively, and going down to 34% and 27% for $J < 19.0$ and $H < 18.5$. While incompleteness in our masking catalog leads to additional point source fluctuation power, the main penalty of catalog impurity is a slightly higher masking fraction.

Testing mask predictions with COSMOS

To assess any systematic uncertainties due to distribution shift between our UDS training set and the science fields, we apply the model to multi-band photometry from the COSMOS 2015 catalog [94]. One subtlety is that the optical and infrared photometry in COSMOS come from Suprime-Cam and IRAC rather than PanSTARRS/WISE, adding another layer of distribution shift beyond our application in the science fields. Nonetheless we use our pretrained model to predict J and H band magnitudes and compare these against measured magnitudes from the COSMOS catalog.

We include the distribution of predicted vs. measured magnitudes in the right column of Fig. 2.5. Due to saturation in some optical bands for the COSMOS catalog, our results are limited to $J > 16$, which is our main focus in any case. Our predictions match the COSMOS15 measured magnitudes closely for a range of fluxes, however there is larger scatter and a mild negative bias on the predicted magnitudes. For sources down to $J < 18.5$ and $H < 17.0$, the error RMS for each band is 20%/30% higher than that of the UDS validation sets, which corresponds to lower purity in the test set results. This may be due to differences in source photometry across catalogs or calibration discrepancies between PanSTARRS/Suprime-Cam and WISE/IRAC.

Masking depth Vega	N_{src} (UKIDSS)	N_{src}, C, $\mathcal{P}$ all, predicted	C, $\mathcal{P}$ PS+unWISE	C, $\mathcal{P}$ PS only	C, $\mathcal{P}$ unWISE only	C, $\mathcal{P}$ 2MASS only	$\delta C_\ell^{SN}/C_\ell^{SN}$ C15 test set
$J < 17.5$	370	(371, 0.98, 0.98)	(0.98, 0.98)	(0.90, 1.0)	(0.0, 0.0)[a]	(0.89, -)	-5.2%
$J < 18.0$	498	(481, 0.93, 0.97)	(0.95, 0.98)	(0.88, 0.87)	(0.16, 0.33)	(0.65, -)	-10.2%
$J < 18.5$	684	(640, 0.92, 0.98)	(0.95, 0.99)	(0.82, 0.97)	(0.57, 0.88)	(0.47, -)	-12.2%
$J < 19.0$	977	(947, 0.93, 0.95)	(0.96, 0.96)	(0.88, 0.92)	(0.45, 0.9)	(0.34, -)	-21.7%
$H < 17.0$	422	(416, 0.96, 0.97)	(0.97, 0.98)	(0.79, 0.85)	(0.33, 0.25)	(0.82, -)	-6.9%
$H < 17.5$	574	(544, 0.91, 0.97)	(0.93, 0.98)	(0.81, 0.83)	(0.33, 0.60)	(0.58, -)	-2.2%
$H < 18.0$	828	(829, 0.94, 0.94)	(0.98, 0.94)	(0.73, 0.92)	(0.83, 0.68)	(0.40, -)	-8.8%
$H < 18.5$	1206	(1167, 0.90, 0.93)	(0.94, 0.95)	(0.84, 0.84)	(0.54, 0.73)	(0.27, -)	-16.4%

Table 2.2: Random forest regression performance on our UDS validation set, for various masking selections. The table shows the total number of sources brighter than each masking threshold from UKIDSS, along with the completeness (C) and purity ($\mathcal{P}$) of the predicted catalogs. We also include the mean fractional power spectrum bias on Poisson fluctuations ($\ell > 10000$), which we estimated by applying our pre-trained model on UDS to the COSMOS 2015 catalog and computing the true vs. estimated fluctuation power of sources fainter than each magnitude cut. The negative signs indicate that the masking method removes more Poisson fluctuations than an ideal mask at the specified J- or H-band magnitude.

[a]There is only one source which satisfies this condition in our validation set.

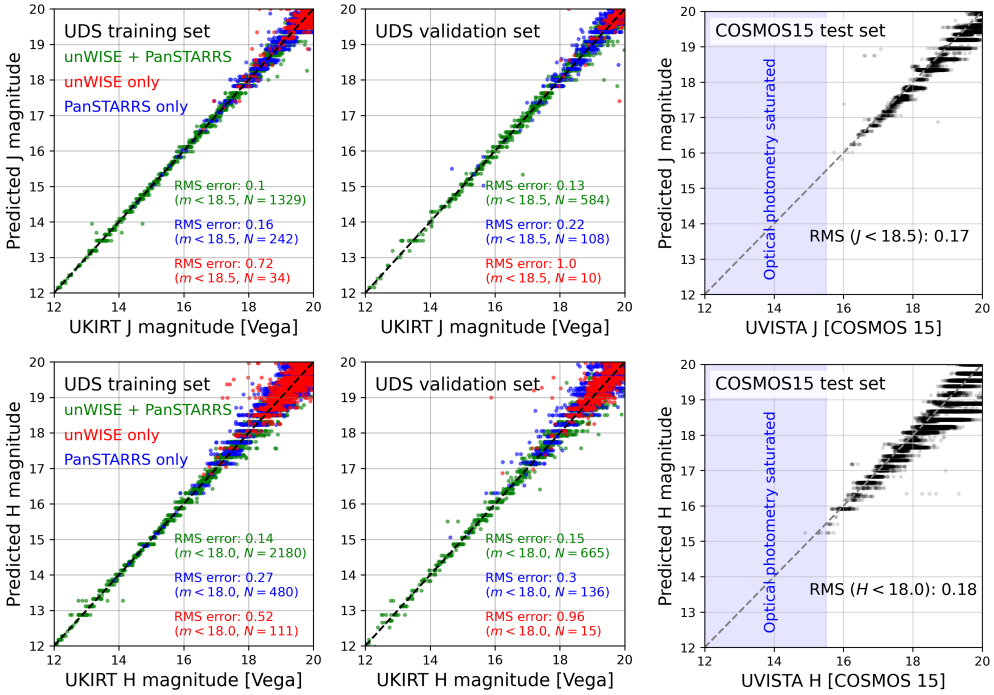

Figure 2.5: Comparison of measured magnitudes and random forest-predicted magnitudes using ancillary photometry, for J-band (top row) and H-band (bottom). The left and middle columns show our results for UKIDSS training and validation sets, respectively. UKIDSS sources with both unWISE and PanSTARRS counterparts are plotted in green, while those with only unWISE or only PanSTARRS are plotted in red and blue, respectively. The right-hand column shows the results of applying our models trained on UKIDSS to COSMOS 2015 photometry, as a test of distribution shift. The COSMOS field used Suprime-Cam and IRAC for the optical and IR data.

COSMOS15 is more complete in the optical/IR than the PanSTARRS/WISE catalog, leading to well-determined photometry for some sources that would be otherwise labeled as non-detections in the training set.

Given our predicted masking selections, we then calculate the sub-threshold shot noise of sources and compare against the "true" shot noise at fixed masking depth. These quantifications are included in Table 2.2. Our results suggest a slight over-removal of point source power in the predicted catalogs, though the fractional difference in power is small ($< 13\%$ for $J < 18.5$ and $< 9\%$ for $H < 18.0$). For the deepest masking depths ($J < 19.0$ and $H < 18.5$) the departures are slightly larger (22% and 16% for J- and H-band, respectively). A full interpretation of these discrepancies needs to take into account systematic differences between the training and test sets, however in general these results suggest that our source

masking procedure is robust, extending over two Vega magnitudes deeper than through 2MASS alone.

Recovered source counts

Figure 2.6 shows the J- and H-band cumulative number counts for sources in our five *CIBER* science fields as recovered by different catalogs. We limit our comparisons to $J > 12$, the brightest magnitude available in our UDS catalog. The 2MASS catalogs are in broad agreement with both our derived catalogs and those from UKIDSS down to $J = 16.5$ and $H = 16$. Beyond these depths the 2MASS catalog becomes incomplete, which is quantified in Table 2.2. In both bands, our predicted counts for the elat10 and elat30 fields are consistent with UKIDSS LAS to within 5% for $J < 18.5$ and $H < 18$. In the SWIRE field, we see larger differences, with our predicted integrated counts higher than UKIDSS by 20-35% going from $J = 16$ to $J = 18.5$.

For the brightest sources ($12 \leq m \leq 15$), our predicted catalogs exhibit larger discrepancies with respect to 2MASS and UKIDSS. We attribute this to the small training set of bright sources within the UDS field. Rather than develop a larger training set to train our random forest model, we simply merge the bright end of the 2MASS catalog ($J < 16$ and $H < 15$) with our random forest-derived catalogs to obtain our final masking catalogs.

Masking catalog consistency with simulations

In Figure 2.7 we compare the cumulative magnitude distributions of our final masking catalogs with those from predicted from simulations. The simulated catalogs combine the TRILEGAL stellar model for each field with realizations of the [79] galaxy model. We find that for $16 < m < 18.5$, our final counts are slightly higher than simulations. For the four non-SWIRE fields the counts are consistently higher by 10-20%, while for SWIRE our counts are 30-40% higher. The discrepancy with SWIRE is of similar magnitude to that seen between our predicted catalogs and UKIDSS DXS for J band, suggesting potential errors in our predicted catalog. The counts of our IGL mocks are constrained by the [79] best-fit luminosity functions, which appear to underestimate measured counts in this magnitude range (c.f. Fig. 12 of [79]).

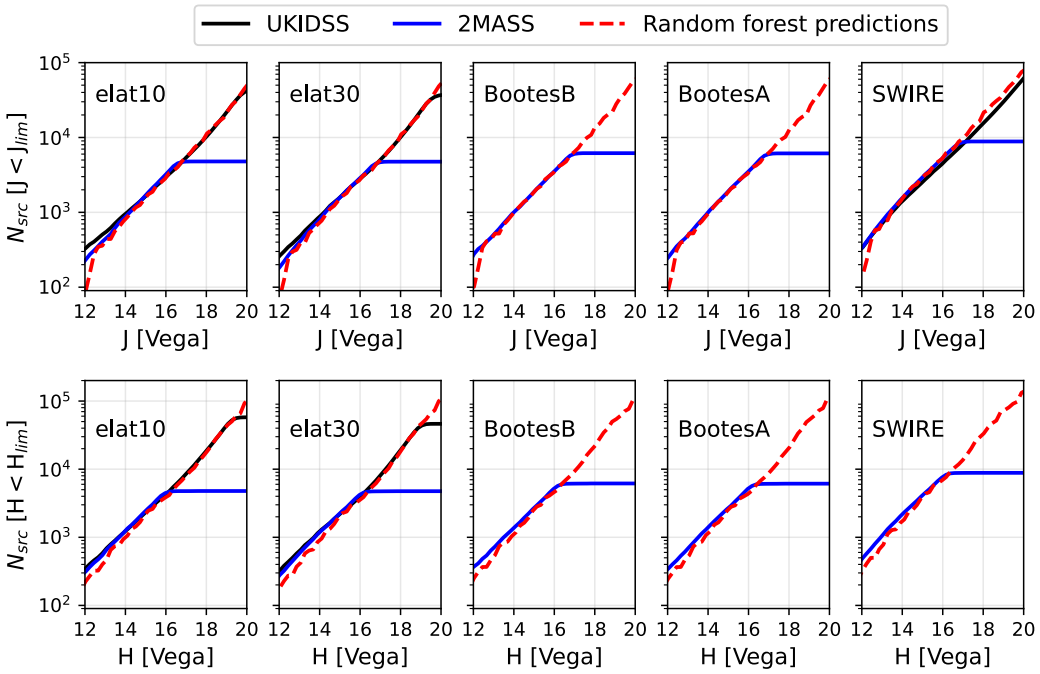

Figure 2.6: Cumulative number counts from 2MASS (blue), UKIDSS (black) and random forest-predicted magnitudes using PanSTARRS+unWISE photometry (red). Our predicted catalogs extend several magnitudes beyond 2MASS and have consistent number density to UKIDSS Large Area Survey (LAS) counts available in two of the five *CIBER* fields (elat10, elat30). The SWIRE field is covered by the UKIDSS Deep Extragalactic Survey (DXS) for J band, for which our random forest predicts slightly higher counts.

Source mask radius prescription

To construct the *CIBER* astronomical masks we model the masking radius for each source as a function of magnitude. For bright sources ($J < 14$ or $H < 14$), we parameterize the masking function as a function of magnitude m

$$r(m)[''] = A \exp\left[-\frac{(m_J - b)^2}{c^2}\right], \tag{2.29}$$

where $b = 3.6$, $c = 8.5$ and $A = 160$. For fainter sources the masking radius is determined non-parametrically. For each magnitude bin with $\Delta m = 0.5$ in the range $14 < J < 19$, we generate a model image of sources in that bin using the measured *CIBER* PSF. The masking radius for those sources is increased until the masked image has power $C_\ell^{ePSF} < 10^{-9}$ nW^2 m^{-4} sr^{-1} for all bandpowers. While the bright end masking parameters $\{A, b, c\}$ are the same across fields, our non-parametric approach on the faint end masking radii is performed for each field. The fact that the

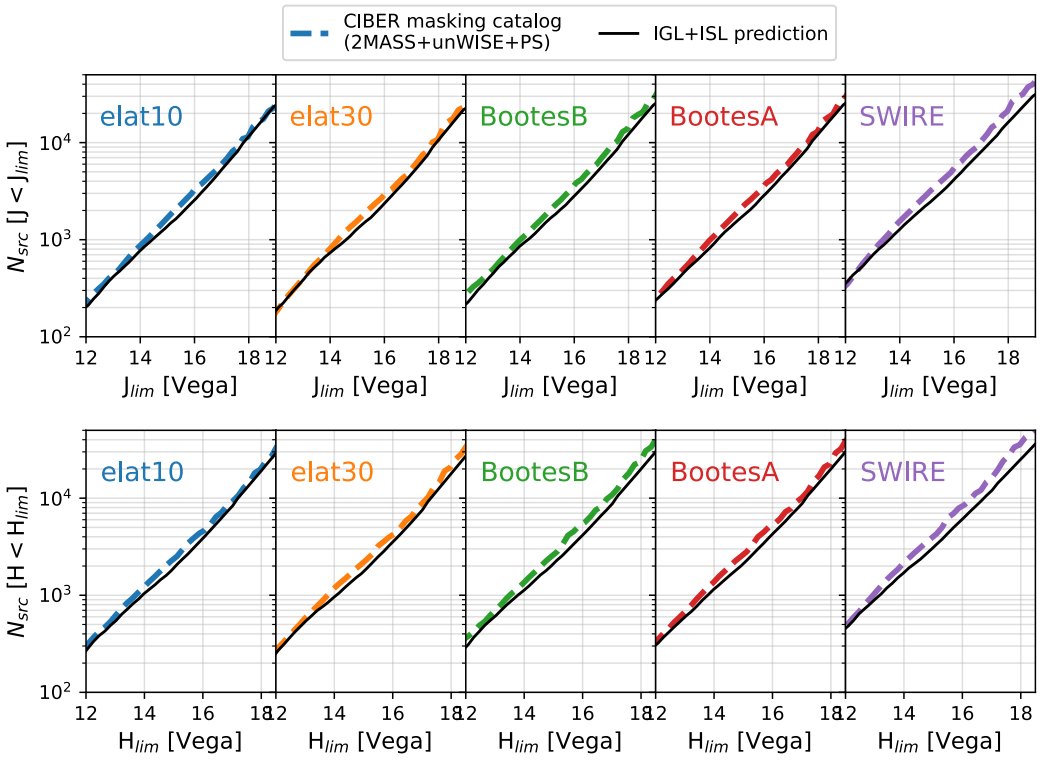

Figure 2.7: Cumulative magnitude distributions of the final *CIBER* masking catalogs (dashed lines) for J-band (top) and H-band (bottom), compared with those derived from IGL+ISL mock simulations of the same fields. The black solid line indicates the mean of the five mock fields while the shaded region bounds the minimum/maximum range in counts spanned by the fields.

CIBER PSF is undersampled means most of the flux for faint sources can be masked across a few pixels. This masking is more aggressive for bright sources than Z14.

We then combine the resulting source masks with the nominal *CIBER* instrument masks and use these to estimate the total contribution from extended PSF of masked sources. Specifically, for each realization we generate a map of sources down to our fiducial masking depth ($J = 17.5$ and $H = 17.0$ for 1.1 μm and 1.8 μm, respectively), apply its corresponding mask, and calculate the resulting power spectrum. We then correct for the effects of mode coupling and the beam transfer function. After performing this on 100 sets of mocks, we show the results of this exercise in Fig. 2.8. In the limit with no astrometric errors, the residual power from masked source halos is more than two orders of magnitude below the IGL+ISL signal and can be considered negligible. To simulate astrometry errors in the *CIBER* pointing solution, we perturb the positions of the injected sources in each dimension by

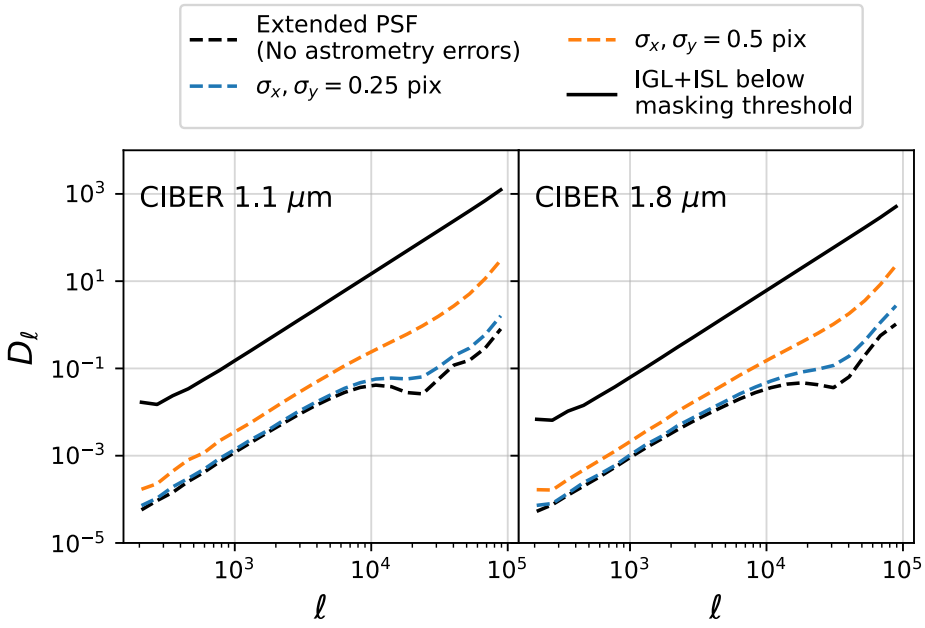

Figure 2.8: Comparison of power from extended PSF of masked sources in each field (dashed curves) and sub-threshold clustering from ISL and IGL (black points). These are obtained from 100 realizations per field and their respective source masks. The case of no astrometry errors is shown in black, while results for non-zero astrometry errors are shown in blue and orange.

$\sigma_x = 0.25$ and 0.5 *CIBER* pixels (3.5″ and 7″ respectively). In the presence of these astrometry errors the residual power from masked sources increases most significantly at $\ell > 10000$, however the residual power is still much lower than the IGL+ISL signal.

2.7 Mock power spectrum recovery tests

Using the mock *CIBER* observations described in §2.3 and the PS formalism from §2.5 we test our ability to recover sky fluctuations. Running the pipeline as implemented on several independent mock observations helps to identify any biases in the PS estimation and to quantify measurement uncertainties. For each test configuration we run our pipeline on one thousand sets of *CIBER* mocks. We note that in these mock tests we assume perfect knowledge for source masking, i.e., we do not directly emulate masking errors.

Field-averaged power spectrum

To optimally combine power spectrum estimates from the five *CIBER* fields we apply per-bandpower inverse variance weights derived from the dispersion of recovered mock power spectra. We show these weights as a function of multipole in Appendix .1. The degradation in statistical precision due to weighting is largest at intermediate scales $10^3 < \ell < 10^4$ where the read noise varies significantly across fields.

Effect of FF errors

We validate our pipeline with two test cases at the fiducial masking depths from Z14, namely $J < 17.5$ for 1.1 μm and $H < 17.0$ for 1.8 μm. The first case assumes perfect knowledge of the FF (i.e., no FF correction is needed), while the second incorporates the FF estimation and bias corrections. In Figures 2.9 and 2.10 we show the results of these tests. In both cases, the recovered power spectra for elat30 (orange) are much noisier than the other fields, due to the field's short exposure time (17.8 seconds compared to ~ 50 for the other science fields). On intermediate scales, the SWIRE field (purple) has large uncertainties due to its higher stellar density and thus masking fraction, despite having the lowest photon noise.

Averaged over the ensemble of mocks, our per-field and averaged power spectra are unbiased on large and small scales, with some exceptions. On large scales in all cases, the fifth bandpower is negatively biased at the $1 - 2\sigma$ level in both the $\delta[\hat{FF}] = 0$ and $\delta[\hat{FF}] \neq 0$ cases. We attribute this to effects of strong mode coupling between the low-ℓ bandpowers introduced by our image filtering (see bottom right panel of Fig. 2.4). In the $\delta[\hat{FF}] = 0$ case we also find a $\sim 1\sigma$ positive bias in the third bandpower for both bands. The fact that this bias is not seen in the $\delta[\hat{FF}] \neq 0$ case may suggest that the additional mode coupling from FF errors more smoothly distributes power across bandpowers, however more detailed tests are needed to confirm this hypothesis. On scales $\ell > 50000$ in the $\delta[\hat{FF}] \neq 0$ case, we find a slight positive bias. We do not see such a bias in the $\delta[\hat{FF}] = 0$ case, which suggests some instability in the FF noise bias correction. This is corroborated by the fact that the power spectrum bias is largest for fields elat10 and elat30, which are most sensitive to errors in the noise bias correction due to their high mean sky intensities that couple to instrumental errors in the FF correction. However, their larger dispersion across mock realizations also means they are more strongly down-weighted in each field average.

To understand the sensitivity of our power spectrum measurements as a function

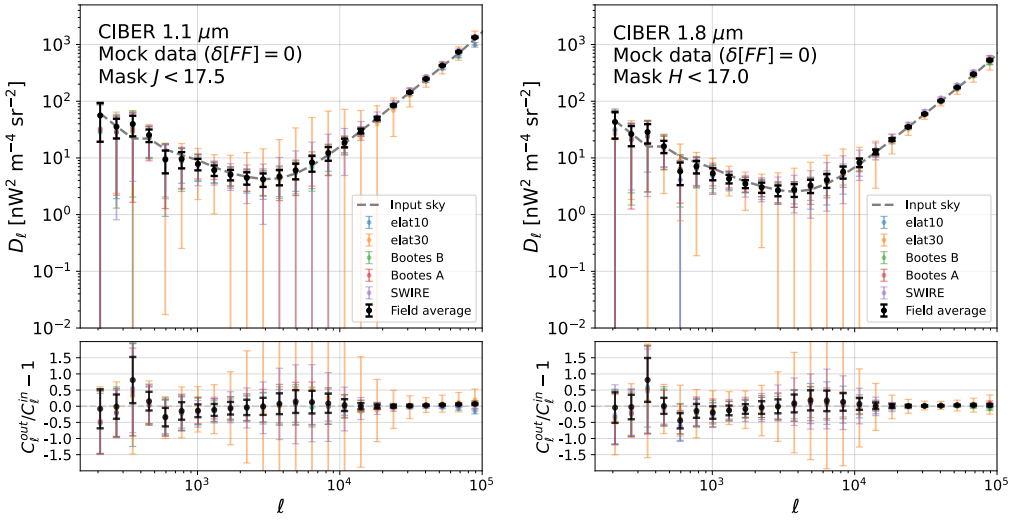

Figure 2.9: Mock power spectrum recovery with no FF errors ($\delta[\hat{FF}] = 0$), for individual fields (colored points) and field average (black points). This is plotted for 1.1 μm (left) and 1.8 μm (right) The errorbars on the black points are computed from the mean and dispersion of recovered power spectra from one thousand sets of maps, where each set denotes a realization of five *CIBER* fields. The bottom row shows the fractional power spectrum error relative to the input power spectra.

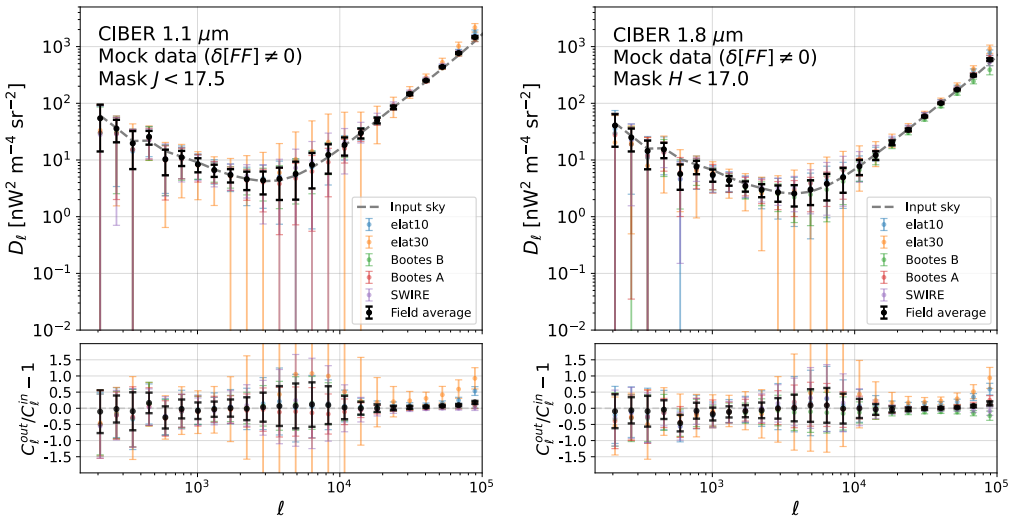

Figure 2.10: Mock power spectrum recovery with estimated FFs ($\delta[\hat{FF}] \neq 0$) using the stacking estimator from §2.5. In these tests we use laboratory FF templates (see Fig. 2.1) when generating mock *CIBER* observations.

of angular scale, in Fig. 2.11 we plot the fractional power spectrum uncertainties for both bands and test cases. On scales $\ell < 1000$, the uncertainties are driven by sample variance and the filtering transfer function. Our uncertainties peak again on intermediate scales near $\ell \sim 6000$, the result of concentrated power from read noise that spreads to other modes through the mask(s). The power spectra are sensitive to noise that is amplified by the beam transfer function correction, explaining the decrease in sensitivity at $\ell > 50000$. These results suggest the large-scale sensitivity peaks between $1000 < \ell < 2000$, corresponding to angular scales $5' < \theta < 10'$.

These mock recovery tests enable us to isolate the impact of FF errors on our final measurements. We highlight this in Figure 2.12, plotting the ratio of power spectrum uncertainties between test cases. While the uncertainties are consistently larger in the presence of FF errors as one would expect, we do find an exception in the third lowest bandpower ($\ell \sim 350$) for both bands. This is the same bandpower for which we find a 1σ bias and likewise may be explained by the mode couplings induced by FF errors. The degradation in sensitivity is generally larger for 1.1 μm than 1.8 μm due to higher photon and read noise levels. Fortuitously, the degradation in sensitivity due to FF errors is modest on large scales, remaining at the $< 20\%$ level for $\ell < 1000$ and $20 - 30\%$ for $1000 < \ell < 2000$. While in this work we do not study FF errors as a function of scale, these results suggest that large-scale modes of our FF estimates have lower error than scales dominated by read and photon noise.

Bandpower correlations

For each test configuration we calculate the bandpower covariance matrix describing departures of each field's recovered power spectrum (indexed by i) with respect to the field average from set j:

$$\hat{C}_{mock} = \langle (C^j_{\ell,i} - C^j_{\ell,av})^2 \rangle. \tag{2.30}$$

Note that this is different than the covariance computed relative to the true underlying sky power spectra. To highlight the difference in correlation structure with and without FF errors, we show the correlation matrices $\rho(\{\hat{C}_\ell\})$ for both bands in Fig. 2.13. In each plot, the upper triangular component is the correlation matrix for $\delta[\hat{FF}] = 0$, while the lower triangular component shows the full $\delta[\hat{FF}] \neq 0$ case. The within-field bandpower covariance (block-diagonal matrices) shows similar structure in both cases, with strong correlations from mode coupling and read noise on intermediate and small scales. These correlations are stronger for 1.1 μm than 1.8 μm which we attribute to the different noise levels across bands. Unlike the

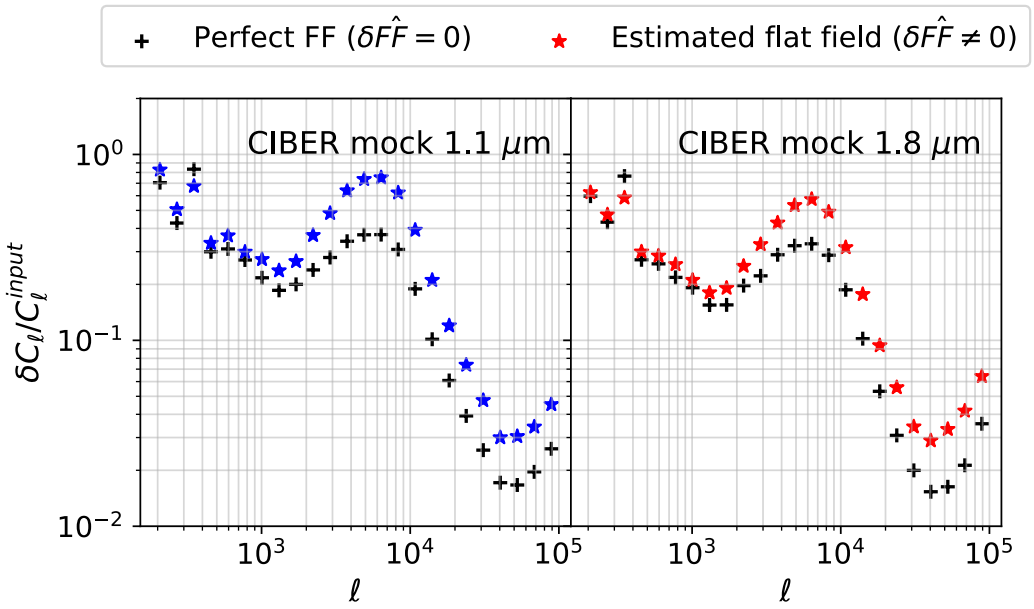

Figure 2.11: Fractional power spectrum uncertainties at 1.1 μm (left) and 1.8 μm (right) derived from the dispersion of recovered mock power spectra. We indicate results with and without FF errors using stars and crosses, respectively.

$\delta[\hat{FF}] = 0$ case, for which each field is treated separately, the $\delta[\hat{FF}] \neq 0$ case shows significant correlations between fields. The cross-field covariance arises because of our FF correction, which mixes the information from all the fields into each field's power spectrum estimate. This demonstrates the importance of accounting for the full field-field covariance when assessing internal consistency of the observed power spectrum measurements.

Power spectrum recovery for varying masking depths

Having demonstrated power spectrum estimation for our fiducial masking case, we now test our full pipeline on the same mocks but over a much broader range of source masking thresholds. This includes recovery of both point source-dominated power spectra (i.e., shallow masking cuts) and much deeper cuts, in total spanning three orders of magnitude in Poisson noise. For sources with $J < 11$, non-linear detector response and saturation in the observed data preclude reliable measurements without detailed corrections. Our deepest masking cuts ($J < 18.5$ and $H < 18.0$ for 1.1 μm and 1.8 μm, respectively) are informed by the reliability of our source masking algorithm as demonstrated in §2.6. Although it is not our science focus to measure point source-dominated fluctuations, this exercise enables us to test the consistency

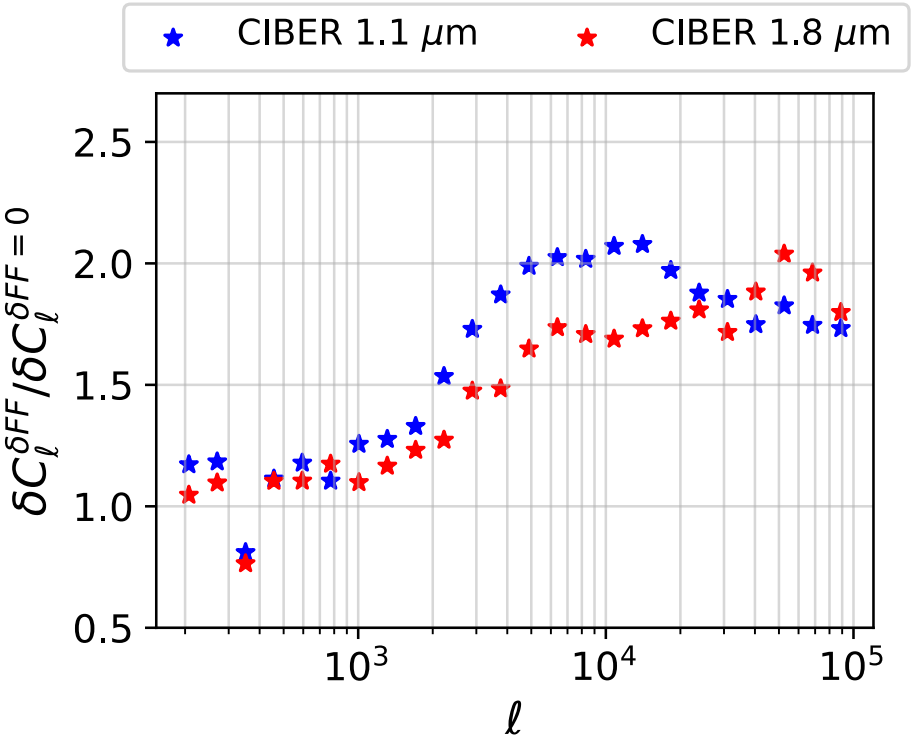

Figure 2.12: Ratio of power spectrum uncertainties in recovery with and without FF errors. While there is a clear degradation in sensitivity on scales $\ell > 3000$, the increased uncertainty on large scales is relatively modest.

of large-angle fluctuations in the observed data as a function of masking depth.

We note that our formalism for handling FF errors breaks down in the presence of bright unmasked point sources. This is a result of using the FF stacking estimator, in which bright point sources need to be masked regardless of masking depth to avoid large FF errors. As a result, the FF errors driven by sky signal differ from that of the target signal (which does contain bright point sources), meaning that the linear FF bias correction is not appropriate. In place of a full treatment, which would require an iterative or simultaneous estimation of power spectra at several masking depths, we characterize this effect empirically using the mocks. We determine that the $M_{\ell\ell'}$ correction without FF errors recovers more accurate power spectra down to $(J_{lim}, H_{lim}) = 15$, while $M_{\ell\ell'}^{mask+FF+filt}$ is more accurate for deeper source masking cuts. We use the fiducial masks ($J < 17.5$ and $H < 17.0$) to calculate $\hat{FF}$ for all shallower masking cuts.

We show the recovered power spectra as a function of masking depth in Figure

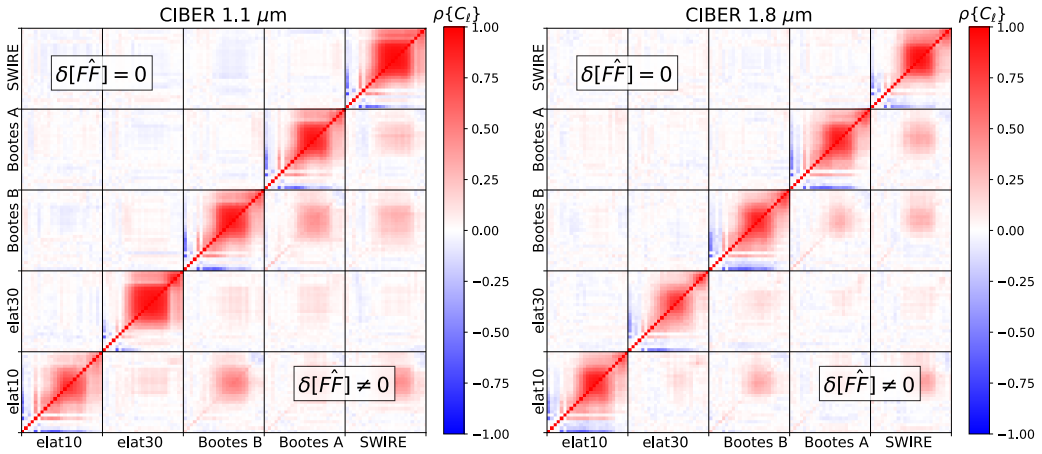

Figure 2.13: Mock correlation matrices for *CIBER* 1.1 μm (left) and 1.8 μm (right). These correlation matrices are derived from 1000 sets of mock *CIBER* observations including instrumental noise, foregrounds, masking and FF estimation (for $\delta[\hat{FF}] \neq 0$). These observational effects induce strong mode coupling on intermediate and small scales within individual fields and between pairs of fields for $\delta[\hat{FF}] \neq 0$. The mild correlations between fields in the $\delta[\hat{FF}] = 0$ case reflect deviations from each weighted field average, which contains information from all fields; these correlations approach zero in the limit of more simulations.

2.14. These results validate our ability to measure large-angle fluctuations across all masking cases. As expected, the fractional power spectrum errors on scales $\ell > 1000$ are largest near the masking depth where we transition from $M_{\ell\ell'}^{mask+filt}$ to $M_{\ell\ell'}^{mask+filt+FF}$. The slight underestimation for $(J_{lim}, H_{lim}) = 16$ is due to the fact that the masks used to estimate the FF are deeper than those used to compute the power spectrum, such that $M_{\ell\ell'}^{mask+filt+FF}$ slightly overcorrects the target signal.

The overcorrection on small scales is due to the fact that the masks used to estimate the FF are deeper than those used to compute the power spectrum, meaning the linear FF bias is not fully coupled to the target signal. We do not pursue shot noise corrections using the estimated power spectra at the FF masking depth, however this may be possible and useful in cases with stronger requirements on estimation of Poisson fluctuations.

2.8 Conclusion

In this work we present an extension of the pseudo-C_ℓ formalism for measurements of NIR EBL anisotropies, with application to imaging data from the Cosmic Infrared Background Experiment. Improving on the methodology in Z14, we address two important effects necessary for measurement of sky fluctuations, namely source

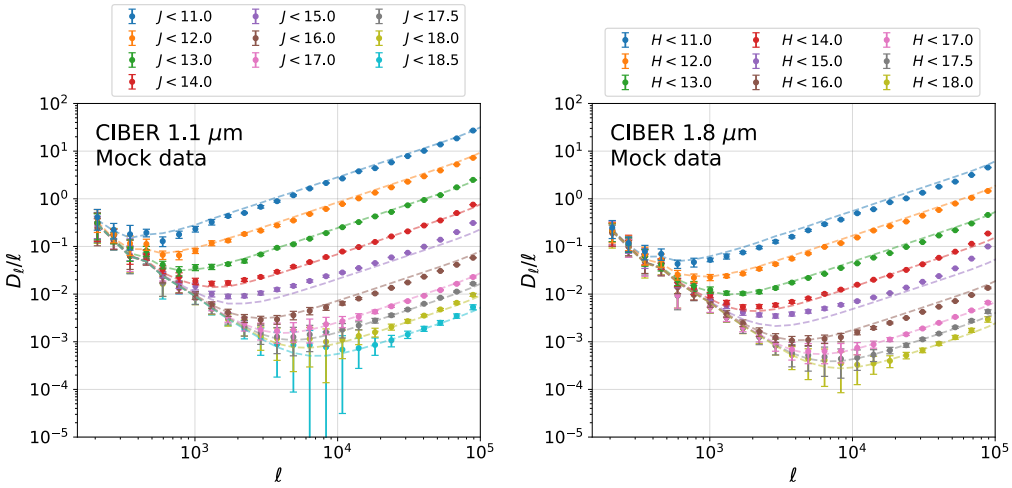

Figure 2.14: Mock power spectrum recovery for a range of masking cuts using the simulations described in §2.3. These results validate our ability to recover unbiased power spectra in the presence of shot noise spanning three orders of magnitude.

masking and FF correction. We derive sky flats directly from the science fields and build on the pseudo-C_ℓ formalism to correct for additive and multiplicative biases sourced by FF errors. Through tests on mock *CIBER* observations with injected FFs obtained from laboratory measurements, we demonstrate that our power spectrum pipeline can recover unbiased power spectra for all but the smallest angular scales. By comparing against similar mock tests where the FF is assumed to be known perfectly, we determine that residual FF errors increase the recovered statistical uncertainties on scales $500 < \ell < 2000$ by less than 20%.

Our scheme bypasses the use of field differences used in Z14, which opens the potential for more aggressive point source masking in individual fields. Our source masking approach is an efficient, data-driven alternative to direct SED fitting, and enables us to mask two magnitudes deeper in the NIR than is possible with existing 2MASS photometry in the *CIBER* fields.

Correcting for mode coupling effects is an important component of this work. As NIR EBL fluctuation measurements become signal dominated there are a number of additional pseudo-C_ℓ corrections that will be important to consider. The first involves the fact that many filtering operations can couple with the astronomical masks, and should be folded into $M_{\ell\ell'}$ estimation. In this work we incorporate image filtering into our mode mixing corrections. The second effect, which we do not correct for, involves the dependence of the mode coupling correction on the shape

of the underlying sky power spectrum. By choosing sufficiently fine bandpowers one can mitigate biases from this effect, however in general the mode coupling will be affected by derivatives of the sky power spectrum $\{\delta^i C_\ell / \delta \ell^i\}$. The shape of the power spectrum can be incorporated as a prior in mode coupling corrections through the bandpower operator P_{bl}, as done in [88]. However, this opens up the much broader question of how one should properly perform inference of pseudo-C_ℓ measurements. It has been shown that the standard MASTER result for computing pseudo-C_ℓ estimates is biased when there is correlation between the signal and the mask [74, 95]. [96] shows this correlation can be calculated analytically and through simulations, with corrections going as the three- and four-point functions of the maps and masks. Estimating such corrections requires the use of more realistic sky models, such as those from MICECAT, since it will be sensitive to the one-halo contributions of IGL and any additional intra-halo light.

In some cases, mode coupling effects can be mitigated at the map level. For example, the fraction of masked pixels can be reduced by subtracting bright stars (or, e.g., extended PSF components) at the map level. With redshift information, which will be available for many *SPHEREx* galaxies [97], it may be possible to perform physically motivated masking, in which the extended light component of bright, low-redshift galaxies can be removed at fixed comoving radius. Targeted point source/CIB de-projection [98, 99] or more general component separation of pointlike and diffuse signals [100] are avenues toward mitigating the effects of mode coupling on future clustering measurements.

The significance of these effects (and mitigation techniques) should be tested directly through tests on realistic mocks in order to assess their impact at fixed experiment sensitivity. By the same token, it may be more reliable in future NIR EBL inferences to instead forward model pseudo-C_ℓ measurements, returning C_ℓ *reconstructions* rather than inverted C_ℓ point estimates. Such a Bayesian approach requires careful consideration of instrumental and observational effects but offers a promising path to obtaining faithful power spectrum uncertainty estimates.

Near-future experiments will map out the NIR EBL over larger regions of sky with significantly broader spectral coverage and resolving power. *CIBER*-2, the second generation of *CIBER*, has three H2RG detectors and six windowpane filters for imaging at $0.5 - 2.5$ μm [101, 102]. The power spectrum formalism from this work will be important for *CIBER*-2 data, which is similar in structure to *CIBER*-1 data. The Spectro-Photometer for the History of the Universe, Epoch of Reionization, and

Ices Explorer (*SPHEREx*) is the next NASA Medium Class Explorer mission which is planned for launch in early 2025. *SPHEREx* will conduct a two-year, all-sky survey in 102 bands spanning 0.75-5 μm, dramatically increasing the volume of data available for intensity mapping. The primary focus for intensity mapping will be in 200 deg^2 centered near the north and south ecliptic poles. *SPHEREx* will be orders of magnitude more sensitive due to higher instrument sensitivity and a daily cadence over the poles throughout the nominal two-year survey.

Improved methods to estimate CNIB fluctuations, a larger analysis toolkit to interpret measurements and a dramatic increase in data quantity and quality will transform our ability to study the history of cosmic light production in the coming years, uncovering features of the low surface brightness universe that may yield unanticipated insights about galaxy evolution and large scale structure formation.

.1 Power spectrum estimation with FF stacking estimator

We introduce the stacking estimator for the FF and propagate FF errors to power spectrum bias. Under the assumption that the sky signals of the fields are drawn from the same underlying distribution, the effect of the FF error after noise bias subtraction is to leave a multiplicative bias on the underlying sky signal.

Stacking FF estimator

Let us consider a single sky realization $I_i(x, y)$ for field i, the sum of a mean normalization $\overline{I}_i^{sky}$ (dominated by ZL), and a fluctuation component S_i which is composed from EBL sky fluctuations, diffuse galactic light (DGL) fluctuations and integrated stellar light (ISL):

$$I_i^{sky}(x, y) = \overline{I}_i^{sky} + S_i = \overline{I}_i^{sky} + S_i^{EBL} + S_i^{DGL} + S_i^{ISL}. \tag{31}$$

The FF responsivity is represented as a scalar field $FF(x, y)$ defined across each detector. The incident sky signal has associated photon noise ϵ_γ, and this signal+noise component is multiplied by $FF(x, y)$, after which read noise (denoted by ϵ_{read}) is imprinted, producing the observed image $I_i^{obs}(x, y)$:

$$I_i^{obs}(x, y) = FF(x, y)\left[\overline{I}_i^{sky} + S_i^{EBL} + S_i^{DGL} + S_i^{ISL} + \epsilon_{\gamma,i}\right] + \epsilon_{read,i}. \tag{32}$$

The FF estimate derived from field i is obtained by dividing the observed image by

the mean surface brightness in unmasked pixels:

$$\hat{FF}_i(x,y) = \frac{I_i^{obs}(x,y)}{\overline{I}_i^{sky}} = \frac{FF(x,y)\left[\overline{I}_i^{sky} + S_i^{EBL} + S_i^{DGL} + S_i^{ISL} + \epsilon_{\gamma,i}\right] + \epsilon_{read,i}}{\overline{I}_i^{sky}}. \tag{33}$$

With this stacking estimator we can model how FF errors propagate to the estimated power spectrum, using relatively minimal assumptions about the observations. This is opposed to using lab-derived FFs, which assume a perfectly uniform illuminating surface but have errors that are harder to quantify. The error on the FF estimate for a single field can be expressed in terms of the assumed sky and noise components:

$$\delta\hat{FF}_i(x,y) = \frac{FF(x,y)(S_i^{EBL} + S_i^{DGL} + S_i^{ISL} + \epsilon_{\gamma,i}) + \epsilon_{read,i}}{\overline{I}_i^{sky}}. \tag{34}$$

This stacking estimator improves as the background levels of the images increase – the FF responsivity will primarily couple with the Zodiacal light, which is spatially smooth over the field of view. Any additional, non-uniform fluctuation components lead to errors in $\hat{FF}$. The FF responsivity only needs to be evaluated in unmasked pixels, but each pixel has a specific number of off-field measurements, and this depends on whether the same pixels are masked in the off fields,

$$\delta\hat{FF}_i(x,y) = M_i(x,y)\left(\frac{FF(x,y)(S_i^{EBL} + S_i^{DGL} + S_i^{ISL} + \epsilon_{\gamma,i}) + \epsilon_{read,i}}{\overline{I}_i^{sky}}\right). \tag{35}$$

We apply pixel weights unique to each "off-field" used in a given stacked FF estimate. Denote the per-pixel RMS of the FF error by σ. The resulting inverse variance weights are

$$w_i \approx \left(\frac{\overline{I}_i^{sky}}{\sigma_{tot}}\right)^2, \tag{36}$$

where

$$\sigma_{tot} = \sqrt{\sigma_\gamma^2 + \sigma_{EBL}^2 + \sigma_{DGL}^2 + \sigma_{read}^2}. \tag{37}$$

Assuming the per-pixel fluctuations from the sky signal S_i are subdominant to the instrumental noise, we can approximate the weights as

$$w_i \approx \left(\frac{\overline{I}_i^{sky}}{\sqrt{\sigma_{\gamma,i}^2 + \sigma_{read,i}^2}}\right)^2. \tag{38}$$

Assuming our field weights are sum-normalized, i.e., $\sum_i w_i = 1$, the stacked FF $\hat{FF}^j(x, y)$ is

$$\hat{FF}^j(x, y) = \sum_{i=1}^{N_f} w_i \hat{FF}_i. \tag{39}$$

The $N_f = 4$ fields that go into each stacked CIBER FF estimate are uncorrelated, such that the variance of the stacked FF can be written as a weighted sum of variances

$$\mathrm{Var}[\hat{FF}^j] = \sum_{i=1}^{N_f} w_i^2 \mathrm{Var}[\hat{FF}_i]. \tag{40}$$

Condensing the sky fluctuation signal for field i into S_i, we can write the FF standard error (given inverse variance weights) as

$$\delta[\hat{FF}]_j = \left(\sum_{i \neq j} \delta[\hat{FF}]_i^{-2} \right)^{-1/2} = \left(\sum_{i \neq j} \left[\frac{M_i \left(FF(S_i + \epsilon_{\gamma,i}) + \epsilon_{read,i} \right)}{\overline{I}_i^{sky}} \right]^{-2} \right)^{-1/2}. \tag{41}$$

Flat field bias

Here we quantify how errors in the FF propagate through the pipeline to estimates of the clustering power spectrum. In this subsection we show that errors on the FF manifest as a *multiplicative bias* on the underlying power spectrum, assuming the underlying sky signals are drawn from the same power spectrum.

Unmasked case

In this case we assume that foreground point sources are perfectly removed from the maps. Following Eq. (31) the observed signal for field j is

$$I_j^{obs} = FF_{true} \left[I_j^{sky} + \epsilon_{\gamma,j} \right] + \epsilon_{read,j}. \tag{42}$$

Let us express our FF estimate in terms of the true FF and the FF error, i.e.,

$$\hat{FF}^j = FF_{true} + \delta[\hat{FF}^j] = FF_{true} \left[1 + \frac{\delta[\hat{FF}^j]}{FF_{true}} \right]. \tag{43}$$

The FF-corrected image is expressed as the following:

$$\frac{I_j^{obs}}{\hat{FF}^j} = \frac{FF_{true} \left[I_j^{sky} + \epsilon_{\gamma,j} \right] + \epsilon_{read,j}}{FF_{true} \left[1 + \frac{\delta[\hat{FF}^j]}{FF_{true}} \right]}. \tag{44}$$

Assuming $\delta[\hat{FF}^j]/FF_{true} \ll 1$, we Taylor expand the expression above:

$$\frac{I_j^{obs}}{\hat{FF}^j} \approx \left(I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{FF_{true}}\right)\left(1 - \frac{\delta[\hat{FF}^j]}{FF_{true}}\right) \tag{45}$$

$$= I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{FF_{true}} - \frac{\delta[\hat{FF}^j]}{FF_{true}}\left(I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{FF_{true}}\right). \tag{46}$$

We then express FF_{true} in terms of the estimated FF and its error. For example, the read noise term is rewritten as

$$\frac{\epsilon_{read,j}}{FF_{true}} = \frac{\epsilon_{read,j}}{\hat{FF}^j - \delta[\hat{FF}^j]} = \frac{\epsilon_{read,j}}{\hat{FF}^j(1 - \frac{\delta[\hat{FF}^j]}{\hat{FF}^j})} \approx \frac{\epsilon_{read,j}}{\hat{FF}^j}\left(1 + \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}\right) \tag{47}$$

such that

$$\begin{aligned}
\frac{I_j^{obs}}{\hat{FF}^j} &\approx I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j}\left(1 + \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}\right) \\
&\qquad - \frac{\delta[\hat{FF}^j]}{\hat{FF}}\left(1 + \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}\right)\left(I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{FF_{true}}\right) \\
&= I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j} + \frac{\epsilon_{read,j}}{\hat{FF}^j}\frac{\delta[\hat{FF}^j]}{\hat{FF}} \\
&\qquad - \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}\left(I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{FF_{true}}\right) + O(\delta^2[\hat{FF}^j]) \\
&\approx I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j} + \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}\left(I_j^{sky} + \epsilon_{\gamma,j}\right) \\
&\qquad + \left[\frac{\epsilon_{read,j}}{\hat{FF}^j}\frac{\delta[\hat{FF}^j]}{\hat{FF}^j} - \frac{\epsilon_{read,j}}{FF_{true}}\frac{\delta[\hat{FF}^j]}{\hat{FF}^j}\right] \\
&\approx I_j^{sky} + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j} + \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}\left(I_j^{sky} + \epsilon_{\gamma,j}\right) + O(\delta^2[\hat{FF}^j]).
\end{aligned}$$

After discarding all terms of order $\delta^2[\hat{FF}^j]$, we express I_j^{sky} as the sum of mean background and fluctuation components,

$$\frac{I_j^{obs}}{\hat{FF}^j} \approx \bar{I}_j^{sky} + S_j + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j} - \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}(\bar{I}_j^{sky} + S_j + \epsilon_{\gamma,j}). \tag{48}$$

This expression for additional fluctuations are in the FF corrected image quantifies the power modulated by each noisy FF when using the FF stacking estimator. In addition to a slight change in the read noise bias, a large portion of the bias corresponds to the fractional FF error $\delta[\hat{FF}^j]/\hat{FF}^j$ coupled with the mean sky brightness.

Noise bias subtraction

Additional fluctuations from read noise and photon noise producing FF errors are to good approximation uncorrelated with the signal, such that the noise power adds linearly. Define the FF estimate in the absence of sky fluctuations as $\hat{FF}_{inst}$:

$$\hat{FF}^j_{inst} = FF\left[1 + \frac{\epsilon_{\gamma,j}}{\bar{I}^{sky}_j}\right] + \frac{\epsilon_{read,j}}{\bar{I}^{sky}_j}. \tag{49}$$

After correcting the mean sky brightness by $\hat{FF}_{inst}$,

$$\frac{\bar{I}^{sky}_j}{\hat{FF}^j_{inst}} = \bar{I}^{sky}_j + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j_{inst}} - \frac{\delta[\hat{FF}^j_{inst}]}{\hat{FF}^j_{inst}}\left(\bar{I}^{sky}_j + \epsilon_{\gamma,j}\right). \tag{50}$$

We use Monte Carlo realizations of (50) to estimate and subtract noise biases from each observed pseudo-C_ℓ estimate, removing leading contributions from $\epsilon_{\gamma,j}$ and $\epsilon_{read,j}$.

We can express the FF error $\delta[\hat{FF}^j]$ in terms of instrument noise and sky fluctuation components,

$$\delta[\hat{FF}^j] = \delta[\hat{FF}^j_{inst}] + \delta[\hat{FF}^j_S]. \tag{51}$$

Using this we can expand (48) and subtract the noise bias from (50). After removing terms $O(\delta\hat{FF}^j_{inst}\delta\hat{FF}^j_S)$, we are left with

$$\frac{I^{obs}_j}{\hat{FF}^j} - \frac{\bar{I}^{sky}_j}{\hat{FF}^j_{inst}} \approx S_j - \frac{\delta[\hat{FF}^j_S]}{\hat{FF}^j}\bar{I}^{sky}_j - \left[\frac{\delta[\hat{FF}^j_S]}{\hat{FF}^j}\left(S_j + \epsilon_{\gamma,j} - \frac{\epsilon_{read,j}}{\hat{FF}^j}\right) + \frac{\delta[\hat{FF}^j_{inst}]}{\hat{FF}^j}S_j\right]. \tag{52}$$

The additional noise terms in brackets are sub-dominant to the second term in (52) and depend on the underlying fluctuations S_j. To assess the importance of the cross terms we use our mocks and noise models to evaluate each term numerically. We confirm that they are small relative to the underlying sky power ($\lesssim 1\%$ of C^{true}_ℓ) and so we do not attempt to remove them in full when applying noise de-biasing.

Predicting FF bias for unmasked case

In the unmasked case it is possible to predict the FF bias analytically. For this derivation we consider different mean surface brightnesses and noise levels across fields. Assume common underlying sky fluctuations denoted by δS, along with instrument noise and mean surface brightnesses unique to each field i (ϵ_i and I_i,

respectively) for N_f fields (excluding the field being corrected). We can express the per-pixel FF error obtained from each field as

$$\delta FF_i = \frac{\sqrt{\epsilon_i^2 + \delta S^2}}{I_i}. \tag{53}$$

The weighted variance from stacking several fields is then

$$\delta[\hat{FF}^j]^2 = \sum_i w_i^2 \delta FF_i^2 = \sum_i w_i^2 \frac{(\epsilon_i^2 + \delta S^2)}{I_i^2}. \tag{54}$$

Following Eq. 46 we can express the FF corrected image for field j as

$$I_j^{obs}/\hat{FF}^j \approx \left(I_j + \sqrt{\epsilon_j^2 + \delta S^2}\right)\left(1 - \sqrt{\sum_i \frac{w_i^2(\epsilon_i^2 + \delta S^2)}{I_i^2}}\right) \tag{55}$$

$$= I_j - I_j\sqrt{\sum_i \frac{w_i^2(\epsilon_i^2 + \delta S^2)}{I_i^2}} + \sqrt{\epsilon_j^2 + \delta S^2} + O(I_i^{-1}). \tag{56}$$

In power units, this is

$$C_{\ell,j} \approx N_{\ell,j} + \delta S^2 + \sum_i \left(\frac{w_i I_j}{I_i}\right)^2 (\epsilon_i^2 + \delta S^2). \tag{57}$$

To calculate the noise bias we expand $\delta[\hat{FF}^j_{inst}]$ in Eq. 50

$$\delta[\hat{FF}^j_{inst}] = \sqrt{\sum_i \left(\frac{w_i \epsilon_i}{I_i}\right)^2}; \quad I_j^{inst} = I_j + \epsilon_j + \sqrt{\sum_i \left(\frac{w_i \epsilon_i I_j}{I_i}\right)^2}. \tag{58}$$

Assuming the instrument noise between fields is independent, we calculate the noise bias with FF errors $N_{\ell,j}^{\delta FF}$:

$$N_{\ell,j}^{\delta FF} = N_{\ell,j} + \sum_i \left(\frac{w_i \epsilon_i I_j}{I_i}\right)^2. \tag{59}$$

After subtracting the noise bias we are left with:

$$C_{\ell,j} - N_{\ell,j}^{\delta FF} = \delta S^2 \left[1 + \sum_i \left(\frac{w_i I_j}{I_i}\right)^2\right]. \tag{60}$$

For equal weights, noise, and ZL levels, the bias reduces to $1 + \frac{1}{N_f}$. This derivation tells us that to leading order, the FF bias depends on the relative mean surface brightness of each target field compared to those used in its FF estimate.

Mode coupling correction in presence of FF errors

The mask couples modes contaminated by FF errors. Denote the mask for field j by M_j, then

$$\begin{aligned}\frac{I_j^{obs}}{\hat{FF}^j} &\approx M_j \left[\bar{I}_j^{sky} + S_j + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j} - \frac{\delta[\hat{FF}^j]}{\hat{FF}^j}(\bar{I}_j^{sky} + S_j + \epsilon_{\gamma,j})\right] \\ &= M_j \left[\bar{I}_j^{sky} + S_j + \epsilon_{\gamma,j} + \frac{\epsilon_{read,j}}{\hat{FF}^j} \right. \\ &\qquad \left. - \frac{\bar{I}_j^{sky} + S_j + \epsilon_{\gamma,j}}{\hat{FF}^j}\left(\sum_{i\neq j}\left[\frac{M_i\left(FF(S_i+\epsilon_{\gamma,i})+\epsilon_{read,i}\right)}{\bar{I}_i^{sky}}\right]^{-2}\right)^{-1/2}\right].\end{aligned}$$

This means that the FF error $\delta[\hat{FF}^j]$ for field j depends on the mode coupling of the off-field masks $\{M_i\}$ and their sky signals $\{S_i\}$. Relying on the fact that the mode coupling operations are linearly separable to good approximation, we estimate the additional mode coupling through a modification of the standard MASTER algorithm.

The procedure for computing a single mask+FF Monte Carlo $M_{\ell\ell'}$ realization is as follows. For each bandpower:

1. Generate n_{field} Gaussian phase realizations with power spectrum equal to one within the bandpower and zero outside. In our case $n_{field} = 5$.
2. Add the mean surface brightness levels corresponding to the fields to their respective realizations.
3. Multiply each field by its respective mask.
4. For each of the n_{field} realizations:
    - Compute a stacked FF estimate from the other $(n_{field} - 1)$ maps, with field weights w_i.
    - Divide the phase realization by the estimated FF.
    - Compute the angular power spectrum from the corrected map.

This effectively decomposes the FF error in Fourier space to estimate each mode's diagonal and off-diagonal multiplicative bias. This is done for 500 Monte Carlo realizations, per bandpower. The implementation has been optimized using the

`pyfftw` package, which allocates memory for efficient computation of discrete Fourier transforms (DFTs) and inverse DFTs. For a set of $n_{field} = 5$ fields and twenty five bandpowers, this takes $\sim$ 20 minutes in wall clock time to run on a Macbook Pro with an M1 processor.

Field weights

The recovered power spectrum of a given field will depend on the masking fraction, the integration time, and any noise on the FF power spectrum bias from other fields. Masking has a large impact at intermediate scales, where noisy Fourier modes from the readout electronics mix with the masks. These weights are shown in Figure .15. There is statistical noise at the few percent level in the weights due to the number of simulations we use. Nonetheless there is clear scale-dependence in the power spectrum weights across different fields, and the weights follow a similar scale dependence for both *CIBER* bands. On large scales, our measurements are limited by statistical noise in the number of modes we sample, and so there is relatively little variation between fields. On intermediate scales where the power spectra are read noise and photon noise dominated, there is much larger variation between the field weights, with the elat30 and SWIRE fields the most downweighted. elat30 has the shortest integration time and therefore the largest instrument noise components, and so the exposure does not contain much information on these scales. SWIRE has an integration 2.5 times that of elat30, along with the lowest ZL level, but it has the highest source density (both stars and galaxies) and so the masking fraction is largest. The relative weights are similar between both bands, however at 1.8 μm the variation is smaller, which are explained by the lower photon noise levels in each exposure.

For the field weights derived from our ensemble of mocks, the fractional reduction in N_{eff} remains below 10% for $\ell < 2000$. On read noise dominated scales this increases to 30% and 20% for 1.1 μm and 1.8 μm, respectively, pleateauing on smaller scales to $\sim$ 10% each.

Cross correlation FF bias

The same effect which introduces a linear bias into our auto spectrum measurements also impacts the *CIBER* cross spectrum. This is due to our use of the FF stacking estimator in both *CIBER* bands. For a cross correlation between *CIBER* and other instruments (e.g., from IRIS, *Spitzer*) or ancillary galaxy catalogs, there is no FF

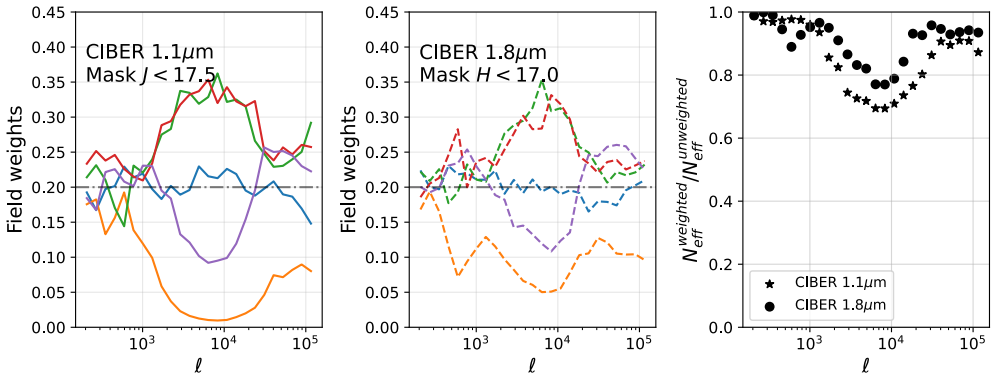

Figure .15: Inverse variance power spectrum weights for the five CIBER fields, for 1.1 μm (left) and 1.8 μm (middle). These weights are computed from the recovered power spectra of 1000 mock realizations. The right panel shows the effective mode reduction as a function of scale due to the field weights. On intermediate scales where read noise and mode coupling are prominent, the weight dispersion is largest.

bias. Any coherent signal across the *CIBER* imagers translate to coherent FF errors across bands, which then couple to the observed maps. Because the instrument noise across imagers is uncorrelated, the resulting FF errors do not introduce a noise bias like in the auto spectra.

We correct for the linear FF cross-spectrum bias and its mode coupling using a similar Monte Carlo estimation as for the auto spectra. To estimate mode coupling matrices for the set of fields, we use the mean surface brightness levels at both 1.1 and 1.8 μm to simulate FF errors in both imagers that depend on the injected tones at each bandpower. In the unmasked case the multiplicative bias depends on the weighted product of sky brightnesses across both bands, i.e., for bands a and b:

$$C_\ell^{j,\delta FF}/C_\ell^{j,true} = 1 + \sum_{i\neq j}\left(\frac{I_j^a I_j^b}{I_i^a I_i^b} w_i^a w_i^b\right)^2 . \tag{61}$$

.2 Auto- and cross-power spectrum statistical uncertainties

We derive expressions for the statistical uncertainty of angular auto- and cross-power spectra, known commonly as the Knox formula [103]. This represents the uncertainty in measurements due to sample variance, which sets a fundamental limit on the number of modes accessible over some portion of the sky.

We start with two real-valued fields I_1 and I_2, observed over some patch of sky with solid angle Ω. Assuming the flat sky limit (i.e., taking the Limber approximation),

we define the angular cross power spectrum as the real part of the product of Fourier transforms $\tilde{I}_1$ and $\tilde{I}_2$:

$$\hat{C_\ell}^{1,2} = \frac{\Omega}{2}\left(\tilde{I}_1\tilde{I}_2^* + \tilde{I}_1^*\tilde{I}_2\right). \tag{62}$$

Each Fourier-transformed field generically has a signal component and a noise component. We can write down the variance of the above cross spectrum as (dropping ℓ from notation for brevity):

$$\delta\hat{C}_{1,2}^2 = \langle\hat{C}_{1,2}^2\rangle - \langle\hat{C}_{1,2}\rangle^2. \tag{63}$$

Expanding this into the Fourier-transformed fields gives

$$\delta\hat{C}_{1,2}^2 = \left\langle\frac{\Omega^2}{4}\left(\tilde{I}_1\tilde{I}_2^* + \tilde{I}_1^*\tilde{I}_2\right)^2\right\rangle - \left\langle\frac{\Omega}{2}\left(\tilde{I}_1\tilde{I}_2^* + \tilde{I}_1^*\tilde{I}_2\right)\right\rangle^2 \tag{64}$$

$$= \frac{\Omega^2}{4}\left[\left\langle\tilde{I}_1\tilde{I}_1\tilde{I}_2^*\tilde{I}_2^* + 2\tilde{I}_1\tilde{I}_1^*\tilde{I}_2\tilde{I}_2^* + \tilde{I}_1^*\tilde{I}_1^*\tilde{I}_2\tilde{I}_2\right\rangle - \left\langle\tilde{I}_1\tilde{I}_2^* + \tilde{I}_1^*\tilde{I}_2\right\rangle^2\right] \tag{65}$$

$$= \frac{\Omega^2}{4}\left[2\left\langle\tilde{I}_1\tilde{I}_1\tilde{I}_2^*\tilde{I}_2^*\right\rangle + 2\left\langle\tilde{I}_1\tilde{I}_1^*\tilde{I}_2\tilde{I}_2^*\right\rangle - \left\langle\tilde{I}_1\tilde{I}_2^* + \tilde{I}_1^*\tilde{I}_2\right\rangle^2\right]. \tag{66}$$

The next step invokes Wick's theorem, to rewrite the four-term products as the sum of two-point cross products. The first is

$$\left\langle\tilde{I}_1\tilde{I}_1\tilde{I}_2^*\tilde{I}_2^*\right\rangle = \left\langle\tilde{I}_1\tilde{I}_1\right\rangle\left\langle\tilde{I}_2^*\tilde{I}_2^*\right\rangle + \left\langle\tilde{I}_1\tilde{I}_2^*\right\rangle\left\langle\tilde{I}_2^*\tilde{I}_1\right\rangle + \left\langle\tilde{I}_1\tilde{I}_2^*\right\rangle\left\langle\tilde{I}_1\tilde{I}_2^*\right\rangle \tag{67}$$

$$= 2\left\langle\tilde{I}_1\tilde{I}_2^*\right\rangle^2, \tag{68}$$

where in the second equality we use the fact that $\langle\tilde{I}_i(\ell)\tilde{I}_j(\ell)\rangle = \langle\tilde{I}_i(\ell)\tilde{I}_j^*(-\ell)\rangle$, which vanishes in expectation for real valued fields. It follows that

$$\left\langle\tilde{I}_1\tilde{I}_1^*\tilde{I}_2\tilde{I}_2^*\right\rangle = \left\langle\tilde{I}_1\tilde{I}_1^*\right\rangle\left\langle\tilde{I}_2\tilde{I}_2^*\right\rangle + \left\langle\tilde{I}_1\tilde{I}_2\right\rangle\left\langle\tilde{I}_1^*\tilde{I}_2^*\right\rangle + \left\langle\tilde{I}_1\tilde{I}_2^*\right\rangle\left\langle\tilde{I}_1^*\tilde{I}_2\right\rangle \tag{69}$$

$$= \left\langle\tilde{I}_1\tilde{I}_1^*\right\rangle\left\langle\tilde{I}_2\tilde{I}_2^*\right\rangle + \left\langle\tilde{I}_1\tilde{I}_2^*\right\rangle\left\langle\tilde{I}_1^*\tilde{I}_2\right\rangle. \tag{70}$$

Now we have the following expression for the variance of the estimator:

$$\delta\hat{C}_{1,2}^2 = \frac{\Omega^2}{4}\left[4\left\langle\tilde{I}_1\tilde{I}_2^*\right\rangle^2 + 2\left(\left\langle\tilde{I}_1\tilde{I}_1^*\right\rangle\left\langle\tilde{I}_2\tilde{I}_2^*\right\rangle + \left\langle\tilde{I}_1\tilde{I}_2^*\right\rangle\left\langle\tilde{I}_1^*\tilde{I}_2\right\rangle\right) - \left\langle\tilde{I}_1\tilde{I}_2^* + \tilde{I}_1^*\tilde{I}_2\right\rangle^2\right]. \tag{71}$$

We can begin identifying more familiar quantities now that we are dealing with two point functions. We have auto spectra $C_{i,i}^2(\ell) = \Omega^2\langle\tilde{I}_i\tilde{I}_i^*\rangle^2$ along with cross spectra $C_{i,j}^2(\ell) = \Omega^2\langle\tilde{I}_i\tilde{I}_j^*\rangle\langle\tilde{I}_i^*\tilde{I}_j\rangle = \Omega^2\langle\tilde{I}_i\tilde{I}_j^*\rangle^2$. We can write the estimated variance in terms of these power spectra:

$$\delta\hat{C}_{1,2}^2 = \frac{1}{4}\left[4C_{1,2}^2 + 2\left(C_1C_2 + C_{1,2}^2\right) - 4C_{1,2}^2\right] = \frac{1}{2}(C_1C_2 + C_{1,2}^2). \tag{72}$$

The statistical error on $\tilde{C}_\ell$ can be calculated by noting that each (ℓ, m) mode is independent. However, because our maps are real valued, the Fourier transform should be symmetric. For one cross spectrum mode, we can see this as

$$(a_\ell b^*_\ell + a^*_\ell b_\ell)_m = a_{\ell m} b^*_{\ell m} + a^*_{\ell m} b_{\ell m} \tag{73}$$

$$= a_{\ell,-m} b^*_{\ell,-m} + a^*_{\ell,-m} b_{\ell,-m} \tag{74}$$

$$= (a_\ell b^*_\ell + a^*_\ell b_\ell)_{-m}. \tag{75}$$

This leads to a reduction in the effective number of independent modes by a factor of two, hence

$$\delta\langle\hat{C}_{1,2}(\ell)\rangle^2_m = \frac{2}{2\ell+1}\delta\hat{C}^2_{1,2}(\ell) = \frac{2}{2\ell+1}\left[\frac{1}{2}(C_1(\ell)C_2(\ell) + C^2_{1,2}(\ell))\right]. \tag{76}$$

In the case where $C_1 = C_2$ (i.e., the auto-spectrum), this reduces to

$$\delta C^2_\ell = \frac{2}{2\ell+1} C^2_\ell \tag{77}$$

while for the cross spectrum this is

$$\delta C^2_{1,2}(\ell) = \frac{2}{2\ell+1}\left[\frac{1}{2}(C_1(\ell)C_2(\ell) + C^2_{1,2}(\ell))\right] \tag{78}$$

$$= \frac{1}{2\ell+1}\left[C_1(\ell)C_2(\ell) + C^2_{1,2}(\ell)\right]. \tag{79}$$

References

[46] Y. T. C. and T. C. C. In: *ApJ* 925.2, 136 (Feb. 2022), p. 136. DOI: 10.3847/1538-4357/ac3aee. arXiv: 2109.10914 [astro-ph.CO].

[61] Y. K. C., B. M., and D. S. In: *ApJ* 877.2, 150 (June 2019), p. 150. DOI: 10.3847/1538-4357/ab1b35. arXiv: 1810.00885 [astro-ph.CO].

[62] S. D., Y. K. C., and S. M. In: *arXiv e-prints*, arXiv:2305.12353 (May 2023), arXiv:2305.12353. DOI: 10.48550/arXiv.2305.12353. arXiv: 2305.12353 [astro-ph.GA].

[63] A. C. In: *Royal Society Open Science* 3.3 (Mar. 2016), p. 150555. DOI: 10.1098/rsos.150555. arXiv: 1602.03512 [astro-ph.CO]. URL: https://ui.adsabs.harvard.edu/abs/2016RSOS....350555C.

[64] T. C. et al. In: *AJ* 164.5, 170 (Nov. 2022), p. 170. DOI: 10.3847/1538-3881/ac8d02. arXiv: 2205.06347 [astro-ph.CO]. URL: https://ui.adsabs.harvard.edu/abs/2022AJ....164..170C.

[65] R. H., K. W. M., and D. S. In: *Applied Spectroscopy* 72.5 (May 2018), pp. 663–688. DOI: 10.1177/0003702818767133. arXiv: 1802.03694 [astro-ph.CO]. URL: https://ui.adsabs.harvard.edu/abs/2018ApSpe..72..663H.

[66] S. M. et al. In: *ApJ* 839.1, 7 (Apr. 2017), p. 7. DOI: 10.3847/1538-4357/aa6843. arXiv: 1704.07166 [astro-ph.GA]. URL: https://ui.adsabs.harvard.edu/abs/2017ApJ...839....7M.

[67] M. S. A. et al. In: *MNRAS* 489.1 (Oct. 2019), pp. 1120–1126. DOI: 10.1093/mnras/stz2186. arXiv: 1908.02260 [astro-ph.GA]. URL: https://ui.adsabs.harvard.edu/abs/2019MNRAS.489.1120A.

[68] H. E. S. S. Collaboration et al. In: *A&A* 606, A59 (Oct. 2017), A59. DOI: 10.1051/0004-6361/201731200. arXiv: 1707.06090 [astro-ph.HE]. URL: https://ui.adsabs.harvard.edu/abs/2017A&A...606A..59H.

[69] V. A. A. et al. In: *MNRAS* 486.3 (July 2019), pp. 4233–4251. DOI: 10.1093/mnras/stz943. arXiv: 1904.00134 [astro-ph.HE]. URL: https://ui.adsabs.harvard.edu/abs/2019MNRAS.486.4233A.

[70] A. K. et al. In: *ApJ* 753, 63 (July 2012), p. 63. DOI: 10.1088/0004-637X/753/1/63. arXiv: 1201.5617. URL: http://adsabs.harvard.edu/abs/2012ApJ...753...63K.

[71] R. G. A. et al. In: *ApJ* 824.1, 26 (June 2016), p. 26. DOI: 10.3847/0004-637X/824/1/26. arXiv: 1604.07291 [astro-ph.CO].

[72] P. A., C. L., and D. L. "Search for brightness fluctuations in the zodiacal light at 25 MU M with ISO". In: *A&A* 328 (Dec. 1997), pp. 702–705.

[73] M. Z. et al. In: *Science* 346.6210 (Nov. 2014), pp. 732–735. DOI: 10.1126/science.1258168. arXiv: 1411.1411 [astro-ph.CO].

[74] Y. T. C. and J. J. B. In: *arXiv e-prints*, arXiv:2207.13712 (July 2022), arXiv:2207.13712. arXiv: 2207.13712 [astro-ph.CO].

[75] Planck Collaboration et al. In: *A&A* 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[76] M. Z. et al. In: *ApJS* 207.2, 31 (Aug. 2013), p. 31. DOI: 10.1088/0067-0049/207/2/31. arXiv: 1112.1424 [astro-ph.IM].

[77] J. B. et al. In: *ApJS* 207.2, 32 (Aug. 2013), p. 32. DOI: 10.1088/0067-0049/207/2/32. arXiv: 1206.4702 [astro-ph.IM].

[78] K. T. et al. In: *ApJS* 207.2, 33 (Aug. 2013), p. 33. DOI: 10.1088/0067-0049/207/2/33. arXiv: 1112.4217 [astro-ph.IM].

[79] K. H., M. R., and A. K. In: *ApJ* 752.2, 113 (June 2012), p. 113. DOI: 10.1088/0004-637X/752/2/113. arXiv: 1201.4398 [astro-ph.CO].

[80] J. C., M. W., and I. S. In: *MNRAS* 444.1 (Oct. 2014), pp. 994–1000. DOI: 10.1093/mnras/stu1527. arXiv: 1406.6072 [astro-ph.CO].

[81] P. C. and B. J. In: *MNRAS* 248 (Jan. 1991), pp. 1–13. DOI: 10.1093/mnras/248.1.1.

[82] T. K. et al. In: *ApJ* 508.1 (Nov. 1998), pp. 44–73. DOI: 10.1086/306380. arXiv: astro-ph/9806250 [astro-ph].

[83] B. P. C. et al. In: *Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave*. Ed. by Makenzie Lystrup et al. SPIE, Dec. 2020. DOI: 10.1117/12.2567224.

[84] L. G. et al. In: *A&A* 436.3 (June 2005), pp. 895–915. DOI: 10.1051/0004-6361:20042352. arXiv: astro-ph/0504047 [astro-ph].

[85] Y. T. C. et al. In: *ApJ* 919.2, 69 (Oct. 2021), p. 69. DOI: 10.3847/1538-4357/ac0f5b. arXiv: 2103.03882 [astro-ph.GA].

[86] J. D. G. and W. J. F. In: *Infrared Detectors and Instrumentation*. Ed. by Albert M. Fowler. Vol. 1946. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Oct. 1993, pp. 395–404. DOI: 10.1117/12.158692.

[87] E. H. et al. In: *ApJ* 567.1 (Mar. 2002), pp. 2–17. DOI: 10.1086/338126. arXiv: astro-ph/0105302 [astro-ph].

[88] J. S. -Y. L. et al. In: *ApJ* 928.2, 109 (Apr. 2022), p. 109. DOI: 10.3847/1538-4357/ac562f. arXiv: 2111.01113 [astro-ph.CO].

[91] D. L. In: *AJ* 147.5, 108 (May 2014), p. 108. DOI: 10.1088/0004-6256/147/5/108. arXiv: 1405.0308 [astro-ph.IM].

[92] E. F. S., A. M. M., and G. M. G. In: *ApJS* 240.2, 30 (Feb. 2019), p. 30. DOI: 10.3847/1538-4365/aafbea. arXiv: 1901.03337 [astro-ph.IM].

[93] A. L. et al. In: *MNRAS* 379.4 (Aug. 2007), pp. 1599–1617. DOI: 10.1111/j.1365-2966.2007.12040.x. arXiv: astro-ph/0604426 [astro-ph].

[94] C. L. et al. In: *ApJS* 224.2, 24 (June 2016), p. 24. DOI: 10.3847/0067-0049/224/2/24. arXiv: 1604.02350 [astro-ph.GA].

[95] M. L. et al. In: *PhRvD* 106.2, 023525 (July 2022), p. 023525. DOI: 10.1103/PhysRevD.106.023525. arXiv: 2109.13911 [astro-ph.CO].

[96] K. M. S., O. H. E. P, and J. C. H. In: *arXiv e-prints*, arXiv:2302.05436 (Feb. 2023), arXiv:2302.05436. arXiv: 2302.05436 [astro-ph.CO].

[97] R. M. F. et al. In: *arXiv e-prints* (Dec. 2023). DOI: 10.48550/arXiv.2312.04636. arXiv: 2312.04636 [astro-ph.CO].

[98] M. S. M. et al. In: *PhRvD* 102.2, 023534 (July 2020), p. 023534. DOI: 10.1103/PhysRevD.102.023534. arXiv: 1911.05717 [astro-ph.CO].

[99] F. M. and J. C. H. In: *PhRvD* 109.2, 023528 (Jan. 2024), p. 023528. DOI: 10.1103/PhysRevD.109.023528. arXiv: 2307.01043 [astro-ph.CO].

[100] R. M. F. et al. In: *AJ* 166.3, 98 (Sept. 2023), p. 98. DOI: 10.3847/1538-3881/ace69b. arXiv: 2307.10385 [astro-ph.IM].

[101] C. H. N. et al. In: *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*. Ed. by Makenzie Lystrup et al. Vol. 10698. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2018, 106984J, 106984J. DOI: 10.1117/12.2311595.

[102] M. S. et al. In: *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*. Ed. by Howard A. MacEwen et al. Vol. 9904. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2016, 99044J, 99044J. DOI: 10.1117/12.2229567.

[103] L. K. In: *PhRvD* 52.8 (Oct. 1995), pp. 4307–4318. DOI: 10.1103/PhysRevD.52.4307. arXiv: astro-ph/9504054 [astro-ph].

[105] A. C. et al. In: *arXiv e-prints*, arXiv:1210.6031 (Oct. 2012), arXiv:1210.6031. DOI: 10.48550/arXiv.1210.6031. arXiv: 1210.6031 [astro-ph.CO].

[107] K. H. et al. In: *ApJ* 785.1, 38 (Apr. 2014), p. 38. DOI: 10.1088/0004-637X/785/1/38. arXiv: 1311.1254 [astro-ph.CO].

[108] A. C. et al. In: *ApJ* 756.1, 92 (Sept. 2012), p. 92. DOI: 10.1088/0004-637X/756/1/92. arXiv: 1205.2316 [astro-ph.CO].

[109] B. Y. et al. In: *MNRAS* 433.2 (Aug. 2013), pp. 1556–1566. DOI: 10.1093/mnras/stt826. arXiv: 1305.5177 [astro-ph.CO].

[110] B. Y. et al. In: *MNRAS* 440.2 (May 2014), pp. 1263–1273. DOI: 10.1093/mnras/stu351. arXiv: 1402.5675 [astro-ph.CO].

[111] Y. G. et al. In: *ApJ* 825.2, 104 (July 2016), p. 104. DOI: 10.3847/0004-637X/825/2/104. arXiv: 1511.01577 [astro-ph.CO].

[112] O. E. K., A. K., and E. V. In: *PhRvD* 99.2, 023002 (Jan. 2019), p. 023002. DOI: 10.1103/PhysRevD.99.023002. arXiv: 1808.05613 [hep-ph].

[113] A. C. et al. In: *JCAP* 2021.5, 046 (May 2021), p. 046. DOI: 10.1088/1475-7516/2021/05/046. arXiv: 2012.09179 [astro-ph.CO].

[114] M. C. et al. In: *MNRAS* 453.2 (Oct. 2015), pp. 1513–1530. DOI: 10.1093/mnras/stv1708. arXiv: 1312.2013 [astro-ph.CO].

[115] P. M. K. et al. In: *ApJS* 207.2, 34 (Aug. 2013), p. 34. DOI: 10.1088/0067-0049/207/2/34. arXiv: 1307.0016 [astro-ph.CO].

[116] P. M. K. et al. In: *ApJ* 926.2, 133 (Feb. 2022), p. 133. DOI: 10.3847/1538-4357/ac44ff. arXiv: 2104.07104 [astro-ph.EP].

[117] C. H. N. PhD thesis. Rochester Institute of Technology, New York, Jan. 2021.

[118] Z. L. W., J. L. H., and F. S. L. In: *ApJS* 199.2, 34 (Apr. 2012), p. 34. DOI: 10.1088/0067-0049/199/2/34. arXiv: 1202.6424 [astro-ph.CO].

[119] Y. K. C. In: *arXiv e-prints*, arXiv:2306.03926 (June 2023), arXiv:2306.03926. DOI: 10.48550/arXiv.2306.03926. arXiv: 2306.03926 [astro-ph.GA].

[121] M. D., M. -A. and D., P. -A. and M., F. and C., J. -C. and D., P. and G., S. and K., E. In: *A&A* 593, A4 (Aug. 2016), A4. DOI: 10.1051/0004-6361/201628503. arXiv: 1605.08360 [astro-ph.GA].

[122] III G. T. N. et al. In: *AJ* 103 (Apr. 1992), p. 1313. DOI: 10.1086/116144.

[123] K. M. W. et al. In: *Nature Communications* 6, 7945 (Sept. 2015), p. 7945. DOI: 10.1038/ncomms8945. arXiv: 1509.02935 [astro-ph.CO].

[124] D. J. S., D. P. F., and M. D. In: *ApJ* 500.2 (June 1998), pp. 525–553. DOI: 10.1086/305772. arXiv: astro-ph/9710327 [astro-ph].

[125] Y. K. C. and B. M. In: *ApJ* 870.2, 120 (Jan. 2019), p. 120. DOI: 10.3847/1538-4357/aaf4f6. arXiv: 1808.03294 [astro-ph.GA].

[127] Y. O. et al. In: *PASJ* 70.4, 76 (Aug. 2018), p. 76. DOI: 10.1093/pasj/psy070. arXiv: 1806.08891 [astro-ph.GA].

[128] M. L. N. A. et al. In: *ApJ* 701.1 (Aug. 2009), pp. 428–453. DOI: 10.1088/0004-637X/701/1/428. arXiv: 0906.0024 [astro-ph.CO].

[129] M. T. S., M. M., and B. M. P. In: ed. by David R. Silva and Rodger E. Doxsey. Vol. 6270. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. June 2006, 627020, p. 627020. DOI: 10.1117/12.673134.

[132] K. L. P. et al. In: (July 2023). arXiv: 2307.12584 [astro-ph.GA].

Chapter 3

MEASURING NIR EXTRAGALACTIC BACKGROUND LIGHT ANISOTROPIES WITH CIBER. II: CLUSTERING RESULTS FROM 4TH FLIGHT DATA

Abstract: We present new clustering results in the near-infrared (NIR) using imaging data at 1.1 and 1.8 μm from the fourth flight of the Cosmic Infrared Background ExpeRiment (*CIBER*). We detect large-angle surface brightness fluctuations at $\theta > 5'$ with *CIBER* at high significance, finding internal consistency in both auto- and cross-power spectra. Through cross-correlations with tracers of diffuse galactic light (DGL), we determine that DGL (as traced by existing extinction maps) does not explain the observed *CIBER* fluctuations, contributing $< 10\%$ to the total fluctuation power on scales $\ell < 1000$. To assess the coherence of sky fluctuations across instruments, we cross-correlate two of our five *CIBER* fields with IRAC mosaics from the *Spitzer* Deep Wide Fast Survey (SDWFS) at 3.6 μm and 4.5 μm. While we are unable to obtain reliable *Spitzer* clustering measurements below $\ell \sim 1000$, we measure clear departures from Poisson clustering in both the *Spitzer* auto spectrum and *CIBER* $\times$ *Spitzer* cross-spectra. The measured spectrum of large-angle fluctuations at $1000 < \ell < 2000$ is bluer than that from small-scale Poisson fluctuations, which, if interpreted astrophysically, could be indicative of a population with active star formation and/or low dust extinction. A full interpretation of these excess fluctuations requires fitting models that realistically incorporate galaxy biasing, non-linear IGL clustering and potentially additional intra-halo light (IHL) clustering, which we plan to do in future work.

3.1 Introduction

Characterizing the NIR extragalactic background light (EBL) through its spatial and spectral distribution offers a window into the astrophysical processes that drive cosmic light production. While many measurements of the EBL have focused on measuring its intensity monopole through absolute photometry. However, this class of measurements is plagued by issues disentangling true extragalactic components from local foregrounds. In particular, errors in zodiacal light (ZL) and diffuse galactic light (DGL) subtraction can lead to order unity errors at optical and NIR wavelengths, since these foregrounds are much brighter than the EBL signal of

interest, even at large heliocentric distances [104].

Fluctuation-based measurements largely bypass these challenges due to the fact that ZL is spatially smooth on sub-degree angular scales [71]. A great deal of information can be extracted from the EBL fluctuation component beyond its monopole, as fluctuation measurements are sensitive to the scale-dependent clustering of galaxies and potentially unidentified EBL components. Fluctuations on small scales are driven by the Poisson noise of discrete sources and provide a form of integral constraint on the galaxy luminosity function (LF) [79]. On arcminute to degree scales, EBL fluctuations encode the linear ("two-halo") clustering of galaxies that trace large-scale structure and non-linear galaxy clustering ("one-halo"). Through cross-correlations, fluctuation measurements are more immune to noise biases and enable a spatial-spectral decomposition that facilitates further component separation. Such infrared fluctuation measurements have been pursued with *Spitzer* [105], the *AKARI* satellite [106], along with second and third flight data from the Cosmic Infrared Background ExpeRiment [*CIBER*; 73]. While EBL fluctuation studies are still at early stages of development and carry their own set of systematics, these measurements display departures in large-angle fluctuation power from Poisson noise of integrated galaxy light (IGL) alone, with a blue spectrum rising from 4 μm to 1.1 μm.

Many astrophysical scenarios have been proposed to explain the measured NIR EBL. One possibility is that some known astrophysical populations (e.g., galaxies, dwarf galaxies) are not fully accounted for by current models [107]. Another explanation is that diffuse intra-halo light (IHL) in the outskirts of galaxies is abundant and contributes to EBL fluctuations on larger angular scales [85]. At higher redshifts, a signal from the Lyman-α break associated with epoch of reionization (EoR) galaxies is expected [108], however its amplitude is expected to be several orders of magnitude smaller than existing *CIBER* fluctuation measurements [73]. More exotic explanations include, but are not limited to, emission from direct collapse black holes (DCBHs) at high redshift [109, 110] and eV-scale axion dark matter with a coupling to light [111, 112, 113].

In this work we apply an improved methodology for measuring EBL anisotropies (detailed in Feder et *al.* 2024a, hereafter Paper I) to 1.1 μm and 1.8 μm imaging data from the fourth and final flight of *CIBER*-1. In Paper I, we develop a power spectrum formalism to quantify and correct for flat field (FF) errors and make improvements in effective source masking, which is required to separate true large-angle clustering

from the Poisson noise fluctuations.

This is not the first study to use imaging data from *CIBER*-1's fourth flight. Recently, [85] performed galaxy stacking on this dataset and, through detailed modeling of the sub-pixel point spread function, placed constraints on the 1- and 2-halo terms of an IHL model painted on MICECAT simulations [114] assuming a NFW light profile.

The paper is organized as follows. We introduce *CIBER* in §2 and describe the observations and map pre-processing steps in §3. In §4 we detail the construction of our science masks, while in §5 we describe our read and photon noise models with corresponding validation using flight data. We summarize the pseudo-power spectrum pipeline used in this analysis in §6. We then present clustering results from observed data in §7 and §8, including *CIBER* auto- and cross-spectra, cross-correlations with tracers of DGL, and *Spitzer* mosaics in the Boötes fields. We conclude with a discussion of our results and avenues for future work in §9.

Throughout this work we assume a flat ΛCDM cosmology with $n_s = 0.97$, $\sigma_8 = 0.82$, $\Omega_m = 0.26$, $\Omega_b = 0.049$, $\Omega_\Lambda = 0.69$ and $h = 0.68$, consistent with measurements from *Planck* [75]. All fluxes are quoted in the Vega magnitude system unless otherwise specified.

3.2 Cosmic Infrared Background Experiment (CIBER)

CIBER[1] is a rocket-borne instrument designed to characterize the NIR EBL through measurements of its spatial fluctuations and electromagnetic spectrum [76]. Four instruments were flown as part of *CIBER*: two wide-field imagers [77], a narrow-band spectrometer [115, 116], and a low-resolution spectrometer [78]. In this paper we concentrate on data from the final (non-recovered) *CIBER* flight that launched at 3:05 UTC 2013 June 6 from Wallops Flight Facility on a four-stage Black Brant XII rocket. Unlike in the first three flights, the payload during the fourth flight achieved altitudes of 550 km (compared to $\sim$ 330 km), resulting in a longer total exposure time and lower levels of airglow contamination. Our analysis utilizes the last five of eight observed *CIBER* fields, following three which are removed due to airglow and sunlight contamination. Each imager is comprised of a 1024×1024 pixel Hawaii-1 detector with a 2×2 deg^2 instantaneous field of view. Table 3.1 contains information on the coordinates and integration times of the five observed fields we retain for this analysis.

[1] `https://ciberrocket.github.io/`

Field	(R.A., Decl.) (deg.)	(l, b) (deg.)	(λ, β) (deg.)	t_{exp} (s)	N_{frame}	ZL at 1.1 μm (nW m^{-2} sr^{-1})	ZL at 1.8 μm (nW m^{-2} sr^{-1})
elat10	(190.5, 8.0)	(295.8, 70.7)	(186.5,11.5)	42.7	24	660	350
elat30	(193.1, 28.0)	(109.2, 89.1)	(179.8,30.7)	17.8	10	470	260
Boötes A	(218.4, 34.8)	(58.6, 66.9)	(200.5,46.6)	51.6	29	350	190
Boötes B	(217.2, 33.2)	(55.0, 68.2)	(200.3,44.7)	49.8	28	345	185
SWIRE	(242.8, 54.8)	(84.6, 44.7)	(208.9,72.5)	44.5	25	285	150

Table 3.1: Table of science fields used in analysis. We include the celestial, galactic and ecliptic coordinates of our fields, the exposure times t_{exp}, and the mean Zodiacal light levels as determined by the [82] model. Coordinates for each field are in reference to the central pointing of the 1.1 μm imager.

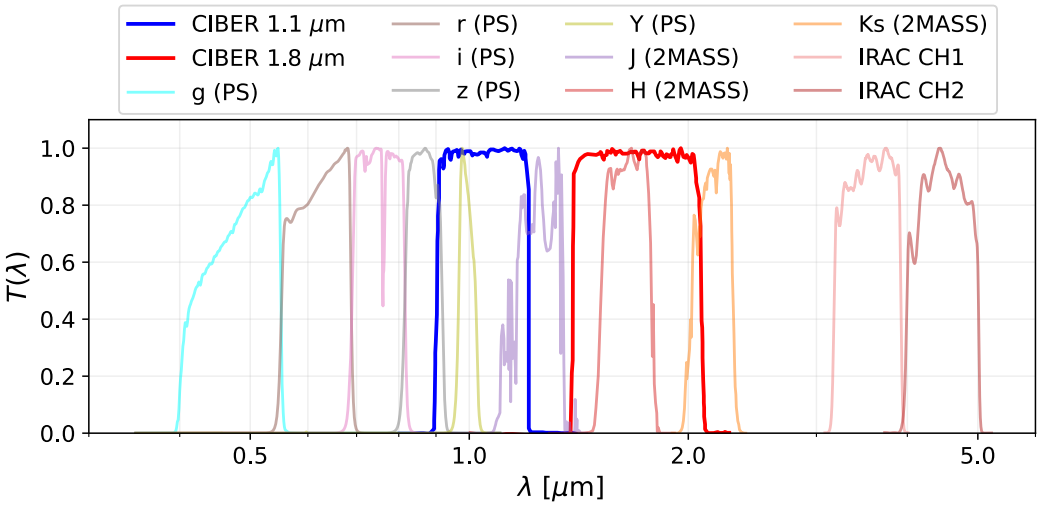

Figure 3.1: Normalized filter profiles for different photometric data used in this work. The *CIBER* imagers (blue, red) are centered at 1.05 μm and 1.79 μm, closest to yJH bands from PanSTARRS and 2MASS. In §3.8 we cross-correlate against IRAC mosaics from the Spitzer Deep Wide Fast Survey (SDWFS) at 3.6 μm and 4.5 μm.

3.3 Data pre-processing

Time stream filtering and slope fits

We employ time stream filtering to remove pickup noise from the readout electronics, which is typically most severe in dark exposures taken with ground power supply. The CIBER-1 Hawaii detectors are read out in four-channel mode, where each channel corresponds to one detector quadrant. Inspecting the time-ordered data from each separate quadrant and computing its 1D power spectrum, we identify noisy frequencies for the TM1 (1.1 μm) and TM2 (1.8 μm) detectors at $\sim$ 9.503 Hz and 9.538 Hz, respectively, corresponding to an angular scale of $\ell \sim 6200$ [117]. We use a notch filter centered on the noisy frequencies with a bandwidth of 0.1 Hz, which reduces the associated noise power by a factor of $\gtrsim 10^2$. The transfer function for these narrow-band filters is assumed to be near unity.

We then perform slope fits starting two frames after each global reset. Due to rocket pointing instability during the first half of the elat30 field exposure, we only use the last ten frames of the integration. For a very small fraction of pixels with high photocurrent, we correct the slope fits for ADU register overflows.

Dark current subtraction

It is necessary to correct for the non-zero response of the detectors in the absence of photons, known as dark current, which is sourced by thermally produced charge carriers and multiplexer glow. To correct for the dark current from each imager we average the set of dark exposures taken on the rail before flight and subtract the mean template from each photocurrent map. The mean dark current level is ~ 0.2 e$^-$ s^{-1} for both detectors. Given the relatively low dark current levels, we assume that any residual errors in our templates, which can introduce noise biases, are negligible.

Gain calibration

The observed surface brightness λI_λ is related to the digitized detector output i through

$$\lambda I_\lambda \, [\text{nW m}^{-2}\text{sr}^{-1}] = g_a g_1 g_2 i \, [\text{ADU fr}^{-1}], \tag{3.1}$$

where g_a is the amplification gain from Volts to ADU frame^{-1} and is chosen as part of the electronics system design [76]. The second factor, g_1, converts from units of ADU frame^{-1} to photocurrent (e$^-$ s^{-1}), while g_2 converts between photocurrent to surface brightness units.

Estimation of g_1

Following App. B in [117] we estimate the gain factor g_1 using the noise statistics of the flight data. In the photon-noise dominated limit, the per-pixel root-mean-squared (RMS) in digital units, σ_{pix}^{dig} (ADU frame^{-1}) is linearly proportional to the photocurrent noise RMS by g_1, such that measuring the σ_{pix}^{dig} in exposures with varying mean levels allows us to fit directly for g_1.

We estimate g_1 using 22 readout frames from each of the science fields For each field, we compute a difference image from 11-frame masked half-exposures. We restrict the set of pixels from each exposure to those with relative gain in the range $0.95 \leq FF \leq 1.05$ and also make cut on pixels with correlated double sample (CDS) noise more than 3σ away from the median CDS. Once the noise variance in each field is calculated we perform jackknife resampling across the fields to estimate g_1 and its uncertainty to be -2.67 ± 0.02 at 1.1 μm and -3.04 ± 0.02 at 1.8 μm. We find similar results using pre-flight optics test data ($g_1 = -2.5 \pm 0.02$ and -2.8 ± 0.1 for 1.1 and 1.8 μm, respectively), however we use g_1 estimates from the flight data as they capture the detector condition most faithfully. In §3.7 we show the impact of different assumed g_1 factors on the *CIBER* auto power spectra.

Absolute gain calibration

We perform point source flux calibration for each imager to estimate the total calibration factor (i.e., $g_1 g_2$) going from ADU frame^{-1} to nW m^{-2} sr^{-1}. We begin by identifying all 2MASS sources with magnitudes $11 < J < 14.5$. The bright end is chosen to avoid non-linear detector flux response above an integrated charge of 7.5×10^4 e-, while the faint end ensures well-determined fluxes for calibrator sources. We then perform aperture photometry in 13×13 pixel sub-regions centered on the sources. The sub-region size is chosen to be large enough to obtain reliable background estimates, however to mitigate biases from nearby bright sources we remove any calibration sources with bright ($J < 16$) neighbors in the same regions. After cutting on sources with neighbors and postage stamps and requiring the masking fraction to be $< 5\%$ in each postage stamp, we are left with between $\sim 500 - 800$ sources per field in each band. The aperture fluxes, combined with the predicted *CIBER* source flux yield an estimate for the absolute calibration.

We derive the absolute calibration by comparing predicted *CIBER* fluxes with fluxes estimated from aperture photometry on *CIBER* images at the positions of 2MASS sources, which are related by

$$\lambda F_\lambda [\text{nW m}^{-2}] = \int \lambda I_\lambda(\theta, \phi) d\Omega. \tag{3.2}$$

Choosing a suitable region of integration around the source and invoking the flat-sky approximation, we perform a change of variables to detector coordinates,

$$\lambda F_\lambda = \int \int \lambda I_\lambda(x, y) \Omega_{pix} dx dy. \tag{3.3}$$

While we have access to fluxes from the 2MASS catalog, the *CIBER* maps are in digital units. The intensity map is then $\lambda I_\lambda(x, y) = g_2(\lambda) i_{phot}^{CIBER}(x, y)$, meaning we can solve the gain for each 2MASS source i with

$$g_1 g_2(\lambda) = \frac{\lambda \hat{F}_\lambda^{2MASS}}{\Omega_{pix} \int \int i_{phot}^{CIBER}(x, y) dx dy}. \tag{3.4}$$

We assume the *CIBER* PSF is sum normalized, i.e., $\int_{d\Omega} P(\theta, \phi) = 1$.

Accurate surface brightness calibration is complicated both because the 2MASS fluxes are observed over a different integrated bandpass than the *CIBER* filters, and because the diffuse sky component may in general have a different spectrum, and thus different effective wavelength, compared to the point sources used for

calibration. For each calibration source, we use cross-matched photometry from PanSTARRS and 2MASS (which are described further in §6 of Paper I), convert to monochromatic flux densities, and fit a smooth spline to model the notional source spectrum. We use the model SEDs to interpolate to the *CIBER* filter central wavelengths and to calculate the bandpass correction needed to predict the flux response R,

$$R = \frac{\int \lambda I(\lambda) F(\lambda) d\lambda}{I(\lambda_{eff})\Delta\lambda}. \tag{3.5}$$

The interpolation to *CIBER* fluxes affects the calibration at the 10-20% level and varies depending on the spectral type of the calibration source.

There is an additional correction related to field distortions across each detector which modify the per-pixel etendue $(A\Omega)_{pix}$. We use the WCS solutions for each field to estimate this field distortion and find $d\theta/dx = 6.99'' - 7.03''$, with smooth variation as a function of detector position. This corresponds to a variation in $(A\Omega)_{pix}$ of $< 1\%$, suggesting the effect is negligible at our absolute calibration accuracy. We compare the set of measured *CIBER* fluxes against predicted fluxes in Fig. 3.3. We fit a slope to each set of sources and perform iterative sigma clipping to reduce the impact of flux outliers on each gain solution. We obtain consistent gain estimates from the different science fields and also validate that the gain is relatively unaffected by the chosen minimum flux cut. We then calculate the final gains by taking a weighted mean of the per-field estimates.

To validate our absolute gain calibration we convert the *CIBER* maps to surface brightness units, mask bright sources (source masking is detailed in 3.4) calculate the mean surface brightness in each map and correlate these with ZL estimates predicted by the [82] model. We compute an additional bandpass correction given the notional ZL spectrum. The spectrum is blue at *CIBER* wavelengths however in practice the bandpass correction is at the few percent level. In Figure 3.3 we show the results of this comparison. The five science fields have different ZL levels which gives a lever arm to measure *CIBER* × Kelsall slopes. We fit an offset and slope for each band and find slopes consistent with unity within $\sim 10\%$.

We perform the same exercises using separate fits from the four quadrants of each detector. The best fit slopes are shown in Table 3.2 along with those derived from the full arrays, use jackknife resampling to estimate uncertainties. While the per quadrant fits have slightly lower precision, the distribution of slopes indicate internal consistency of our gain solutions, though this is not strictly expected.

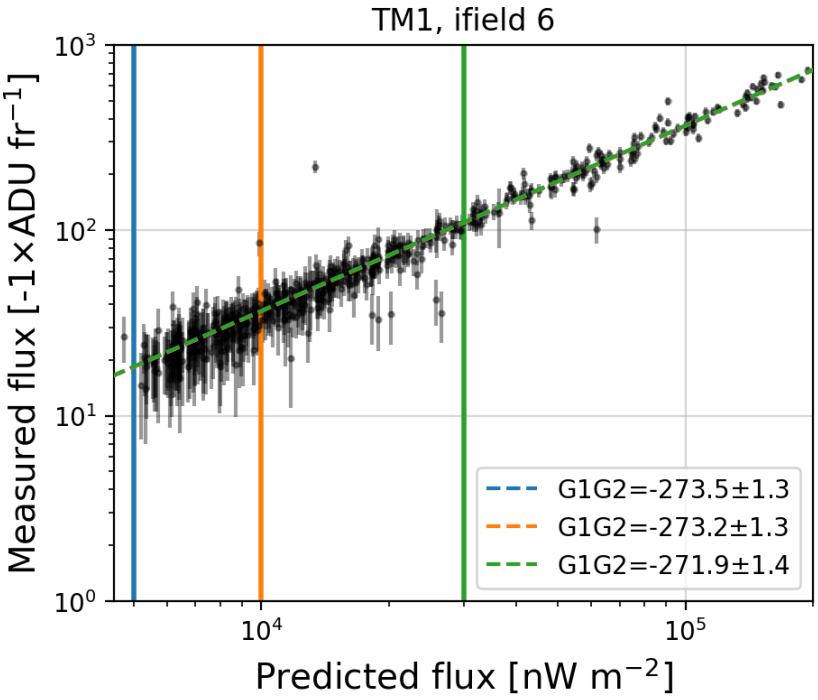

Figure 3.2: Comparison of predicted and measured *CIBER* fluxes for 1.05 μm in the Boötes field. The colored vertical lines indicate the minimum flux thresholds corresponding to the gain estimates in the legend. $g_1 g_2$ has units of (nW m^{-2} sr^{-1})/(ADU frame^{-1}).

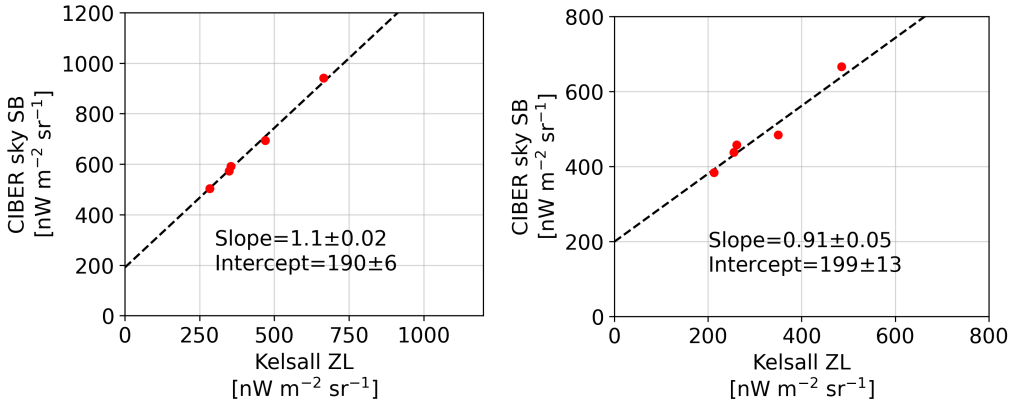

Figure 3.3: *CIBER* - Kelsall surface brightness calibration, for 1.05 μm (left) and 1.79 μm (right). The correlation between *CIBER* and Kelsall across fields indicates consistency at the 10% level for both imagers with DIRBE.

CIBER band	Detector subset	N_{src}	Best fit slope
TM1 (1.1 μm)	Quadrant A	735	1.04 ± 0.03
	Quadrant B	715	1.09 ± 0.03
	Quadrant C	668	1.13 ± 0.02
	Quadrant D	710	1.17 ± 0.03
	Full array	2828	**1.10 ± 0.02**
TM2 (1.8 μm)	Quadrant A	818	0.89 ± 0.05
	Quadrant B	638	0.95 ± 0.06
	Quadrant C	730	0.90 ± 0.05
	Quadrant D	588	0.99 ± 0.06
	Full array	2774	**0.91 ± 0.05**

Table 3.2: *CIBER* - Kelsall surface brightness comparison, estimated in separate quadrants and from the full arrays. Slopes consistent with unity indicate strong correlation between the measured and model surface brightness. N_{src} denotes the total number of sources across the five fields in each subset.

The extrapolation of our fits in Fig. 3.3 to $I_{ZL}^{Kelsall} \rightarrow 0$ implies the presence of a mean surface brightness component that is $\sim$ 200 nW m^{-2} sr^{-1} in both *CIBER* bands. This is higher than that expected from the combined monopole intensity of DGL, stars and galaxies. One potential explanation is that the off-axis response to bright stars near the field of view introduces additional signal in our maps. This can be estimated with knowledge of the off-axis telescope gain $g(\theta)$ and the ISL distribution. However, such estimates are sensitive to the off-axis response for small θ, which is difficult to estimate through laboratory measurements, along with the specific distribution of bright stars near the FOV.

FF correction and image filtering

Two more corrections are needed to estimate the *CIBER* power spectrum. The first, removal of fluctuations from Zodiacal light, is done by filtering out low ℓ modes with a gradient subtraction step. This assumes that ZL (and any other foregrounds) are relatively smooth on the scales probed by *CIBER*. The second is correction for relative gain variations across each imager, otherwise known as the FF responsivity.

As discussed in Paper I, our ability to reliably measure the power spectra of individual fields comes from our revised treatment of the FF and its errors. We use the flight data directly to estimate and correct for the FF with a stacking estimator, in which each field's mean background (which is dominated by ZL) acts as an approximately uniform source of illumination. This stacking estimator has errors related to the

scale-dependent sky and instrument fluctuations in the exposures and scale inversely with the total number of exposures. In Paper I, we model the effects of FF errors on the pseudo-power spectrum and show that the underlying fluctuation power can be recovered through a modified noise bias and mode mixing correction.

Building a FF estimator using flight data involves masking out bright sources within each field, and then stacking combinations of exposures normalized by their means. While the astrophysical signals will vary across images, the FF responsivity will not, and so in the limit of many independent exposures this estimator should converge to the true FF. However, any sky gradients in the stacked fields will be partially degenerate with large-scale FF responsivity variations. For this reason, we employ an iterative procedure where at each step we correct the maps with an estimate of the FF, fit and subtract gradients from the maps, and undo the initial FF estimate. By repeating this process until convergence in the gradient parameters (typically < 5 iterations), we debias the impact of ZL on the FF estimate. This has the caveat that any array-scale variations in the FF may be partially absorbed by the final gradient estimate, however in the limit where these are small the two corrections have roughly the same effect at leading order. More concretely, for a FF variation δFF,

$$\frac{I^{obs}}{1+\delta FF} \approx I^{obs}(1-\delta FF), \tag{3.6}$$

i.e., $I^{obs}\delta FF$ can be approximated by a gradient.

While inspecting the slope fits we identified a quadrant-specific electrical effect in the science field integrations. In particular for one quadrant of the 1.8 μm detector, we observe a "flickering" effect that coherently shifts the ADU values across frames. This resembles a form of two-state noise, potentially the result of a voltage biasing issue in the detectors or readout. We find that the observed effect is coherent across pixels within each separate quadrant, which leads to different DC offsets after performing the slope fits. This then sources spurious anisotropic power on large scales and can bias array-scale gradient estimates. As a result we choose to pre-process each quadrant separately. We show in App. .2 that the per-quadrant treatment improves the fidelity of our read noise model, since the effect is present in dark integrations as well.

In Figure 3.4 we show an example of our pre-processing for the Boötes B field at 1.1 μm. Our filtering procedure removes the majority of fluctuations sourced by the FF variation and array-scale sky gradients, however FF errors imprint additional signal and noise fluctuations on the processed CIBER maps.

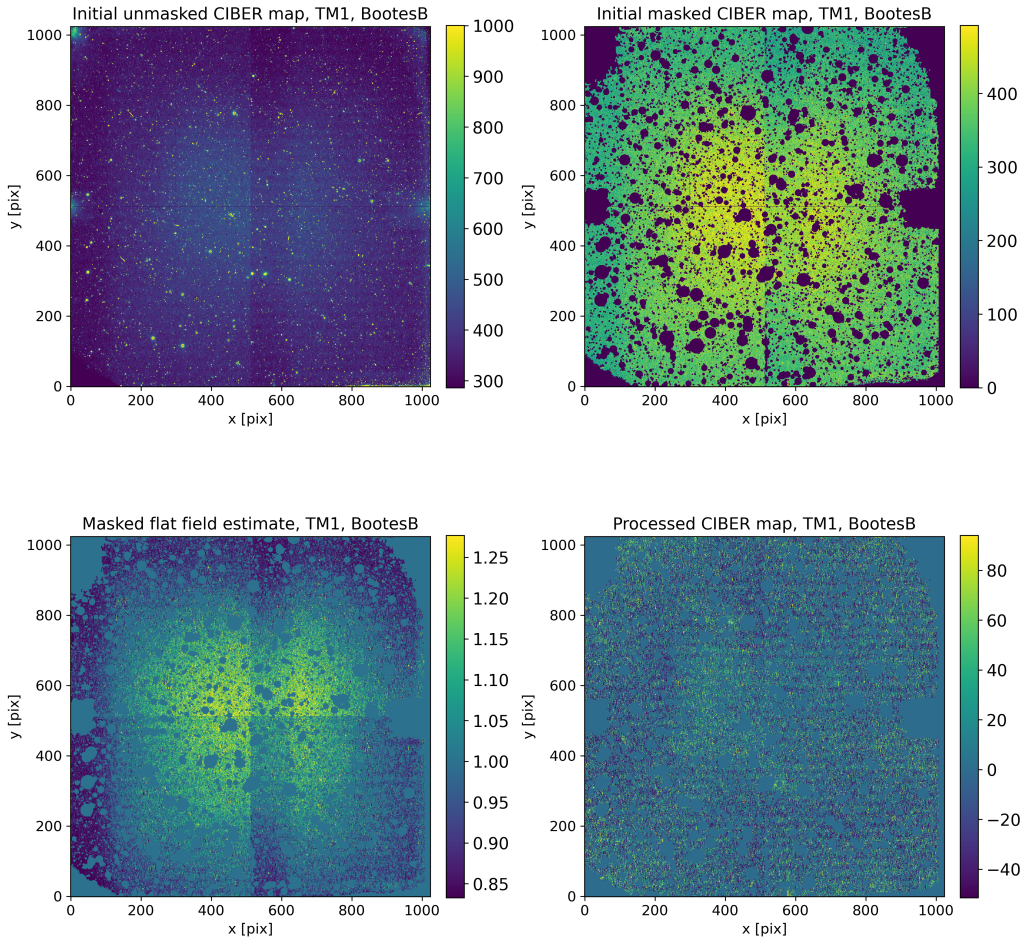

Figure 3.4: Example of data processing for *CIBER* TM1 (1.1 μm) observations of the Boötes B field. The top left panel shows the original unmasked image, while the top right is the same field after applying the instrument and astronomical masks. The instrument mask removes errant individual pixels and regions affected by multiplexer glow (seen in the corners of the unmasked maps and near $y = 500$). The bottom left panel shows the FF estimated from the other four fields. The bottom right panel shows the FF corrected map which is then mean-subtracted. The three *CIBER* maps are plotted in units of surface brightness (nW m^{-2} sr^{-1}).

3.4 Mask construction

Masking is an important part of our analysis, which is required to mitigate spurious fluctuation power from errant pixels and Poisson noise from bright stars and galaxies.

Astronomical mask

Each set of astronomical source masks is defined down to fixed depth in J- and H-band. We refer the reader to §6 of Paper I, for details regarding our source masking procedure. In brief, we use a combination of direct NIR photometry from 2MASS along with ancillary PanSTARRS/unWISE photometry that we use to predict NIR magnitudes through random forest regression. Through validation tests and comparisons with directly measured/model-based number counts, we determine that we are able to reliably mask over two magnitudes deeper than Z14, which relied on 2MASS photometry alone.

Stacking validation on CIBER images

We perform a final validation of our faint-end source masks by comparing the fluxes predicted by our masking catalogs to direct aperture photometric fluxes from the *CIBER* maps. We consider sources with $J \in [16, 18.5]$ and $H \in [15.5, 18.0]$. While individual sources in this range have low SNR ($\leq 3-5$, with variation across fields due to ZL and read noise), we stack sources in magnitude bins of width $\Delta m = 0.5$ to obtain reliable estimates of the mean flux within each bin. For each field, we omit sources within 50 pixels of the detector edge and any source with a neighbor brighter than $J = 20$ or $H = 19$. After these cuts we are left in each field with 100-1000 sources for each magnitude bin.

In Figure 3.5 we show the results of this test for both imagers. We perform an additional color correction to relate the observed *CIBER* fluxes at 1.05 and 1.79 μm to the predictions which are reported in UVISTA magnitudes (1.25 μm and 1.65 μm for J and H band, respectively). The mean fluxes from individual fields are consistent with one another, though elat30 has much noisier estimates due to the shorter integration time. For both bands we find generally close agreement between measurements and predictions, with the exception of bins $J \in [17.5, 18.0]$ and $J \in [18.0, 18.5]$ where our measured fluxes are ~ 0.2 mag brighter. Our measured fluxes have an additional uncertainty related to errors in our derived absolute calibration from §3.3.

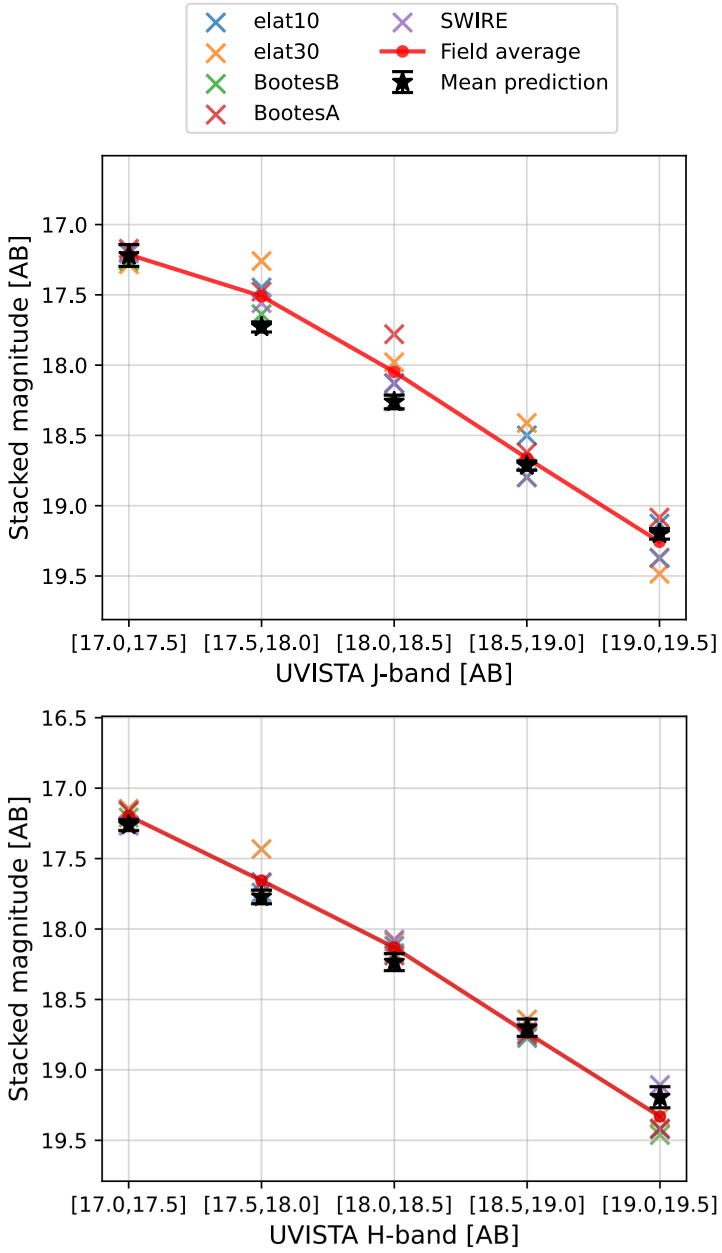

Figure 3.5: Comparison of stacked source aperture fluxes in magnitude bins (colored crosses) with mean predictions from our masking catalogs (black). We find close agreement between measurements and predictions across the range $m_{AB} = 17-19.5$, with the exception of $J = 17.5 - 18.5$ in which we measure slightly higher average fluxes.

SDSS Cluster catalog

We additionally mask galaxy clusters identified in [118] using SDSS photometric data. The number of clusters in each *CIBER* field is small, ranging from 32 to 52. We find that these have a negligible contribution to the observed *CIBER* fluctuation power.

Bright stars

We perform a search for the brightest stars in or near each field of view (FOV). The large fluxes of bright stars source intensity spread over many pixels due to the extended PSF. In addition, sources slightly outside of the detector FOV can scatter off of components in the focal plane, leading to reflected images onto the detectors. Within each FOV, we find an average of $\sim$ 3 J-band sources brighter than 7th magnitude and only one field (elat30) with two sources brighter than 5th magnitude.

There are a handful of bright stars that fall slightly outside the detector FOV but which could reflect onto the detector. These include:

- A magnitude $J = 3.6$, $H = 2.7$ star near elat10, which is $\sim 14'$ from the nearest detector edge.
- A magnitude $J = 5.0$, $H = 4.3$ star near elat30, $15'$ from the nearest detector edge.
- A $J = 5.2$, $H = 4.4$ star near Boötes B that is $8.8'$ from the detector edge, along with a $J = 5.5$, $H = 5.3$ star $7'$ from the detector edge.
- A $J = 5.8$, $H = 5.4$ star near Boötes A that is $7'$from the detector edge as well as a $J = 5.4$, $H = 4.8$ star $7'$ from the nearest detector edge.
- A $J = 5.0$, $H = 4.4$ star near SWIRE that is $3.5'$ from the detector edge, along with a $J = 5.3$, $H = 4.9$ source $9'$ from the detector edge.

We perform a systematic search for excess light that may be associated with these bright sources. In one corner of Boötes B we identify additional signal, with diffuse structure reminiscent of the "dragon's breath" effect seen in H2RG detectors, as observed by JWST[2]. We do not see the same diffuse structure in images from

[2] https://jwst-docs.stsci.edu/jwst-near-infrared-camera/nircam-features-and-caveats/nircam-dragon-s-breath

DECaLS, WISE, or 2MASS images, however we do find the color of the excess light is similar to that of the nearby red star with $J = 5.2$ and $H = 4.4$, which resides at the same detector latitude but ~ 75 pixels (1.4 mm assuming pixel pitch of 18 μm) from the detector edge.

Instrument and FF mask

The instrument mask flags pixels that are dead, have unusually high photocurrent, or reside near detector edge effects (e.g., multiplexer glow). Some pixels are identified by computing outliers from dark exposure differences. The instrument mask comprises $\sim 10\%$ of pixels in both imagers.

An additional mask, unique to each field, is required for pixels with an undefined FF estimate (i.e., all "off-field" pixels that would contribute to the FF are already masked). This increases the total masking fraction by $\lesssim 0.5\%$, depending on the field. In a similar fashion we mask pixels with FF estimates that deviate by $> 3\sigma$ relative to the mean local FF estimate, so to avoid non-linear effects from large FF errors. This affects an additional $1 - 2\%$ of pixels in each field.

Lastly, in the observed data there are a handful of low-redshift, extended sources that we identify and mask.

3.5 Noise model

Our noise model is constructed from two components. The first is read noise from the detector and readout electronics, which we estimate from dark exposures. The second is photon noise due to the Poisson statistics of sky signals incident on the *CIBER* imagers. We use the derived models in this section to simulate noise in our mock observations (see Paper I) and to estimate noise biases and uncertainties on the observed auto- and cross-power spectra.

Noise model construction

To construct our read noise model, we use a series of twelve exposures taken while the rocket was on the launcher shortly before flight and follow a similar procedure to that employed in Z14. These exposures most closely match the electrical environment expected during flight, and were taken with the shutter closed, such that read noise is the dominant contribution. To obtain a read noise model for each science field, we perform time stream filtering and slope fits on the exposures matching the number of frames in each science field integration. We then compute the difference between dark exposure pairs and compute the two-dimensional power spectrum. In these

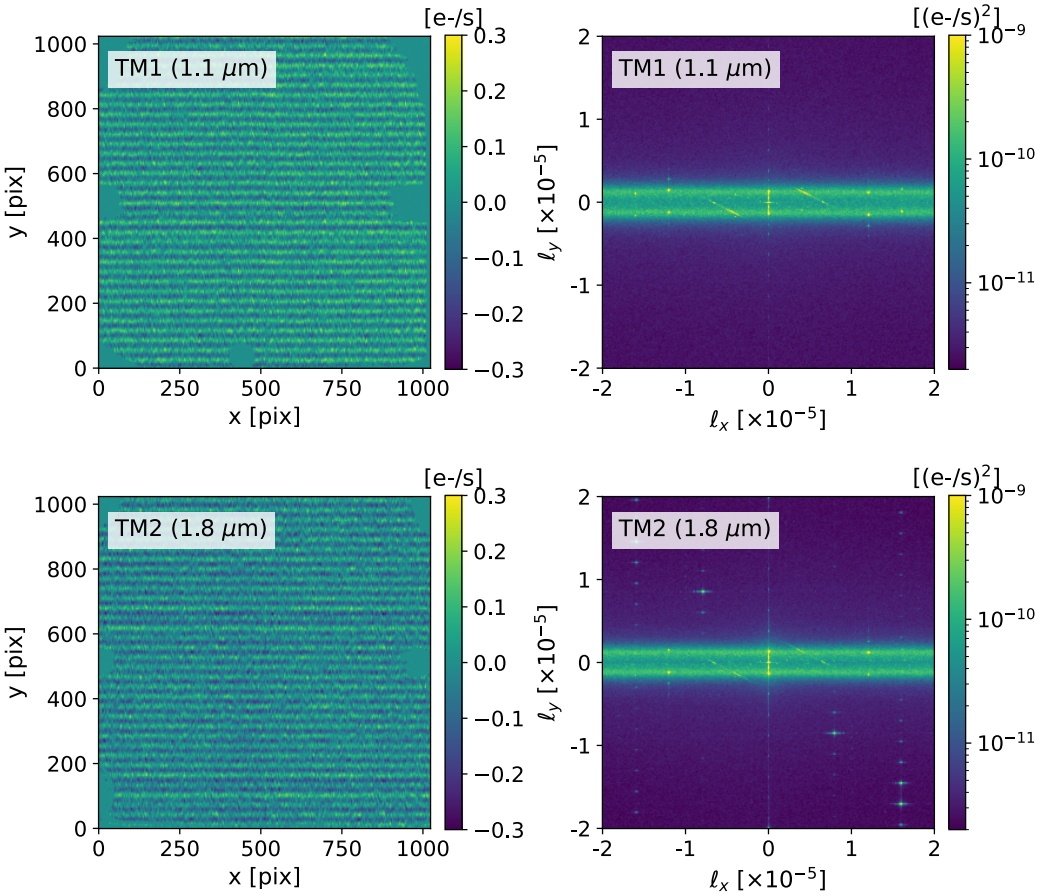

Figure 3.6: Individual dark exposure differences (left column) and two-dimensional power spectra (right column) averaged over pairs for 1.1 μm (top row) and 1.8 μm (bottom row). These match the integration time of science field elat10. Read noise is highly anisotropic in Fourier space, which we take advantage of through two-dimensional Fourier weighting of the observed power spectrum.

differences, any ambient non-zero signals should cancel out, while the noise power should double.

Figure 3.6 shows pairwise dark exposure differences and the derived read noise models for the elat10 field for both *CIBER* imagers. Due to the orientation of the readout amplifiers, there is anisotropic power in the two-dimensional power spectra along ℓ_y, peaking with specific (ℓ_x, ℓ_y) modes, which then mix through the *CIBER* masks. This motivates Fourier weighting of individual (ℓ_x, ℓ_y) modes before computing azimuthally-averaged bandpowers.

We infer the photon noise component of each *CIBER* field by converting the images to units of photocurrent using g_1 (derived in §3.3), computing the mean photocurrent

in unmasked pixels and relating this to the photon noise RMS as in [86],

$$\sigma_\gamma^2 = \frac{6}{5}\frac{i_{phot}}{T_{int}}\left(\frac{N^2+1}{N^2-1}\right), \tag{3.7}$$

where i_{phot} is the photocurrent, T_{int} is the integration time and N is the number of frames in the integration.

Validation tests

Read noise model consistency with dark data

For each field integration, we validate the read noise model by generating an ensemble of 500 Gaussian noise realizations drawn from the underlying mean two-dimensional power spectrum of the dark exposure differences and compute the ensemble of 1D power spectra. We then compare the 1D power spectra of the noise realizations against those from the initial set of dark differences in Fig. 3.7. Our model is able to reproduce the power spectra of the dark data, including sharp, anisotropic features such as the peak near $\ell \sim 6200$, which comes from residual noise imprinted by the readout amplifiers. The paucity of dark exposures taken on battery power limits our ability to estimate the validate the covariance of the read noise power spectra from our model. With only six exposure difference pairs, the uncertainty on the power spectrum covariance is $\sim 40\%$. Our simulated read noise realizations have variance consistent within at this level or better.

Consistency with flight data

Due to non-destructive readout of the Hawaii detector electronics, it is possible to construct flight half-exposures from subsets of the readout frames. In half-exposure differences, the power from coherent sky signal should cancel out while the noise power adds linearly. For each field we compute the power spectrum of the flight half-exposure differences and compare with similar differences derived from dark exposure integrations. We then apply the full mask associated with each field (detailed in previous section) to remove shot noise contributions from point sources in the flight images and compute masked one-dimensional power spectra. This bounds errors in the noise model at the level of half-exposure integrations.

This consistency test assumes that sky signal cancels out perfectly across half-exposures. However, this assumption is broken due to pointing drifts during the course of each science integration. We use the tool `astrometry.net` to compute separate astrometry for each set of flight half-exposures and differences as large as

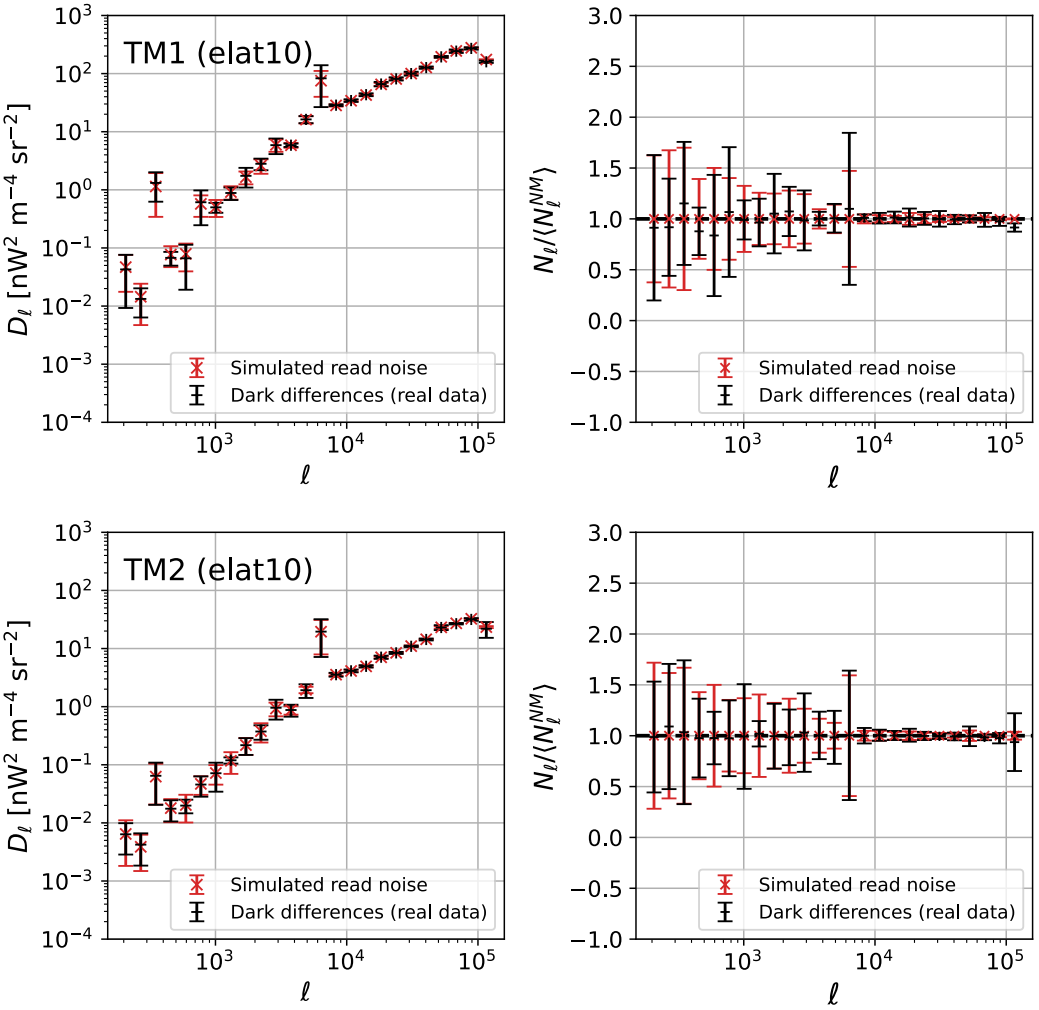

Figure 3.7: Validation of read noise models for *CIBER* imagers (1.1 μm detector in top row, 1.8 μm in bottom row). Left: comparison of the mean unweighted one-dimensional power spectrum derived from exposure differences taken before flight (black) and from the distribution of power spectra from read noise model realizations (red). Right: Dispersion for the two sets of power spectra, normalized to the mean power spectrum of the simulated noise.

10″ between pointing solutions. We attempted to mitigate small-scale power leakage due to pointing offsets by reprojecting the second half of each flight exposure onto the coordinate system of the first before taking differences. For *CIBER* images this reprojection is challenging because the instrument beam is undersampled (FWHM≈ 9″), introducing an additional transfer function which affects small scales. As such, we instead use our set of point source mocks to estimate the level of leakage power due to pointing offsets. We generate two versions of the *CIBER* point source mocks using the first- and second-half astrometry solutions, after which we mask each set of mocks, take their difference and compute the resulting power spectrum. This is then subtracted from the flight difference power spectra for each field.

We show the results of this consistency test in Figs. 3.8 and 3.9, comparing the mean and dispersion of noise difference realizations (black) to those from flight (colored points). On small scales ($\ell > 10^4$) the difference spectra are photon noise dominated and find agreement at the $10-15\%$ level. The estimated contribution from residual source structure is up to 20% and correcting for its effect reduces the discrepancy between the noise model and flight differences on small scales.

On intermediate scales ($10^3 < \ell < 10^4$) the difference spectra are dominated by read noise. On these scales our mean noise model predictions are up to 20% larger than that of the flight differences. One possibility is that pickup noise during flight was lower than during ground measurements. However, it is difficult to assess differences because the variance across noise model realizations on these scales is large and has correlated structure. We find mild inconsistency on large scales for the field elat10 in both imagers. This could be due to the presence of airglow contamination (elat10 is the first observed science field of the five we use).

For scales $\ell < 10^4$ we compute the χ^2 statistic between our noise model and the flight data, using a covariance matrix for each field derived from the ensemble of 1000 weighted noise realizations. We find deconvolving the noise spectra with the inverse mode coupling matrix significantly reduces cross-bandpower correlations, leading to better conditioned covariance matrices. Using these covariance matrices we calculate 1) χ^2 of the flight difference spectra relative to their noise models; and 2) the distribution of χ^2 values for each of the 1000 noise realizations for each field. These are both shown in the inset of each panel. We compute probability to exceed (PTE) values for each field using the empirical CDF of the flight χ^2 relative to the distribution from noise realizations.

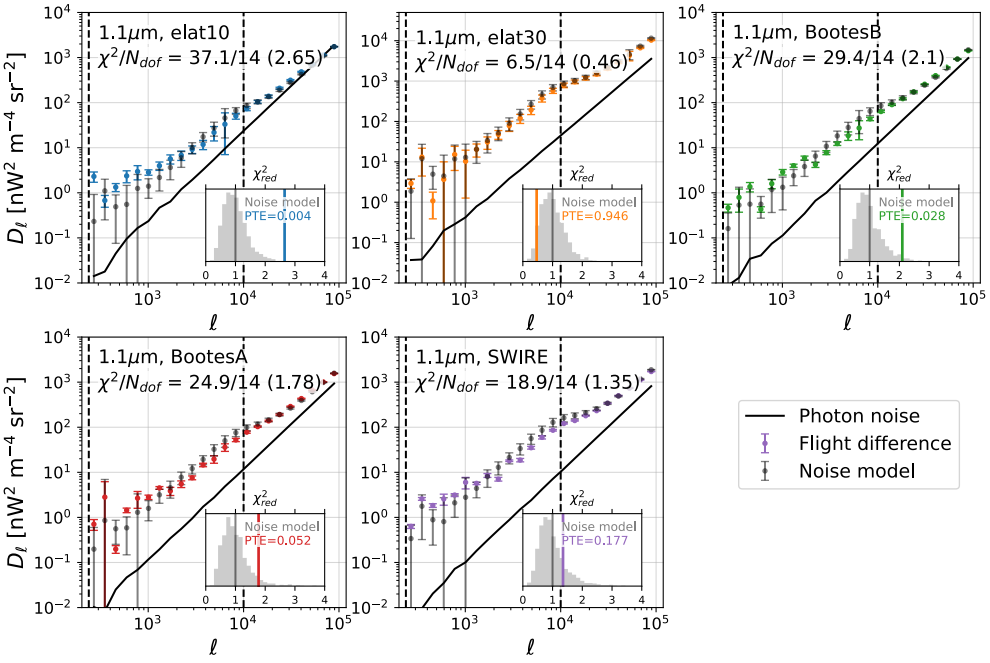

Figure 3.8: Noise model validation for 1.1 μm using flight and dark half-integration differences. Plotted are the one-dimensional angular power spectra derived from these differences, along with the photon noise level for each field indicated by solid lines. The dashed lines bracket the multipole range for which we evaluate deviations of the flight data from our model.

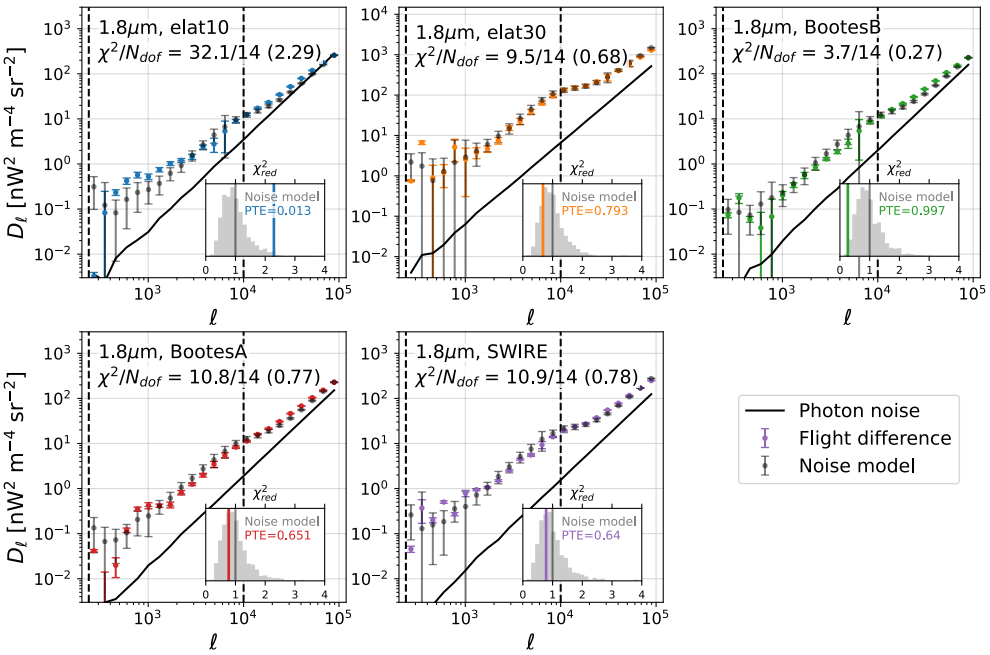

Figure 3.9: Same noise model validation but for 1.8 μm.

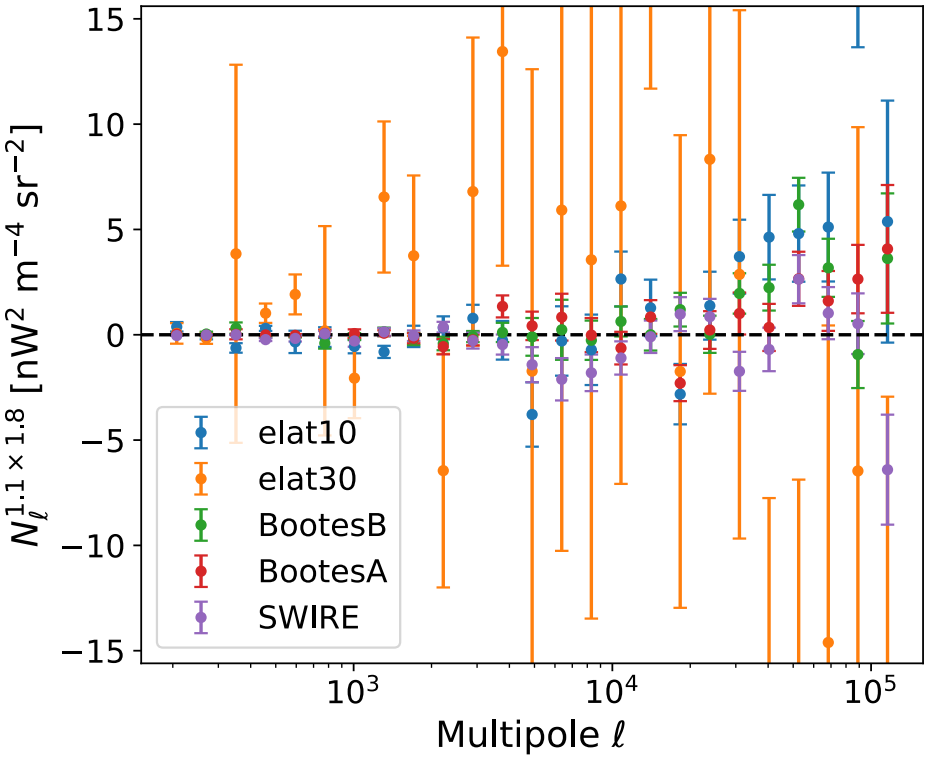

Figure 3.10: Cross-power spectra of *CIBER* 1.1 μm and 1.8 μm exposure half-differences. This constrains the presence of any time-variable fluctuations over the course of each integration which are correlated across bands. On large scales we find the cross-spectra to be consistent with zero.

Cross-correlation of half-exposure differences

If there are time-variable foreground components (on the time scale of individual exposures), these can contaminate the observed clustering in the full exposure maps. An example of such a component would be airglow from Earth's atmosphere. Our noise model validation tests using flight differences (§3.5) probe this to some degree, however we can also make use of the flight half-exposure differences cross-correlated across *CIBER* bands to check for coherent cross-power (within some range of signal colors). We calculate the cross power between 1.1 μm and 1.8 μm flight half-differences and show the results in Figure 3.10. For our five science fields, the cross-spectra are consistent with zero on all scales. Note that this test does not rule out all potential time-variable foregrounds.

3.6 Power spectrum estimation

In this Section we describe the steps used to relate an underlying sky power spectrum to bandpower estimates of the observed pseudo power spectrum. This is detailed in §4 and 5 from Paper I, however we summarize them briefly here. The sky signal with power spectrum C_ℓ^{sky} is related to the observed pseudo power spectrum C_ℓ^{obs} as

$$C_\ell^{obs} = \sum_{\ell'} M_{\ell\ell'} T_{\ell'} (B_{\ell'}^2 C_{\ell'}^{sky} + N_{\ell'}). \tag{3.8}$$

where $M_{\ell\ell'}$ is the mode coupling matrix, T_ℓ is the filtering transfer function, B_ℓ is the beam transfer function and N_ℓ is the noise bias. With estimates for each of these terms we can invert the equation above to solve for C_ℓ^{sky}:

$$\hat{C}_\ell^{sky} = T_\ell^{-1} B_\ell^{-2} \sum_{\ell'} M_{\ell\ell'}^{-1} (\hat{C}_{\ell'}^{obs} - \hat{N}_{\ell'}). \tag{3.9}$$

The following steps are required to estimate the underlying sky power spectrum:

1. Estimate and correct for the FF responsivity and large-scale gradients. In practice the two are estimated together using the iterative procedure described in §3.3 of Paper I. For this step we process each detector quadrant separately.

2. Estimate the noise bias (and 2D power spectrum noise weights) using an ensemble of read+photon noise realizations. We also include Monte Carlo estimates of the FF noise bias, which comes from instrument noise in the stacked FF estimates coupled to the mean sky brightnesses of the target fields.

$$N_\ell = \sum_{\ell'} M_{\ell\ell'} (N_{\ell'}^{read} + N_{\ell'}^{\gamma} + N_{\ell'}^{\delta FF}). \tag{3.10}$$

   We estimate the noise bias using an ensemble of 1000 noise realizations.

3. Compute the initial masked 2D pseudo-power spectrum and apply Fourier noise weights to obtain bandpower estimates:

$$C_\ell = \frac{\sum_{(\ell_x,\ell_y)} w(\ell_x,\ell_y) M(\ell_x,\ell_y)}{\sum_{(\ell_x,\ell_y)} w(\ell_x,\ell_y)}, \tag{3.11}$$

   where $M(\ell_x,\ell_y)$ is the two-dimensional observed power spectrum.

4. Correct for the effects of mode coupling introduced by the masks and FF errors by computing mixing matrices $M_{\ell\ell'}$, which are obtained through Gaussian Monte Carlo realizations, and applying the inverse $M_{\ell\ell'}^{-1}$ to each pseudo-C_ℓ.

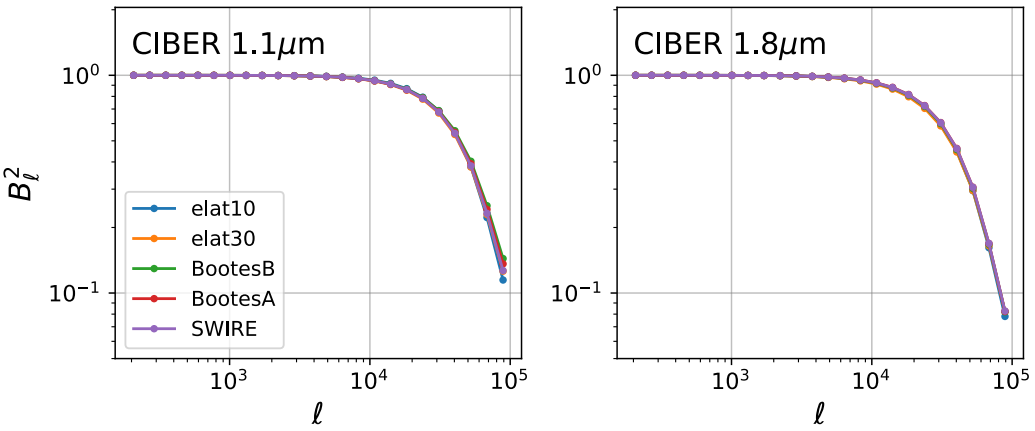

Figure 3.11: *CIBER* pixel-convolved beam functions for both imagers. These are derived from the best-fit PSF models obtained from stacking in [85].

As described in Paper I, it is necessary to calculate the mode coupling of FF errors from our stacking estimator through the science masks, which is done for each bandpower following the MASTER formalism.

5. Apply correction for the beam and filtering transfer functions. The *CIBER* PSF smoothes the observed sky signal, resulting in a roll-off in power on small scales, while the filtering transfer function primarily impacts large scales. These are shown for individual fields in Figure 3.11.

Mock power spectrum recovery

In Paper I we detail the synthetic mocks constructed for this analysis. The mocks are designed to match the realism of the observed data as closely as possible, including random realizations of galaxies and stars, DGL, ZL and EBL fluctuations to match the observed power spectrum from Z14. We note that our mocks use the observed mean surface brightness levels, rather than those from the Kelsall ZL model, in order to simulate representative photon noise. Testing our pipeline on one thousand sets of mocks, we validate our ability to measure unbiased estimates of sky fluctuation power down to masking depths of $J = 18.5$ and $H = 18.0$. We then use the dispersion of mock recovered power spectra to define per-bandpower weights that are used to combine power spectrum estimates from our five fields, which have varying levels of instrument noise, and to quantify statistical measurement uncertainties of the observed power spectra. The latter application is particularly important at low-ℓ, where estimating uncertainties from the observed set of modes in each bandpower

suffers from large sample variance. Lastly, we use the mocks to estimate power spectrum covariances that we utilize in §3.7 to assess internal consistency across the five *CIBER* fields.

3.7 CIBER clustering results

CIBER auto power spectrum measurements

Figure 3.12 shows the per-field *CIBER* auto power spectra at 1.1 μm and 1.8 μm, along with each weighted field average. We make predictions for the Poisson noise contribution to the *CIBER* power spectra that utilize the 2MASS and COSMOS catalogs described in Paper I. For frequencies A and B,

$$C_{\ell,A\times B}^{SN} = \mathcal{A}^{-1} \int_{I_\nu^{A,min}}^{I_\nu^{A,max}} (\nu_A I_\nu^A)(\nu_B I_\nu^B) \frac{dN}{dI_\nu^A} dI_\nu^A, \tag{3.12}$$

where $\mathcal{A}^{-1}$ normalizes the area of the measured counts (sr^{-1}) and dN/dI_ν is the differential number counts of the sample. This expression reduces to the auto spectrum shot noise for $A = B$. We use integrated 2MASS counts to compute the shot noise down to J and $H = 14$ along with COSMOS 2015 for $J \geq 15$. We then calculate color corrections to the shot noise between the catalog central wavelengths (~1.25 μm and 1.65 μm for 2MASS and UltraVISTA) and those of *CIBER*. This is done in magnitude bins of $\Delta m = 1.0$ to capture variation of source types and colors that correlates with a given magnitude limited sample. For example, many faint sources are galaxies with red spectra, while the bright end is dominated by stars with red and blue spectra at *CIBER* wavelengths. As shown in Fig. 3.12, our shot noise predictions at $J > 17.5$ and $H > 17.0$ underestimate the observed small-scale power by roughly a factor of two, which is coherent across the five science fields. We conduct several tests to consider different explanations for this in the following sections, however for the purpose of studying large-angle fluctuations the objective is to remove as much Poisson noise as is possible. Overall we are able to reduce the effective shot noise by a factor of $\sim 10^3$ between our shallowest and deepest masks.

There is a clear measurement of fluctuation power in both *CIBER* bands on scales $\ell < 5000$, exceeding expectations from Poisson noise. For scales $5' < \theta < 20'$, we detect non-zero fluctuation power at 12σ and 15σ for 1.1 and 1.8 μm, respectively. We also for the first time measure angular fluctuations on scales $\ell < 500$, which were unconstrained in [73] due to FF error and use of field differences. We also include updated estimates for the clustering contribution from DGL to the *CIBER* auto spectra. These are derived from cross-correlations from the CSFD extinction

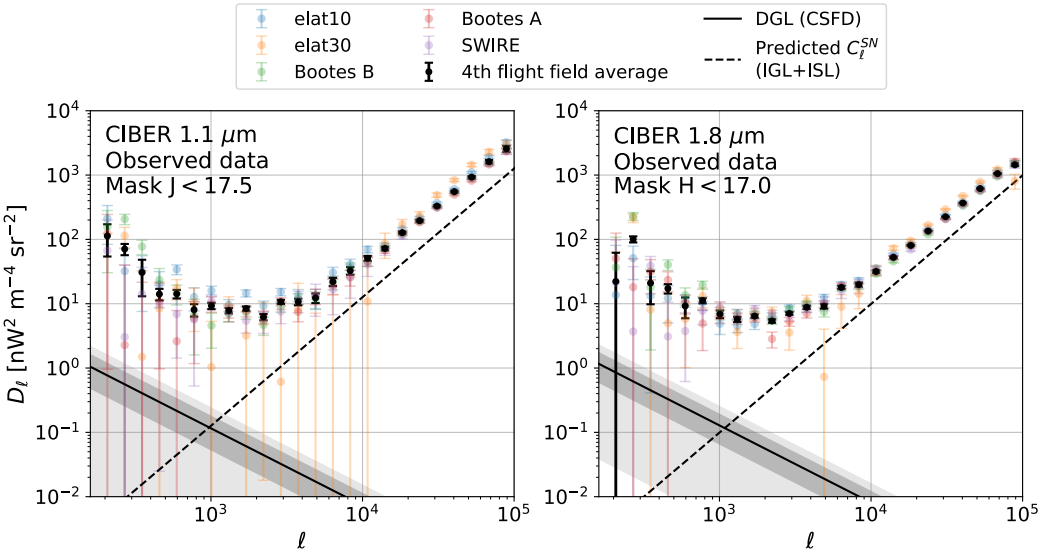

Figure 3.12: *CIBER* auto power spectrum measurements for imagers centered on 1.1 μm and 1.8 μm. The individual field power spectra are shown with colored points while the field averages (weighted within each bandpower) are shown in black. Also included are the predicted DGL contributions (black solid lines) with 1σ and 2σ uncertainties bounded by the shaded regions. IGL+ISL predictions using the COSMOS 2015 catalog (black dashed lines) are plotted for sources below each masking threshold. These are derived from cross-correlations with CSFD extinction maps (see §3.8). Our measurements show clear excess large-scale fluctuations beyond those expected from Poisson noise and DGL.

maps [119] which we present in §3.8.

Data consistency checks and systematic uncertainties

Field-field consistency

An important check for isotropy of the measured signal is testing that the observed fluctuations across different fields are consistent with one another. In Figures 3.13 and 3.14, we plot the deviations of the per-field 1.1 μm and 1.8 μm auto power spectra from their weighted field averages. We estimate the power spectrum covariance using from the ensemble of mock recovered power spectra. On small scales, the dispersion across fields is roughly ±15% and ±10% for 1.1 μm and 1.8 μm, respectively, with the exception of elat30 which has slightly larger departures – this field has the shortest integration of our five (only 10 frames, compared to 25-29 for the other fields), meaning it is most sensitive to any noise biases. The noise bias for elat30 is $\sim$ 10$\times$ the amplitude of the sky signal on intermediate to

small scales, meaning the recovered power spectra for this field are very sensitive to biases in our noise models. While we expect ISL variation across our fields, as well as sample variance in the IGL Poisson noise, the observed small-scale field consistency suggests these variations are small at our masking depth.

For bandpowers $\ell < 10000$, we compute a χ^2 statistic for each field i using the aforementioned mock covariance matrix:

$$\chi^2 = (C_{\ell,i}^{obs} - C_{\ell,av}^{obs})^T \hat{C}_{mock}^{-1} (C_{\ell,i}^{obs} - C_{\ell,av}^{obs}) \tag{3.13}$$

where the mock covariance is defined by

$$\hat{C}_{mock} = \langle (C_{\ell,i}^{j} - C_{\ell,av}^{j})^2 \rangle_j \tag{3.14}$$

averaged over realizations j. To compute χ^2 for individual fields, we use the block diagonal components of $\hat{C}_{mock}$. For 1.1 μm the reduced χ^2 of our fields range between 0.55 (SWIRE) and 2.6 (elat10), while for 1.8 μm they span 1.14 (SWIRE) and 2.8 (Bootes B). To obtain a probability-to-exceed (PTE) statistic for each field, we perform the same χ^2 calculation on the set of 1000 mocks and then compute the rank statistic of the observed data relative to the mocks. While we do not directly use the cross-field covariance in calculating χ^2, this is compensated by using the mocks which are treated consistently. Indeed we find that the mean χ^2_{red} of the mock ensemble is slightly below one (~ 0.9). Our PTE values span a relatively uniform range between 0 and 1, with the exception of elat10 at 1.1 μm ($p = 0.02$) and Bootes B at 1.8 μm ($p = 0.01$). These low PTE values are driven by departures on the largest scales ($\ell < 500$).

Sensitivity to masking function

If the observed signal is uncorrelated with the mask function then the deconvolved power spectra from different mask choices should be consistent with one another. In order to probe the correlation between the observed signals and masks, we perturb the faint end ($J > 14$) masking radii by ±25% for our fiducial masks $J < 17.5$ and $H < 17.0$. This changes the masking fractions by $\pm 2-5\%$ in both bands, depending on the field. In each case we recompute the mode mixing matrices for the five fields. Figure 3.15 shows the fractional deviation of the *CIBER* power spectra for different mask choices relative to the fiducial cases (denoted $C_\ell^{fiducial}$). For both bands, reducing the source mask radii has the effect of increasing the observed power on small scales by 5-10%, though the change on large scales is negligible. Increasing

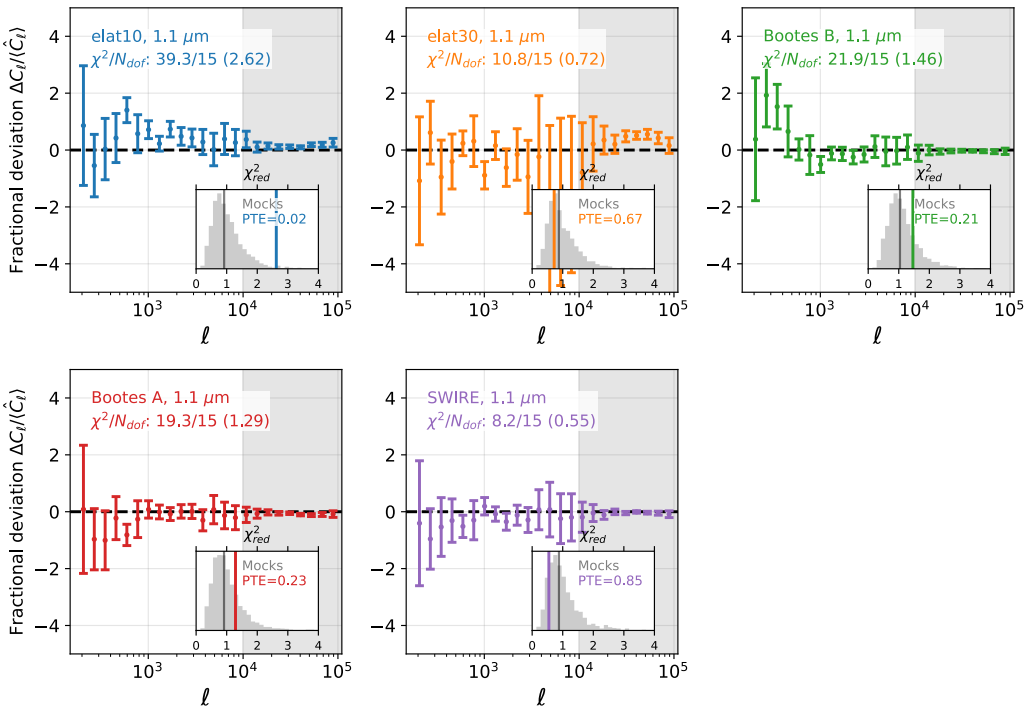

Figure 3.13: Fractional deviation of per-field *CIBER* 1.1 μm power spectra relative to the field average. We test field consistency on scales $\ell < 10000$ by computing the χ^2 statistic between each field and the mean, with bandpower covariances derived from an ensemble of recovered mock power spectra.

the masking radius lowers the observed 1.1 μm power spectra by a similar fraction, though for 1.8 μm there is no appreciable change.

Auto power spectra with varying source masking depth

To test the consistency of our measurements we compute auto power spectra for a range of masking depths and compare against IGL+ISL predictions. We plot these 1.1 μm and 1.8 μm auto spectra in Fig. 3.16. For the brightest sources, non-linearity in the detector response suppresses the measured flux of bright pixels. Non-linearity suppresses the flux by 10% for an integrated charge of 7.5×10^4 electrons, which corresponds to sources with Vega magnitudes $J, H = 11.5 - 12.0$ depending on the exact subpixel source position and assuming no pointing jitter[3].

The bandpowers in both bands decrease monotonically as a function of masking depth, with some exceptions on large scales that have large statistical variance. The

[3]Pointing jitter additionally suppresses the peak pixel integrated charge, shifting the threshold for non-linearity to slightly brighter magnitudes.

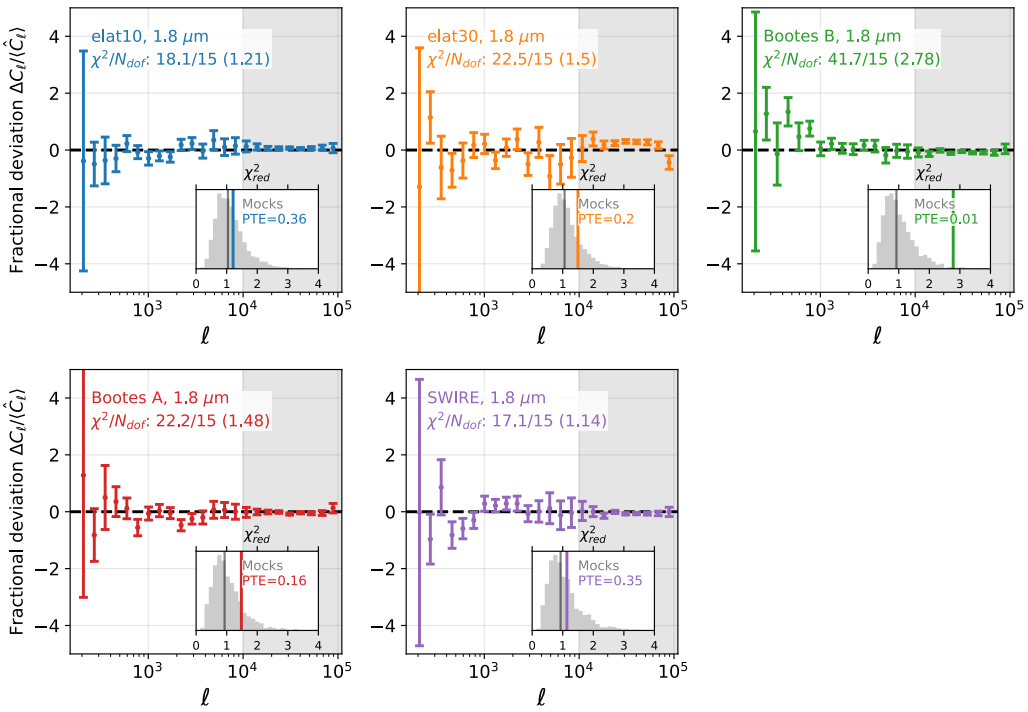

Figure 3.14: Same as Fig. 3.13 but for 1.8 μm.

observed power spectra have higher recovered small-scale power beyond $J = 16.0$ and $H = 16.0$, with departures from predictions increasing with mask depth. We consider several possibilities for this discrepancy:

- Incompleteness in our masking catalog could bias the observed power spectra high (at fixed depth). We validate our masking procedure down to $J = 18.5$ and $H = 18.0$ using the COSMOS 2015 catalog as a test set and determine that masking errors have a $< 10\%$ effect on the recovered shot noise relative to ground truth. This, along with consistency with mocks at the population level through number counts provide evidence that masking errors are not driving the observed small-scale discrepancy. Furthermore, our cross-spectrum measurements (next section) are much closer to shot noise expectations from IGL+ISL.

- Errors in the noise bias subtraction may bias the recovered PS amplitude. In §3.5, we find that noise in flight half-exposure differences is consistent with our noise model within the variance expected across noise realizations. Nonetheless, one possibility we explore is that errors in our read noise model on intermediate scales could bias the small-scale power spectrum through

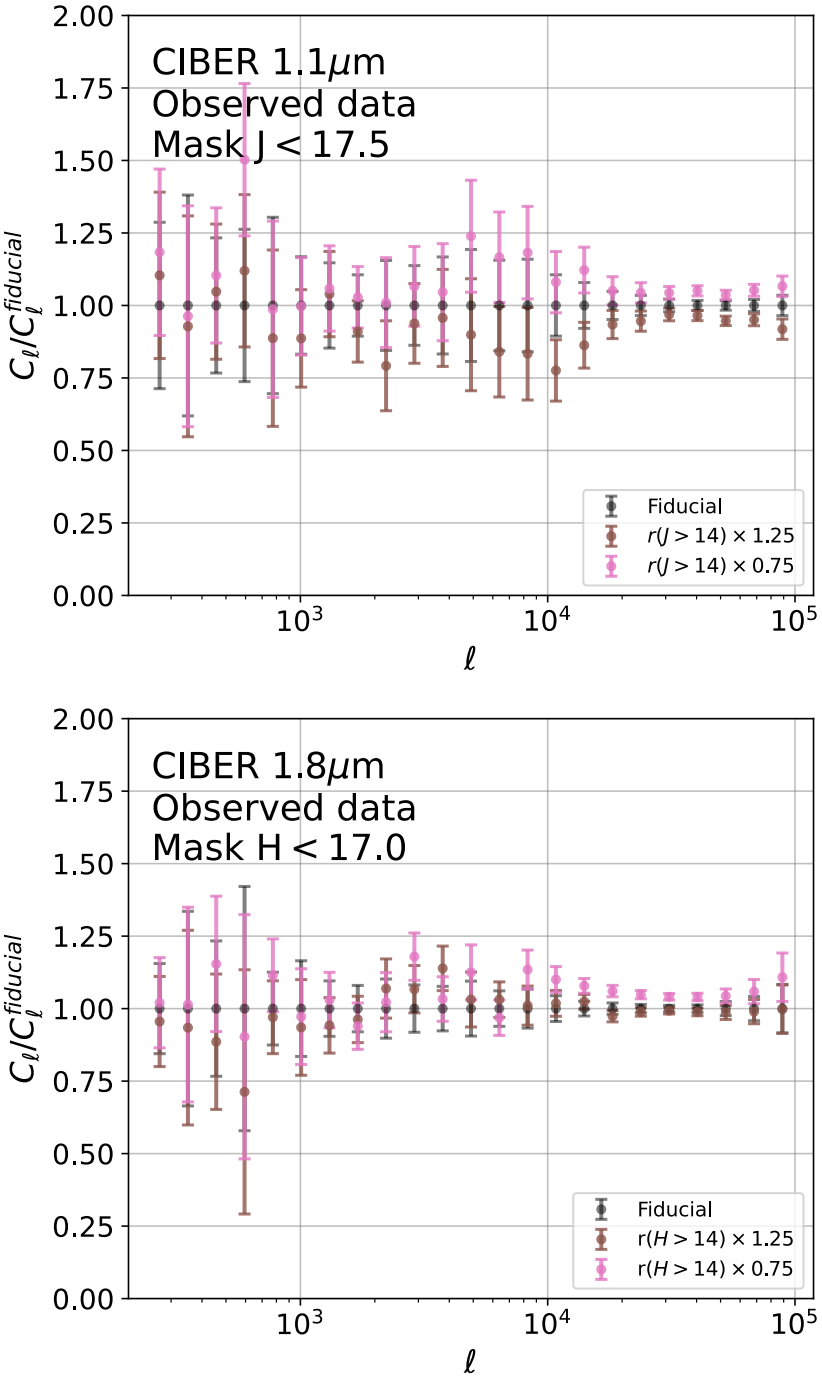

Figure 3.15: Sensitivity of *CIBER* observed power spectra to source masking, for 1.1 μm (left) and 1.8 μm (right). Perturbing the mask radius around each source by ±25% has a negligible effect on scales $\ell < 10000$ and changes the small-scale fluctuation power by less than 10%.

bandpower correlations. We test this empirically by scaling the read noise model by ±20% relative to the model used to generate our mock realizations and do not see evidence for changes in recovered power spectra.

- Additional fluctuation components with small-scale power could contribute to the auto spectra. Our cross correlation measurements in the next section suggest that these components would need to be uncorrelated across bands. One astrophysical scenario that could explain this would be the presence of strong emission line population. However given the relatively broad *CIBER* filters the implied equivalent width of the lines is large (> 1000 Angstroms).

- There may be errors in the Poisson noise predictions themselves. At the catalog level, errors may arise from incompleteness effects or flux mis-estimation (e.g., blending effects). The quality and depth of the COSMOS catalog make this explanation unlikely. If the properties of satellite galaxies differ from centrals in their stellar mass - halo mass relations, this may also affect Poisson noise model predictions.

Instrumental systematic uncertainties

Errors in beam correction

The *CIBER* power spectra are sensitive to beam errors as $1/(B_\ell^2 + \delta B_\ell^2)$, which are largest at on small scales. The PSF models we use from [85] are estimated for each separate field, however we do not account for PSF variations across each detector which have some impact on the small-scale power. There is some dispersion in the small-scale 1.1 μm beam transfer function across fields. The fact that the same dispersion is not seen in the TM2 beam functions suggests it is not due to pointing jitter, which would be coherent across imagers.

Sensitivity to gain errors

We test sensitivity of our power spectrum results to the assumed gain calibration by varying g_1, which converts the maps from units of ADU frame^{-1} to electrons per second. For a fixed absolute gain (ADU frame^{-1} to nW m^{-2} sr^{-1}), modifying g_1 has the effect of changing the relative contributions of photon noise and read noise, which affects the estimated noise bias. Figure 3.17 shows the *CIBER* observed power spectra for different assumed g_1 factors around our fiducial case ($g_1 = -2.7$

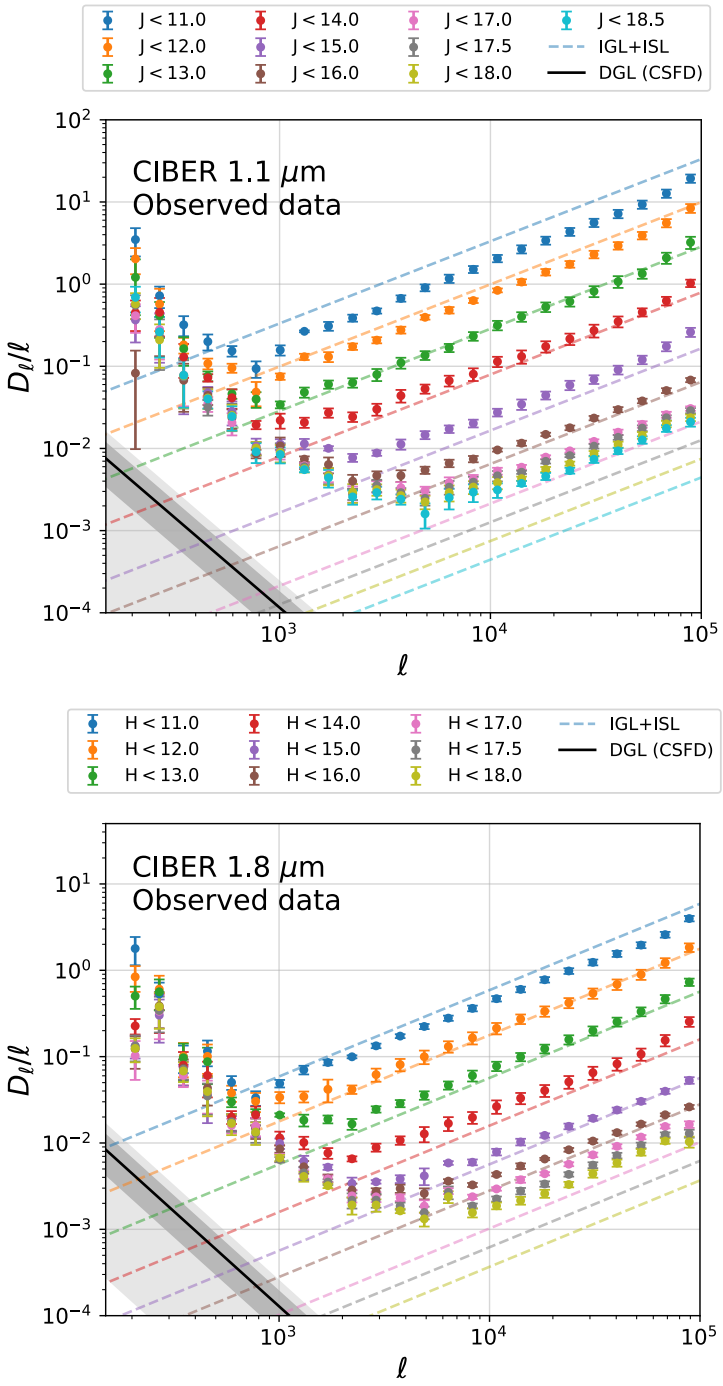

Figure 3.16: *CIBER* 1.1 μm (left) and 1.8 μm (right) power spectra at varying source masking depths, along with IGL+ISL model predictions from 2MASS and COSMOS (dashed lines) and DGL constraints (shaded regions). Note that this plot is for the quantity $\ell C_\ell/2\pi$ rather than D_ℓ in order to reduce the dynamic range between cases. The small-scale power spectra match shot noise predictions down to $J = 16$ and $H = 16$, beyond which we measure higher fluctuation power.

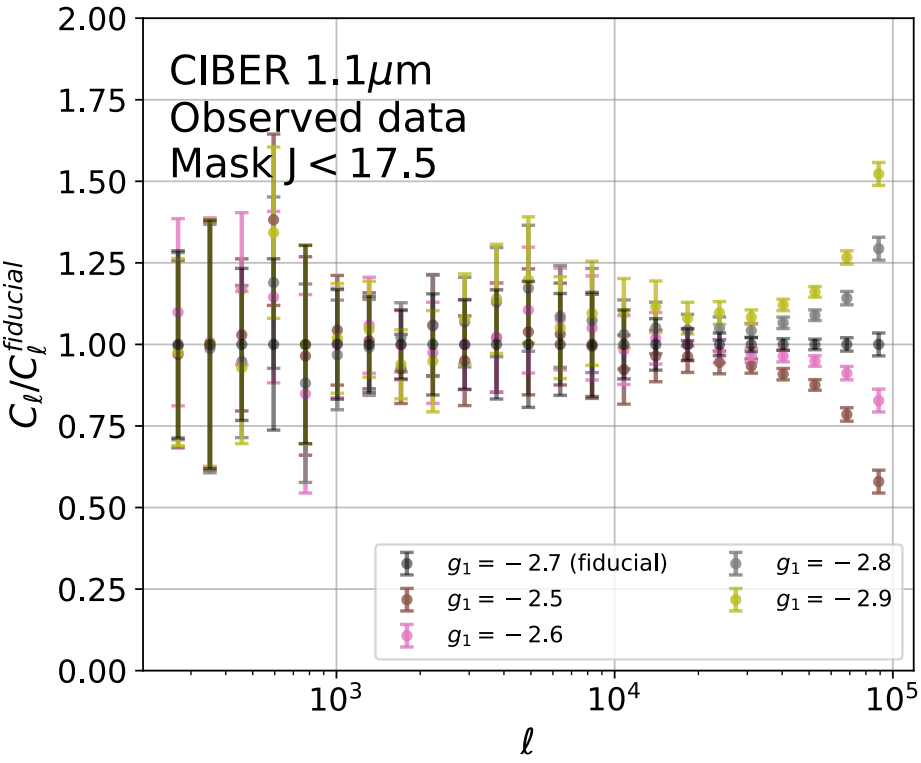

Figure 3.17: Sensitivity of *CIBER* observed power spectra to different assumed gain factors g_1, for 1.1 μm. This gain factor determines the ratio between read noise and photon noise, leading to changes in the noise bias. These couple with the mode mixing and beam correction leading to departures in power on small scales. However, the recovered power spectra on large scales are largely insensitive to changes in g_1.

for 1.1 μm, $g_1 = -3.0$ for 1.8 μm). Varying g_1 has the largest impact on the small scale power spectrum, which is sensitive to the noise bias from photon noise.

There is an additional overall uncertainty from our determination of the absolute calibration, $g_1 g_2$. Because the power scales as the square of the intensity the amplitude of the power spectrum is quadratically sensitive to errors in the absolute calibration, however this does not impact the shape of the power spectrum.

Scattered and reflected stray light

An important systematic for any analysis of diffuse light measurements is the presence of light reflected and/or scattered off the telescope optical elements and focal plane. The contribution from on-axis reflections has been quantified using simulations and is determined to be small (10^{-4} in fractional intensity). As presented in

[77] (see Fig. 12 of that work), the telescope off-axis gain function $g(\theta)$ with baffling is measured to be small ($g(\theta = 5^\circ)/g(0) \sim 4 \times 10^{-3}$ and $g(\theta \geq 15^\circ)/g(0) < 10^{-4}$). However, contributions from off-axis specular and diffuse reflections of bright sources that land within the focal plane are more difficult to quantify, as their structure depends on the geometry of focal plane elements and their reflectance, along with the astronomical scene of each pointing. We are unable to simulate this directly. However, given the independent focal planes and optics of the two *CIBER* bands, stray light across imagers is mitigated in cross-power spectrum measurements that we present in the next Section.

Image persistence

Image persistence describes the effect of remnant charge that remains in the detector following one or several exposures, and is typically explained by traps in the depletion regions of photodiodes within each pixel [120]. The amplitude of persistence decays exponentially with time and is determined to be at the sub-per cent level level for HgCdTe detectors. In place of constructing a full persistence model for the Hawaii-1 detectors, we check for effects of persistence in our power spectrum measurements by taking the brightest sources ($J < 10$) from each field and masking them in the subsequent exposures. The recovered power spectra with these additional persistence masks show no appreciable difference, suggesting the effect is negligible at *CIBER* sensitivity.

Comparison with Z14

Our closest point of comparison is the *CIBER* auto/cross spectrum results from [73] (Z14), which used imaging data from *CIBER*-1's second and third flights. To place the results of this work in context with Z14, we rescale the Z14 power spectra to the absolute gain derived in this analysis. To test the impact of analysis variations, we reprocess our power spectra in two ways:

- Restricting our masking catalog to sources detected in 2MASS only, consistent with that used in Z14. 2MASS has an integrated completeness of 75% down to $J = 17.5$ and $H = 17.0$, however the completeness falls off rapidly beyond $J = 16$ and $H = 15.0$.
- Applying the 2MASS only catalog and matching the masking radius function.

Z14 used the following parameterization:

$$r(m) = \alpha_m m + \beta_m, \tag{3.15}$$

where $\alpha_m = -6.25$ and $\beta_m = 110''$. This is a less aggressive masking prescription than used in our fiducial results.

We show the results of this comparison in Fig. 3.18. Our fiducial results are broadly consistent with Z14, however we do find some differences between results. In particular our clustering measurements are consistently lower by $30 - 50\%$ in both bands for scales $\ell > 1000$. Both measurements share large angle fluctuations which increase on scales $\ell < 1000$. Our results for $\ell < 1000$ are lower than Z14, however given the large uncertainties of the Z14 measurement the significance of the discrepancy is unclear.

Our analysis variations, which are intended to match the analysis of Z14, reduce some discrepancy between measurements. We find that retaining our masking radius prescription but restricting the masking catalog to 2MASS-identified sources increases the small-scale clustering power by 20% and 30% for 1.1 and 1.8 μm, respectively, but does not strongly affect clustering on scales $\ell < 5000$. When we additionally revert to the Z14 masking prescription for our astronomical masks, the clustering on scales $1000 < \ell < 5000$ is 50% higher on average than our fiducial measurements. This analysis variation also increases the small-scale clustering power by $\sim 10\%$ for 1.1 μm. These results highlight the sensitivity of our measurement to choices in source masking. The observed differences may be explained by extended PSF of bright stars in the fields, but also potentially by mask-signal correlations.

Despite these analysis variations there remains some discrepancy between measurements which is largest on scales $5000 < \ell < 20000$, scales for which both analyses are highly sensitive to read noise. Beyond what we have tested there are several choices that differ between Z14 and this analysis which may impact the measured clustering. These include, but are not limited to: the use of field differences in Z14 vs. our use of individual flight exposures; treatment of the FF correction and its errors and modeling of read noise variance used for noise bias subtraction. There is an additional color correction which we do not quantify due to the fact that the *CIBER H*-band filters differed slightly (1.6 μm to 1.8 μm) between the second and third flights.

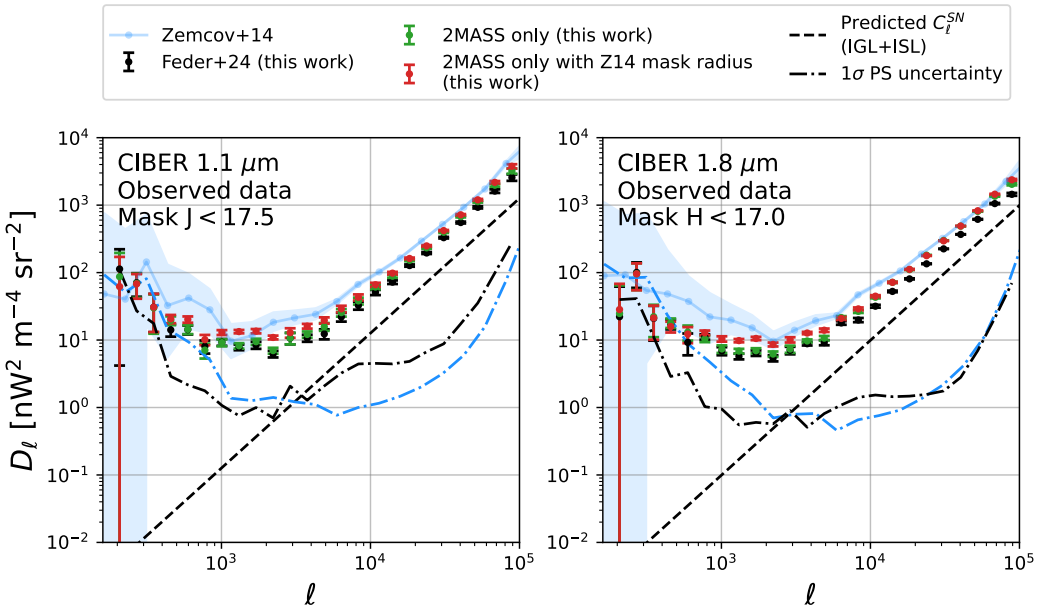

Figure 3.18: Comparison of fourth flight *CIBER* auto spectra (this work) with results from Z14 (blue). We show our fiducial results in black and those from two analysis variations: 1) restricting our masking catalog to 2MASS only (green) and 2) additionally reverting to the masking radius prescription from Z14 (red). These variations reduce some but not all of the discrepancy between measurements on scales $\ell > 1000$.

3.8 Consistency between wavelengths through cross spectra

Cross-power spectra isolate the common signal between imagers and/or instruments while mitigating the effect of noise biases unique to each measurement. For maps A and B, the cross-power spectrum variance is given by

$$(\delta C_\ell^{A\times B})^2 = \frac{1}{n_\ell}\left[(C_\ell^A + N_\ell^A)(C_\ell^B + N_\ell^B) + (C_\ell^{A\times B})^2\right], \tag{3.16}$$

where n_ℓ is the effective number of modes in the field of view considered. For each case, we estimate the noise contributions ($N_\ell^A(C_\ell^B + N_\ell^B)$ and $N_\ell^B(C_\ell^A + N_\ell^A)$) by generating noise realizations from one noise model, computing the cross power spectra against the other map realization, and calculating the dispersion as a function of multipole.

Correlation between CIBER bands

To compute *CIBER* 1.1 μm $\times$ 1.8 μm cross power spectra, we reproject the *CIBER* 1.8 μm maps and instrument mask to the 1.1 μm detector coordinate frame. The re-projection step occurs after the FF correction and gradient filtering in native

detector coordinates. This re-gridding imposes a transfer function on the 1.8 μm maps that primarily impacts the cross-spectra on small scales. We estimate the transfer function by computing the square root of masked pre- and post-reprojection 1.8 μm auto spectra. We find the measured transfer function is insensitive to source masking threshold. There is additional uncertainty in the re-projection due to relative astrometric errors between imagers, which we do not directly quantify in this work. We assume any aliasing effects due to the undersampled *CIBER* PSFs are captured by the re-gridding transfer function. The relative boresight accuracy between imagers is small, however there is a modest increase in the masking fraction due to the individual instrument masks. We mask on both J- and H-band magnitudes in order to probe the sub-threshold fluctuations common to both bands. There is a multiplicative FF bias related to the common sky signal across *CIBER* imagers (as explained in Paper I), which we correct for using a similar prescription to that of the auto spectra.

In Figure 3.19 we show cross-spectrum measurements at two masking depths with individual fields and field-averaged estimates. We find strong internal consistency across fields. Unlike the *CIBER* auto power spectra, the small-scale cross-spectra are much closer to the shot noise level predictions derived from COSMOS 2015, for all cuts down to our most aggressive masking depth ($J = 18.5$ and $H = 18.0$). Notably, the cross spectra exhibit similar excess large-angle fluctuations seen in the auto spectra.

Figure 3.20 shows the field averaged cross spectra for our full range of masking depths from $(J, H) < (12, 11.5)$ to $(J, H) < (18.5, 18.0)$. Similar to Fig. 3.16 the small-scale cross spectrum varies more rapidly with masking depth than on large scales, where the clustering appears to be largely independent of masking cut.

To quantify the coherence of surface brightness fluctuations across imagers, we calculate the power spectrum cross-correlation coefficient as a function of multipole. For maps A and B, the correlation coefficient r_ℓ is defined as

$$r_\ell = \frac{C_\ell^{A\times B}}{\sqrt{C_\ell^A C_\ell^B}}. \tag{3.17}$$

The variance on r_ℓ is then

$$\delta r_\ell^2 = \frac{1}{C_\ell^A C_\ell^B} \times \left[(\delta C_\ell^{A\times B})^2 + \left(\frac{C_\ell^{A\times B} \delta C_\ell^A}{2C_\ell^A} \right)^2 + \left(\frac{C_\ell^{A\times B} \delta C_\ell^B}{2C_\ell^B} \right)^2 \right]. \tag{3.18}$$

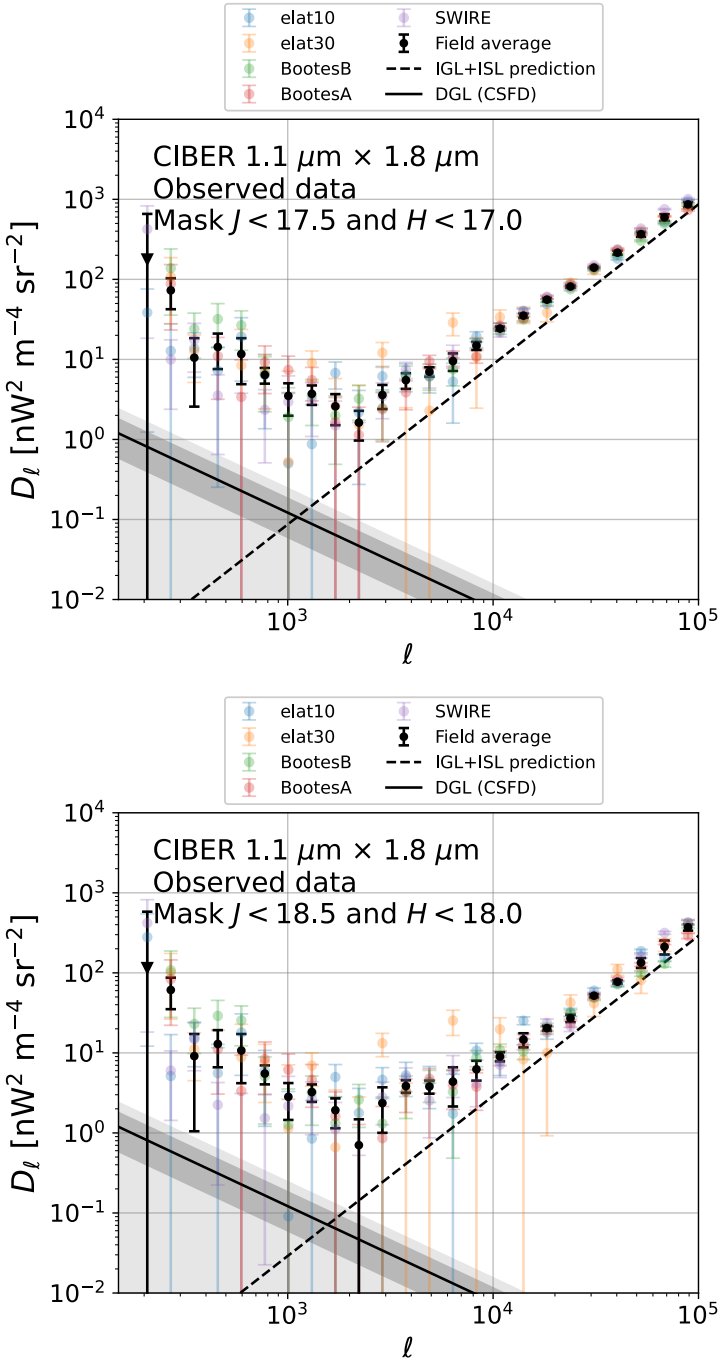

Figure 3.19: *CIBER* 1.1 × 1.8 μm cross-power spectrum measurements for two masking depths. For each panel, the individual field power spectra are shown with colored points while the weighted field average is shown in black. The dashed lines show cross-shot noise predictions obtained using the COSMOS 2015 catalog. DGL predictions on the cross spectrum (shaded regions) are obtained from the geometric mean of DGL 1.1 μm and 1.8 μm auto spectrum constraints. The triangle markers indicate negative bandpowers, for which we plot 95% upper limits.

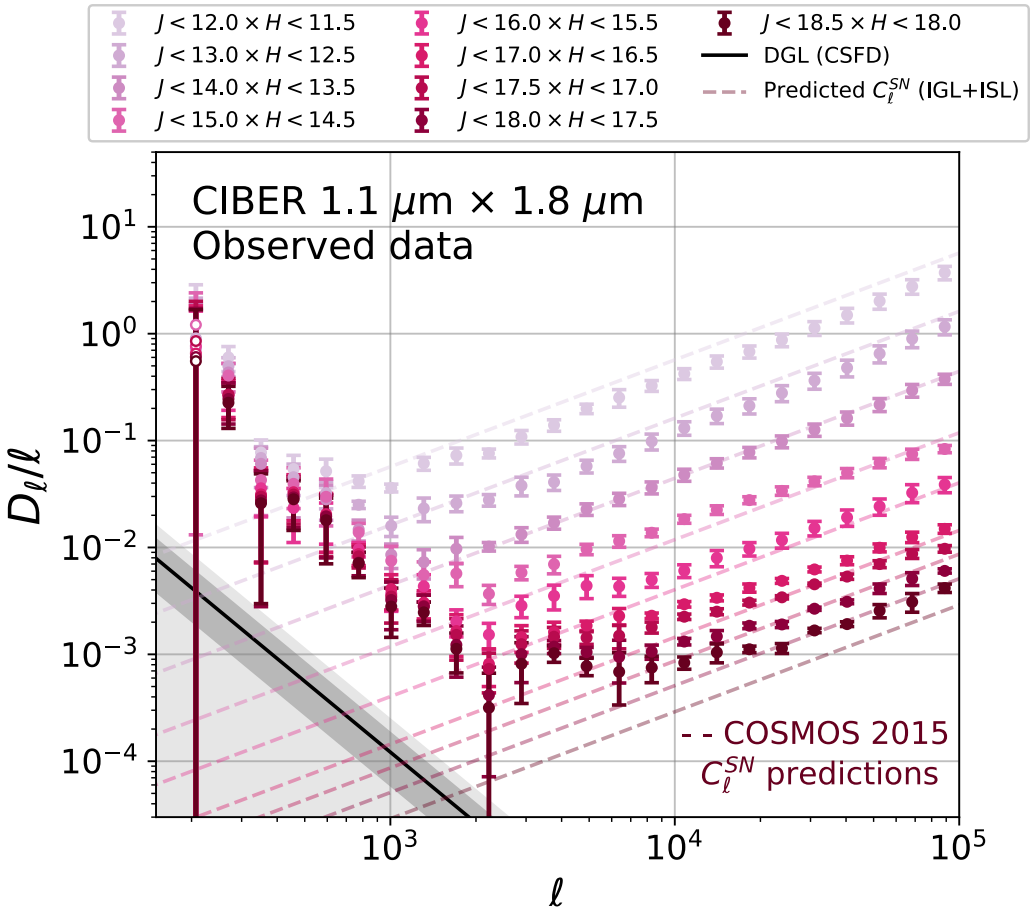

Figure 3.20: *CIBER* 1.1 μm × 1.8 μm cross-spectra for different source masking cuts indicated in the legend. Negative bandpowers are indicated by open circles. While the small-scale cross spectra roughly follow shot noise predictions derived from COSMOS (dashed lines), the large-angle clustering evolves more mildly, suggesting it is independent of the sources being masked.

We compute r_ℓ from auto and cross power spectra using the same masking cuts in Fig. 3.20. To highlight the scale dependence of r_ℓ we average into three broader bandpowers ($10^2 < \ell < 10^3$, $10^3 < \ell < 10^4$ and $10^4 < \ell < 10^5$). We plot these estimates as a function of masking magnitude in Figure 3.21. For $\ell < 1000$, r_ℓ is fairly insensitive to masking depth with $\langle r_\ell^{1.1\times1.8} \rangle = 0.67 \pm 0.14$ at fiducial masking depth ($J < 17.5$, $H < 17.0$). This is slightly lower but consistent with Z14, which reported $r_\ell = 0.76 \pm 0.10$. On small scales ($\ell > 10000$), $\langle r_\ell^{1.1\times1.8} \rangle \sim 0.87 \pm 0.05$ for point source dominated maps but decreases rapidly between $J = 16.0$ and $J = 18.5$ down to $\langle r_\ell^{1.1\times1.8} \rangle = 0.30 \pm 0.02$. This behavior is not seen in predictions from COSMOS and UDS and reflects the milder reduction in shot noise for observed

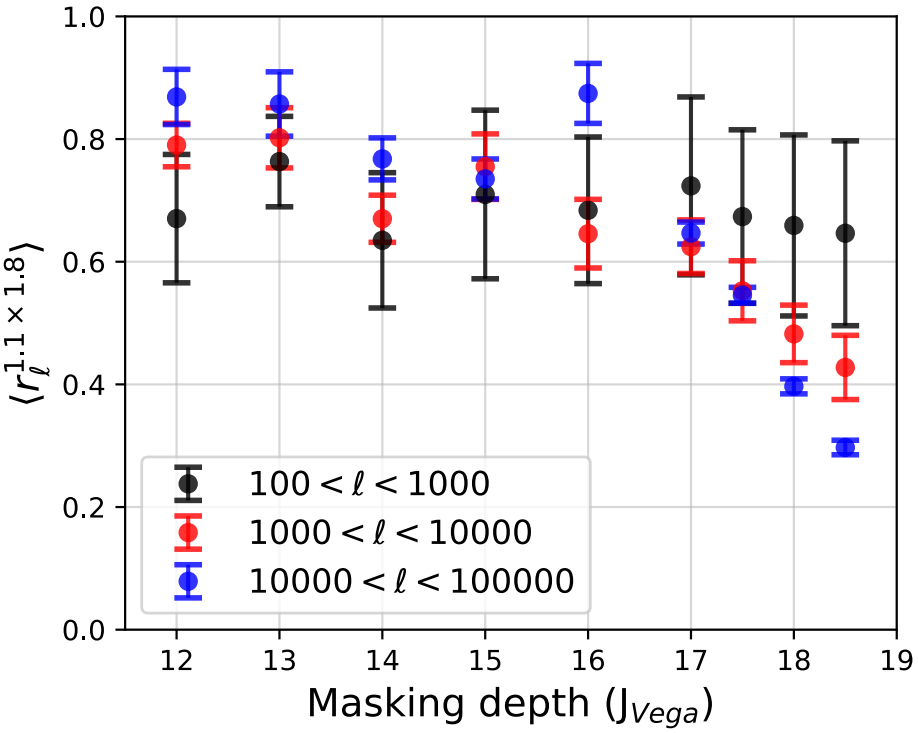

Figure 3.21: *CIBER* 1.1 μm $\times$ 1.8 μm cross-correlation coefficient in broad band-powers as a function of masking depth. While we find strong de-correlation on small scales as masking depth increases, the correlation coefficient on large scales ($\ell < 1000$) remains relatively unaffected.

auto power spectra for deeper masking depths. On intermediate scales ($1000 < \ell < 10000$), r_ℓ decreases beyond $J = 16$, albeit less rapidly than on small scales.

CIBER – DGL correlation

Having assessed the internal consistency of *CIBER* clustering and coherence across bands, we now estimate the contribution from diffuse galactic light (DGL) in the *CIBER* power spectrum. By performing cross-correlations with external tracer maps of DGL we seek to answer whether DGL can explain the large scale fluctuations measured by *CIBER*. This comes with the caveat that no single map is a perfect tracer of DGL. The DGL angular power spectrum falls steeply with multipole, with several studies finding $\gamma \sim -3$ [121, 122]. A joint analysis of HST and *CIBER* clustering data at 1.6 μm yielded a best-fit spatial index $\gamma = -3.05 \pm 0.07$ [123], however interpreting this as the spatial index of DGL assumes no other components to the large-angle *CIBER* surface brightness fluctuations.

We perform cross-correlations against two versions of the well-used SFD 100 μm

dust extinction maps. These include the original maps from [124] along with more recent "CIB-cleaned" SFD maps [CSFD 119]. Recent work has demonstrated the presence of extragalactic large scale structure in SFD and other dust maps [125]. Using cross-correlations with catalogs of known extragalactic sources, [119] isolates and subtracts the two-halo contribution in existing SFD maps, which is validated by stacking the original and corrected maps on the positions of known CIB galaxies. We cross-correlate against both versions of SFD in order to quantify the level of CIB contamination as seen by *CIBER*.

For each set of SFD maps, we apply our fiducial *CIBER* masks ($J < 17.5$ for 1.1 μm and $H < 17.0$ for 1.8 μm) and cross-correlate with both *CIBER* bands. After correcting for mask mode coupling we apply a beam correction given by the geometric mean of the beam functions from each map,

$$B_\ell^\times = \sqrt{B_\ell^{CIBER} B_\ell^{SFD}}. \tag{3.19}$$

Because the *CIBER* beam size is much smaller than the resolution of the SFD maps, the effective beam correction is $B_\ell^\times \approx B_\ell^{SFD}$. We model B_ℓ^{SFD} assuming a Gaussian beam profile with FWHM= 6.1 arcmin. We then estimate the noise associated with each cross-correlation following (3.16).

The resulting cross spectra are shown in Figure 3.22. While our measurements for individual fields are marginal, the field-averaged cross-spectra have total signal to noise ratios (SNRs) ranging between 4.8 and 11.8. The bulk of the total SNR comes from small scales. Our CSFD cross-spectra have lower SNR on average compared with SFD. We model the *CIBER* × SFD cross-spectra by fitting a power-law $C_\ell = A_\times \ell^\gamma$ for $\ell < 3000$. We measure separate spatial indices γ for SFD and CSFD based on fits to $6^\circ \times 6^\circ$ maps centered on each of the *CIBER* fields. Averaging over the five *CIBER* fields we find $\gamma_{SFD} = -2.7$ and $\gamma_{CSFD} = -3.2$. This is consistent with CSFD containing less CIB contamination, which leads to a shallower power spectrum in the uncorrected SFD maps. In each fit we fix γ to these values and float $A_\times$, the results of which are also shown in Fig. 3.22 with 1σ uncertainties overlaid. Table 3.3 summarizes the cross-spectra and power-law fits for each map combination.

In Figure 3.23 we plot our derived DGL color measurements alongside existing measurements in the literature spanning $0.4 - 4$ μm. Our measurements are broadly consistent with [126] and slightly higher than those from [127], in particular at 1.8 μm. For those studies, the mean intensity was correlated against SFD extinction

External map	λ_{CIBER} (μm)	Cross SNR ($\ell < 2500$)	$\delta I_\lambda^{CIBER}/\delta I_{100\mu\mathrm{m}}$
SFD	1.1	6.2	9.1 ± 1.5
($\gamma = -2.7$)	1.8	11.8	7.7 ± 0.6
CSFD	1.1	5.3	**8.2 ± 2.2**
($\gamma = -3.2$)	1.8	4.8	**8.6 ± 2.1**

Table 3.3: Results of power-law fits to *CIBER* × SFD power spectra, where γ indicates the best-fit spatial index for each external set of maps. The conversion factors $\delta I_\lambda^{CIBER}/\delta I_{100\mu\mathrm{m}}$ have units (nW m^{-2} sr^{-1})/(MJy sr^{-1}).

maps re-scaled to 100 μm. Comparing with our measurements assumes that the color of the fluctuation component correlates with the color of the mean intensity, which may or may not be the case. In addition, our fluctuation measurements may be subject to correlations with residual CIB beyond that removed in the CSFD extinction maps.

We then estimate the color of the cross power, which goes as the ratio of cross-spectrum amplitudes i.e., $I_{1.1}/I_{1.8} = A_\times^{1.1\times100}/A_\times^{1.8\times100}$. We measure colors $I_{1.1}/I_{1.8} = 1.5\pm0.4$ and 0.9 ± 0.4 from SFD and CSFD maps, respectively. These intensity ratios are lower than measurements from [127], which determined $I_{1.1}/I_{1.8} = 2.5\pm0.7$. Interpretation of colors from our cross-spectra is difficult due to modest cross-power SNR across the different map combinations, along with potential additional CIB contamination beyond that removed in CSFD with a different color from the DGL.

With estimates of $\delta I_\lambda^{CIBER}/\delta I_{100\mu m}$ from our cross spectra and measured SFD/CSFD auto spectra at 100 μm, we make predictions for the contribution of DGL to observed *CIBER* auto power spectra. These are shown in Fig. 3.24. On large scales, the DGL auto and cross power spectra can be expressed as

$$C_\ell^{SFD} = I_{100\mu m}^2 C_\ell^{DGL} \tag{3.20}$$

$$C_\ell^{CIBER\times SFD} = I_{\lambda_{CIBER}} I_{100\mu m} C_\ell^{DGL} \tag{3.21}$$

$$C_\ell^{\lambda_{CIBER}} = \left(\frac{I_{\lambda_{CIBER}}}{I_{100\mu m}}\right)^2 C_\ell^{SFD}. \tag{3.22}$$

Our DGL predictions are roughly consistent with [73], which marginally detected cross-power between *CIBER* and IRIS in one field difference. Uncertainties in the second and third flight auto spectra were too large to rule out DGL as the

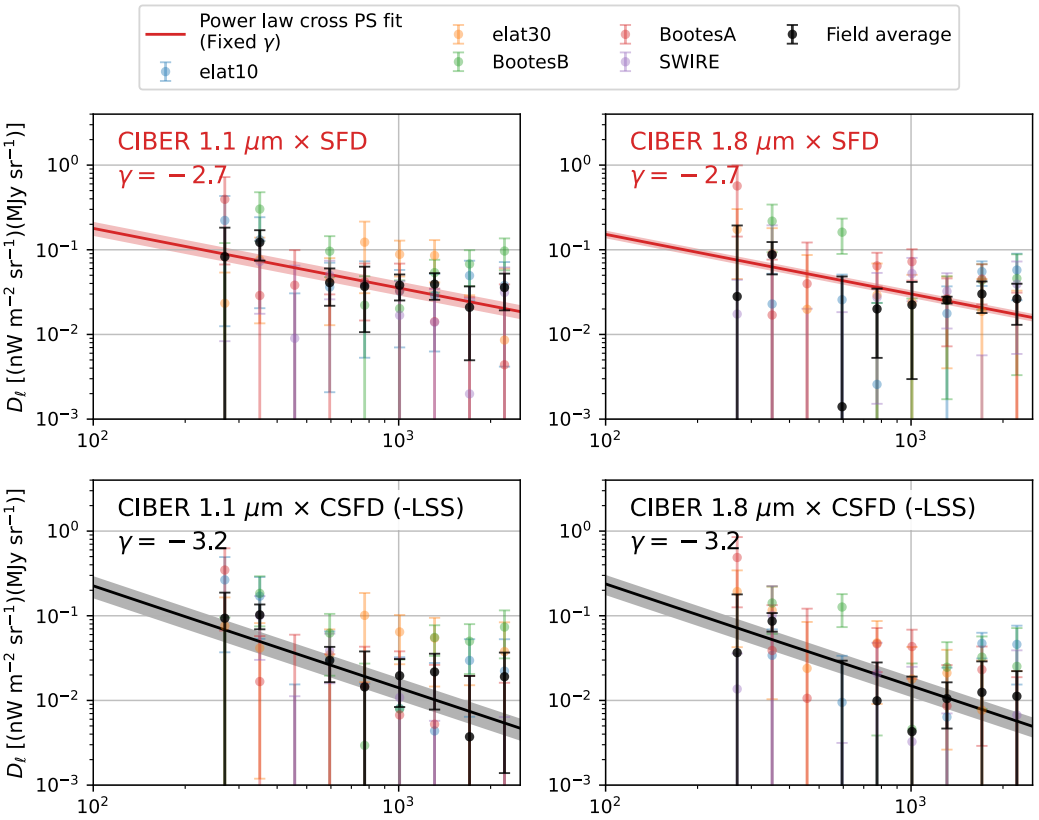

Figure 3.22: *CIBER* cross spectra with SFD extinction maps normalized to 100 μm (top row) and "corrected" SFD maps which have been cleaned of large scale structure (CSFD, bottom row). Individual field measurements are shown in colored points while weighted field averages are shown in black. The best-fit power spectra are calculated assuming a power-law form $C_\ell = A(\ell/\ell_{pivot})^\gamma$, for which the shaded regions span the $\pm 1\sigma$ uncertainties $\sigma(\hat{A})$. We fit the *CIBER* × SFD cross-spectra with free amplitude A but fixed $\gamma = -2.7$, while for the CSFD maps we fix $\gamma = -3.2$.

dominant component contributing to the observed large-scale fluctuations, however our analysis suggests that DGL does *not* explain the observed fluctuations at high significance. Indeed, our limits from CSFD are lower than the measured *CIBER* fluctuations by over an order of magnitude, motivating the presence of an additional sky component to explain the power.

CIBER – Spitzer correlation

For a subset of fields (Boötes A and B) we have access to IRAC infrared imaging. The *Spitzer* Deep, Wide-Field Survey [SDWFS; 128] covers the majority of the Boötes field footprint observed by *CIBER*. We use two versions of the SDWFS mosaics in order to assess consistency of our power sepctrum estimates. The first

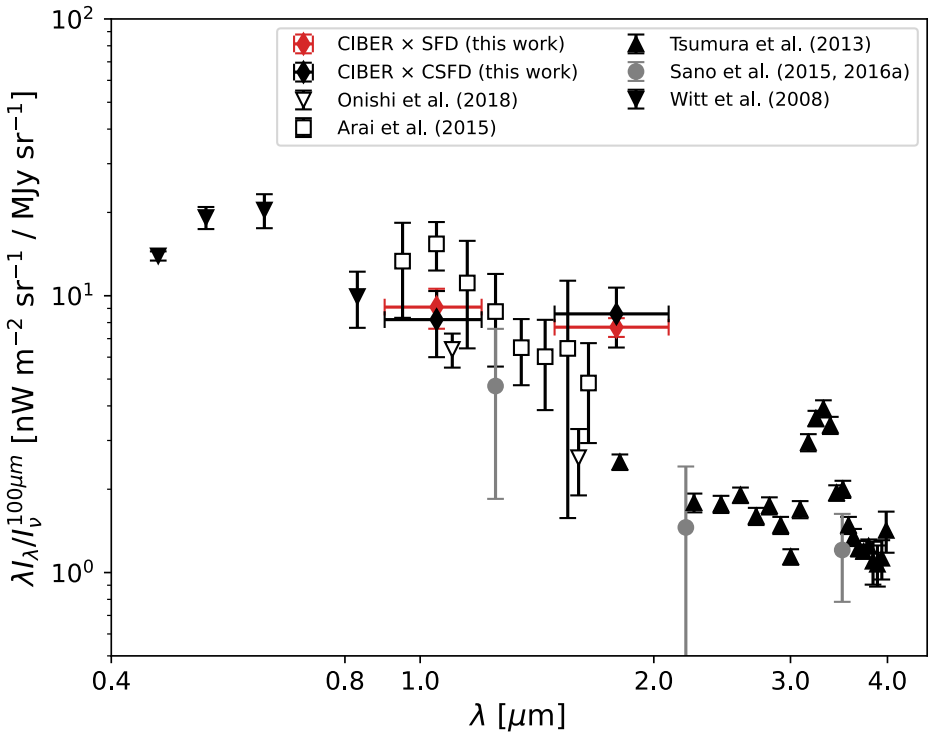

Figure 3.23: Comparison of νb_ν with existing measurements in the literature. Unlike most previous measurements presented, which come from correlations with the DGL mean intensity, our derived values (black and red diamonds) are derived from the cross-power spectrum.

are the official SDWFS mosaics which are constructed using a combination of `IRACProc` [an augmentation of the MOPEX algorithm 129] for 1 deg^2 sub-mosaics and `Montage` [130] for the final 10 deg^2 mosaic. The second version, which we use for our fiducial results (and are also used in [105] and `Z14`, were generated using the self-calibration algorithm of [131]. The co-addition algorithms are similar in spirit, however the self-calibration mosaics use dithered exposures to solve for the background and relative gains simultaneously. We utilize mosaics at 3.6 μm (IRAC Channel 1) and 4.5 μm (Channel 2).

Data reduction and noise estimation

We first reproject the SDWFS mosaics at each epoch onto *CIBER* detector coordinates, interpolating to *CIBER* angular resolution and converting from MJy sr^{-1} to nW m^{-2} sr^{-1}. The initial SDWFS mosaics have a pixel resolution of 0.86″ and PSF FWHM of 1.9″. To place the mosaics in the same surface brightness units as our *CIBER* results we apply a multiplicative normalization factor N_{eff} = 11.2

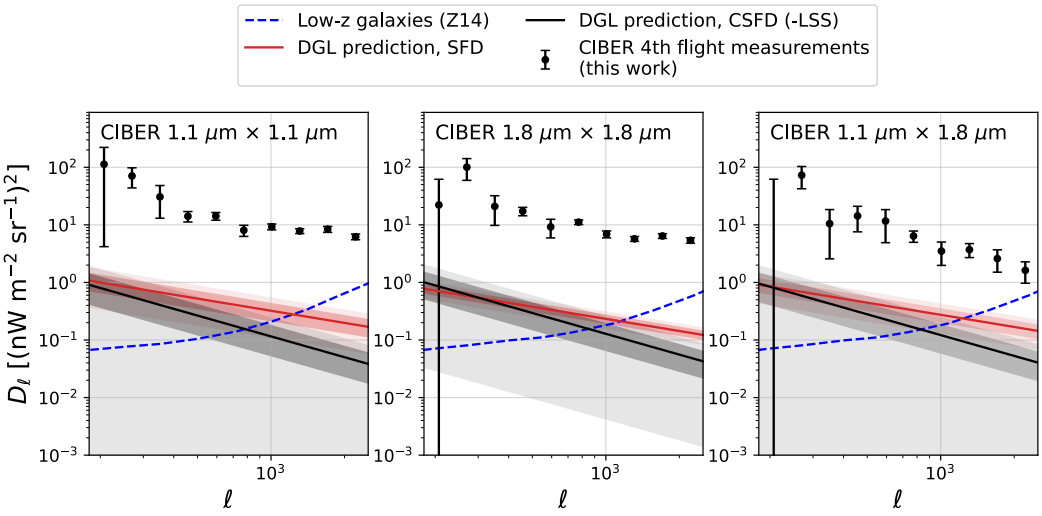

Figure 3.24: Constraints on the DGL contribution to the *CIBER* auto-spectrum, using fits to the SFD and CSFD cross spectra in combination with the tracer map auto spectra (black points), with 68% and 95% confidence regions overlaid. Both sets of constraints suggest the observed *CIBER* clustering far exceeds the predicted contribution from scattered diffuce galactic light (as traced by SFD).

equal to the PSF pixel effective area and assuming a Gaussian PSF. We validate the calibration of our interpolated *Spitzer* maps by comparing aperture photometry around sources with $13 < L < 16$ with the SDWFS catalog. Before averaging the four epoch maps we perform gradient filtering on each individual regridded epoch map. To reduce scatter in our cross-correlation measurements due to bright *Spitzer* sources, we mask sources with $L < 16$ ($m_{lim}^{3.6} = 18.8$ and $m_{lim}^{4.5} = 19.2$ AB mag) in addition to our fiducial $J < 17.5$ and $H < 17.0$ mask. This, combined with the union of the *CIBER* and SDWFS instrument/coverage masks, results in slightly higher masking fractions, the lowest being 45% for Boötes B. In Figure 3.25 we show the *CIBER* and *Spitzer* maps for Boötes A, smoothed by a Gaussian kernel with $\sigma = 5$ arcmin.

We then re-compute mode coupling matrices for each mask. The beam transfer function for the cross-spectrum is taken to be the geometric mean of the individual beam functions, i.e.,

$$B_\ell^\times = \sqrt{B_\ell^{CIBER} B_\ell^{IRAC}} \tag{3.23}$$

where we use the beam correction B_ℓ^{IRAC} determined in [105].

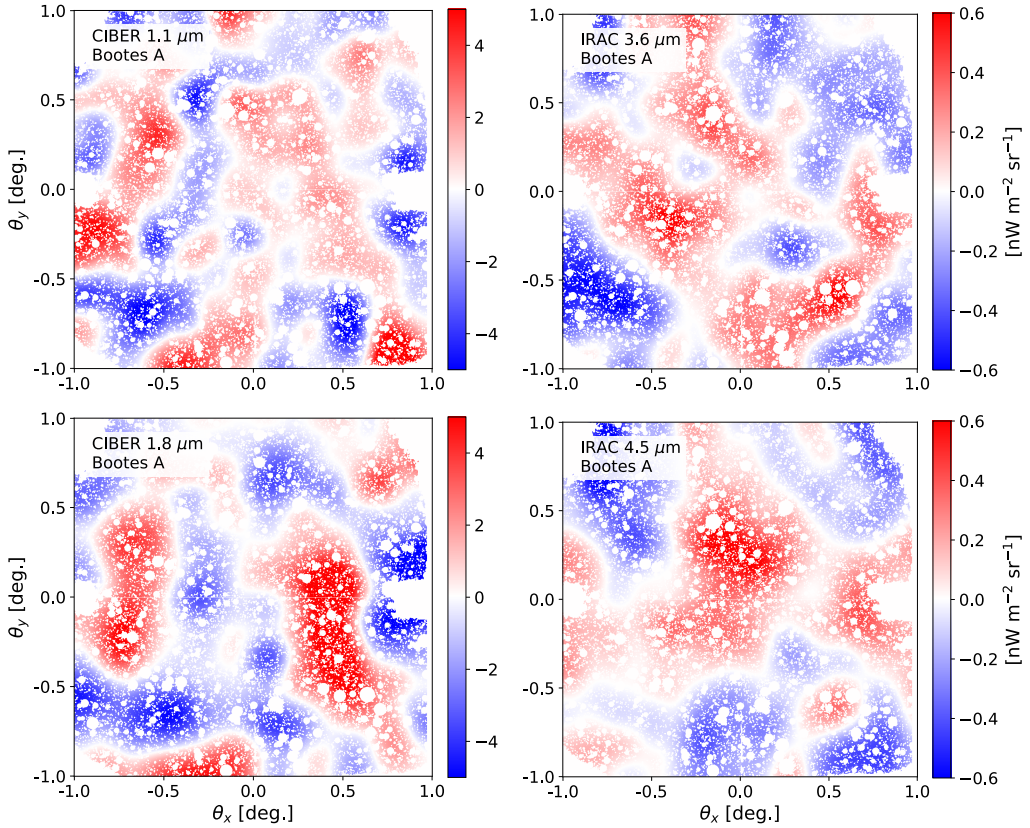

Figure 3.25: *CIBER* (left) and *Spitzer* (right) mean-subtracted maps for Boötes A. Each map is masked and then smoothed with a Gaussian kernel with $\sigma = 5'$. All maps are regridded to 1.1 μm detector coordinates.

Spitzer noise model

We follow the noise model prescription from Z14 which used the average 2D power spectrum of masked cross-epoch map differences, following the fact that any static sky signal should cancel out while noise power should add linearly. This approach has the benefit of capturing the direct noise statistics of the data, however there is additional variance due to detector effects, non-stationary noise across epochs and sample variance on large scales.

We rectify the 2D Fourier modes within each bandpower using the diagonal elements of the $M_{\ell\ell'}$ matrix calculated for each field. Assuming the SDWFS maps are photon noise-limited, we then scale the 2D noise power spectra to match the sensitivity of both the epoch pair averages used in the *Spitzer* auto spectrum and the four-epoch averaged maps that are cross-correlated with *CIBER*.

Spitzer auto- and cross-power spectra

To compute the *Spitzer* auto spectrum, we average two pairs of epochs and cross-correlate these against each other, i.e., $C_\ell^{IRAC} = 0.5(A+B) \times 0.5(C+D)$ for epochs $\{A, B, C, D\}$. This estimator of the auto spectrum is favorable because noise across the per-epoch maps is uncorrelated, meaning the expectation of the noise bias is zero (with the caveat of non-stationary noise). We estimate δC_ℓ^{IRAC} following Eq. (3.16) by generating Gaussian realizations from our *Spitzer* noise model and then cross-correlating them against each other and against the pair-averaged maps. We calculate *CIBER* × *Spitzer* cross spectrum uncertainties in a similar fashion, instead using using the *CIBER* and IRAC noise models. The *CIBER* noise model realizations include read noise, photon noise, and Monte Carlo estimates of the FF error specific to the Boötes fields.

We plot the resulting *Spitzer* auto power spectra and *CIBER* × *Spitzer* cross power spectra in Fig. 3.27. While the *Spitzer* auto spectra derived from the two Boötes fields are consistent with each other on scales $\ell > 1000$, we find evidence for systematics common to both fields on large scales. For example there is a sharp feature in the 3.6 μm auto spectra at $\ell = 450$ that is not expected from typical large-scale galaxy clustering. We find a larger dispersion between fields in the 4.5 μm auto spectra, with several negative bandpowers. The *CIBER* × *Spitzer* cross-spectra show strong consistency between fields for $\ell > 1000$. For $\ell < 1000$ the cross spectrum uncertainty is dominated by *Spitzer* noise (orange dashed curves), while the *CIBER* noise is dominant on small scales. For this reason our 1.8 μm cross-correlations have slightly lower noise than those with 1.1 μm. We note that our noise model is likely an incomplete estimate of the true systematic PS uncertainty, and while some bandpowers may be internally consistent they may be affected coherently by systematics in the *Spitzer* and *CIBER* maps.

A significant source of systematic uncertainty in this measurement is the mosaicing used to stitch together individual IRAC observations, which have an instantaneous field of view of $5' \times 5'$. We probe this by cross-correlating the per-pixel SDWFS exposure maps from each co-add against the IRAC mosaics themselves. This tests the assumption that the observed fluctuation power is independent of the survey footprint. We find that, while $|r_\ell| < 0.1$ for $\ell > 1000$, it grows to order unity for $\ell < 1000$, with some coherence between the two Boötes fields. This is shown in Fig. 3.26. It is beyond the scope of this work to characterize the mosaicing transfer function in detail, however this is possible by passing clustering realizations through

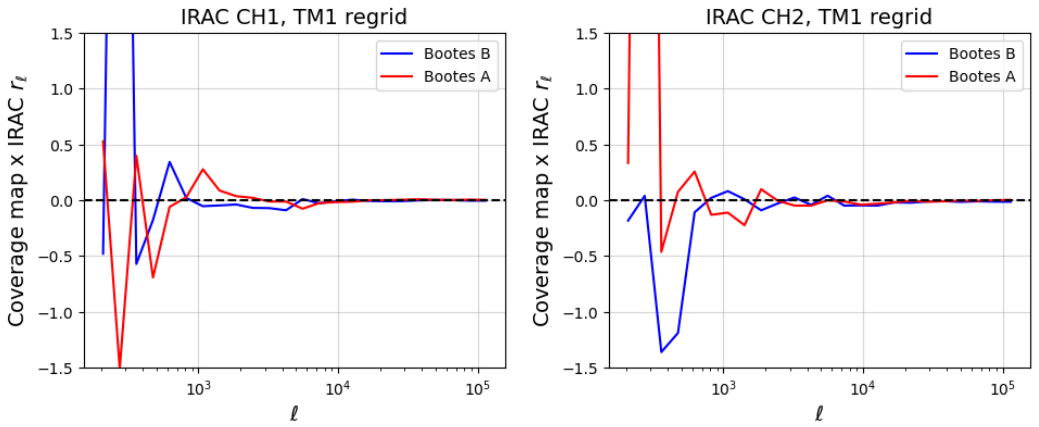

Figure 3.26: Cross-correlation coefficient between *Spitzer* self-cal mosaics and the per-pixel SDWFS exposure maps, for 3.6 μm (left) and 4.5 μm (right). The presence of large correlations and anti-correlations on scales $\ell < 1000$ is indicative of systetmatics in the combined mosaics that preclude reliable measurement of fluctuations on those scales.

the code used to make the mosaics.

Likewise, we compare our fiducial results with those using the official SDWFS mosaics (which use `IRACProc` and `Montage`) as a check of consistency. We show the relative auto- and cross-power spectrum estimates in Fig. .34 of App. .3. We find that results from both mosaicing algorithms begin to diverge significantly for $\ell < 1000$. Without more detailed tests, we are unable to make a judgement of whether one (or both) mosaicing codes drive the observed differences. However, these consistency tests lead us to conclude that measurements for $\ell < 1000$ are unreliable, and we do not use them for any downstream results.

With these caveats, in Figure 3.28 we plot the field-averaged *CIBER* and *Spitzer* auto- and cross-spectra against Poisson noise predictions. For this comparison we recompute the *CIBER* auto spectra using the same *CIBER* $\times$ *Spitzer* union masks used in this section. For $\ell > 1000$ we detect auto- and cross-power at high significance in all band combinations. Our *Spitzer* auto spectra are consistent with expectations from shot noise (as predicted using the COSMOS 2015 catalog). The *CIBER* $\times$ *Spitzer* cross spectra are consistent with shot noise expectations on small scales, however the 1.1 μm cross spectra have slightly lower measured power than predicted. It is notable that all three sets of power spectra exhibit departures from Poisson noise below $\ell = 5000$ ($\theta > 2$ arcmin).

We use the *CIBER* and *Spitzer* auto- and cross-spectra to estimate the cross-

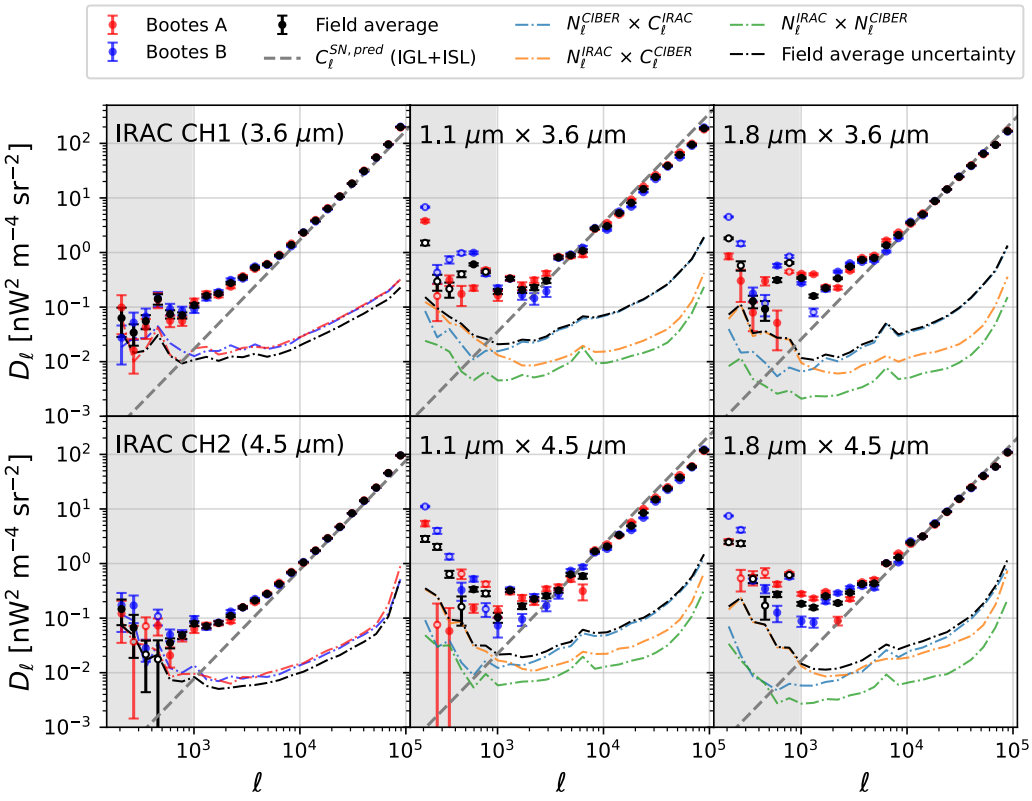

Figure 3.27: *Spitzer* auto-spectra (left column) and *Spitzer* × *CIBER* cross-spectra (middle, right). We show the recovered power spectra for the two individual Boötes fields in blue and red along with the corresponding field averages in black. In each panel, we include Poisson noise predictions derived from COSMOS 2015 (grey dashed lines) and the respective noise components contributing to the reported uncertainties (dash-dotted lines). For the *Spitzer* cross-epoch power spectra we show the total uncertainty of each Boötes field, while for *Spitzer* × *CIBER* we decompose the individual noise terms in (3.16). On large scales our cross-spectrum uncertainties are limited by *Spitzer*, while for $\ell \gtrsim 1000 - 2000$ the *CIBER* noise sources the dominant cross-spectrum uncertainty.

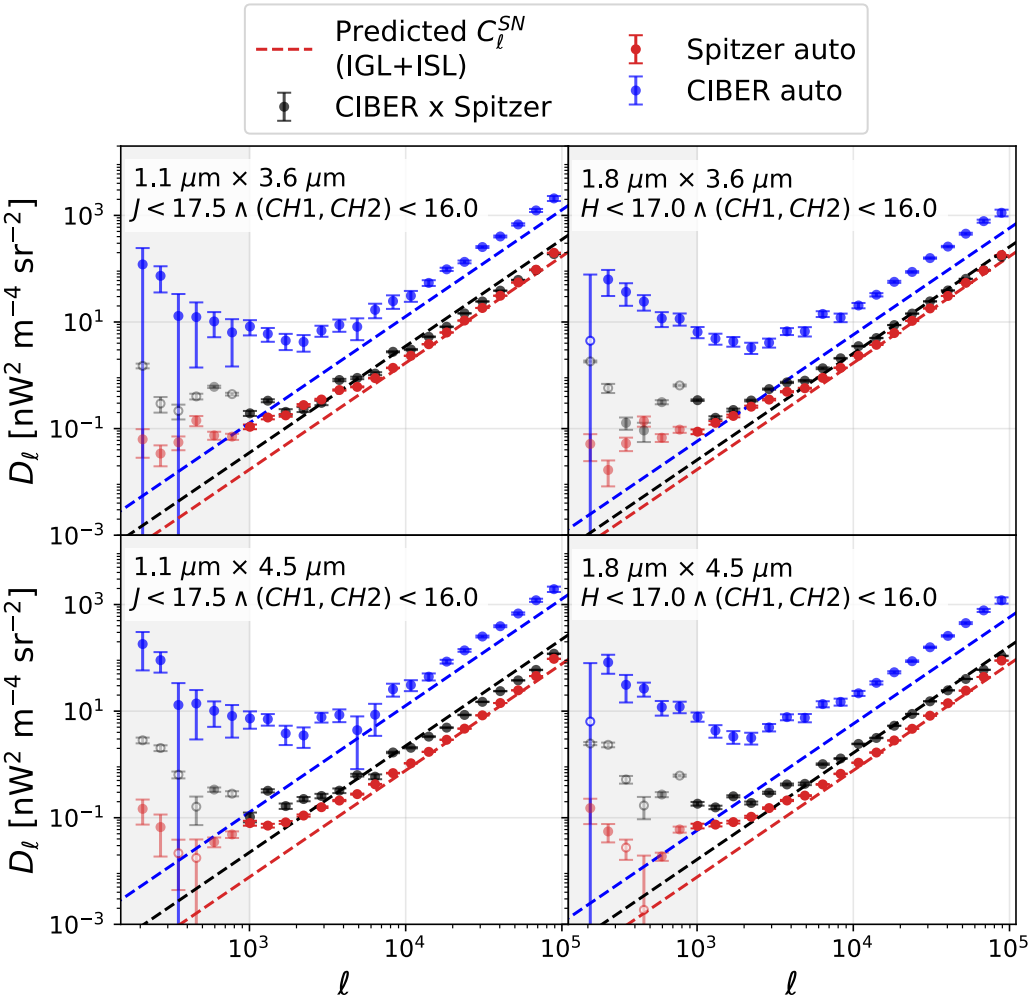

Figure 3.28: *CIBER* and *Spitzer* auto- and cross-power spectra in the Boötes fields (1.1 μm, 1.8 μm, 3.6 μm and 4.5 μm). For this comparison we recompute the *CIBER* auto spectra using the combined *CIBER* × *Spitzer* masks for the Boötes fields, including a mask for bright IRAC 3.6 μm and 4.5 μm sources. The *Spitzer* auto power spectra (red) are calculated through epoch cross-correlations. Our assessment of systematics in the SDWFS mosaics suggest *Spitzer* auto- and cross-spectrum measurements on scales $\ell < 1000$ are unreliable, indicated by the grey shaded regions.

correlation coefficient as a function of scale. When computing uncertainties on r_ℓ with Eq. 3.18, we use the same per-field *CIBER* auto spectrum uncertainties derived from mocks as in §3.7. We show these for the four *CIBER* × *Spitzer* band combinations in the Fig. 3.29.

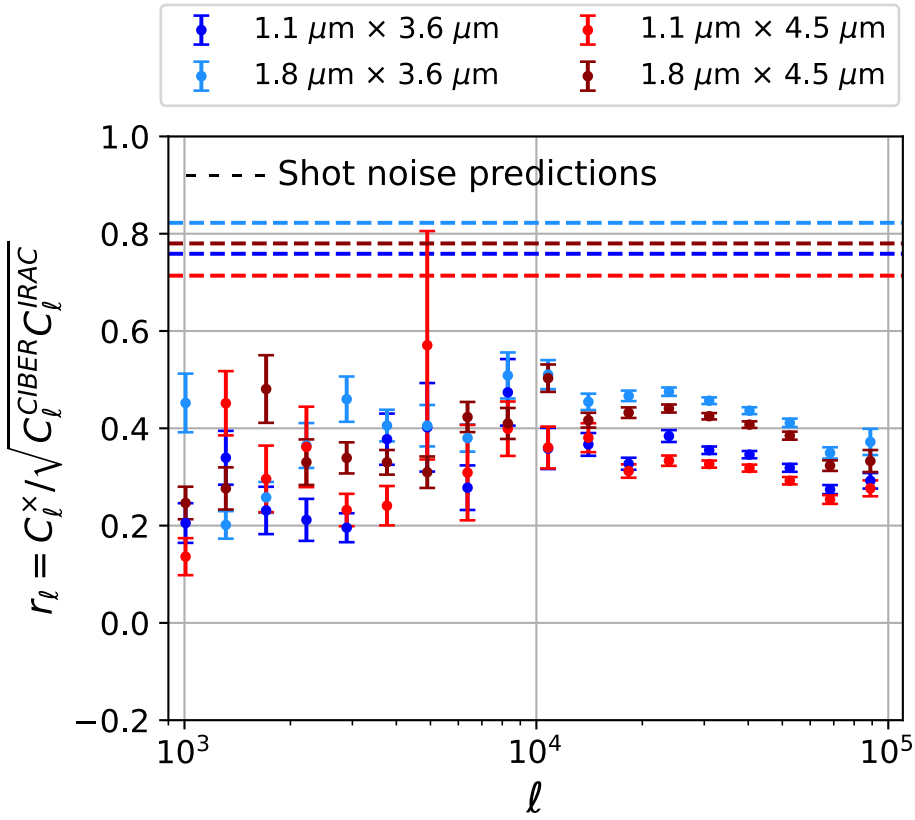

Figure 3.29: Cross-correlation coefficients of *CIBER* and IRAC SDWFS maps as a function of angular multipole. For the reasons described earlier in this section, we report cross-correlations for scales $\ell \geq 1000$. The dashed lines indicate predictions from Poisson noise based on the COSMOS 2015 catalog.

Spectrum of diffuse surface brightness fluctuations from CIBER and Spitzer

In Figure 3.30 we show the wavelength dependence of $\theta = 5' - 10'$ fluctuations measured through auto- and cross-power spectra at 1.1 μm, 1.8 μm, 3.6 μm and 4.5 μm. We note that all maps used to produce these results have consistent source masking depth. While these utilize *Spitzer* data that are also used in previous studies, our independent measurements confirm that the amplitude of fluctuations increases as one goes to shorter NIR wavelengths. Our cross-correlations show various degrees of coherence between bands and instruments, hinting that either multiple components contributing to the fluctuation power on these scales or a single component with some redshift dependence.

In Figure 3.31 we show a similar plot but for auto-spectrum measurements in four bandpowers spanning $\ell = 1000$ to $\ell = 100000$. From these we find that the spectrum

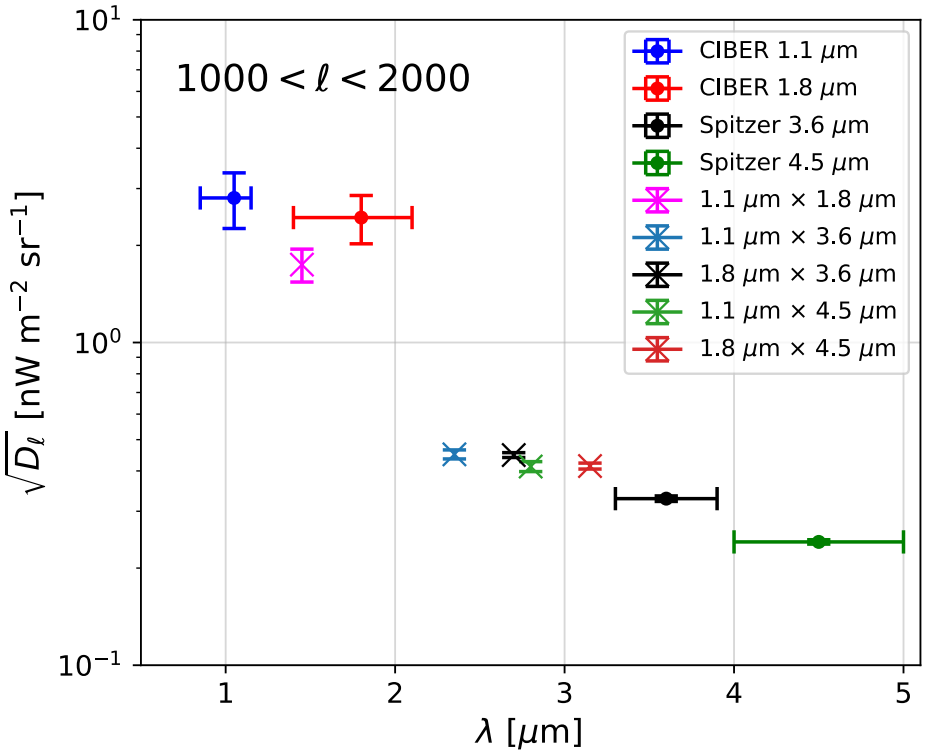

Figure 3.30: Spectrum of large-angle surface brightness fluctuations from *CIBER* and *Spitzer* auto- and cross-spectra.

of fluctuations on large angular scales is considerably bluer than that from Poisson fluctuations. While our *CIBER* auto-spectrum measurements overproduce Poisson noise fluctuations relative to IGL+ISL model predictions, agreement with such models would only increase the level of color variation going from small to large angular scale.

3.9 Conclusion

In this work we present new fluctuation measurements in the NIR using imaging data from *CIBER*'s fourth and final flight. Building on the analysis of [73], we address two important systematics necessary for unbiased recovery of sky fluctuations, namely errors in source masking and correction of FF errors. By using estimates of the FF from stacking the science fields, we bypass the use of field differences used in previous analysis and determine that after FF bias correction, the remaining FF errors increase our recovered power spectrum uncertainties on scales $500 < \ell < 2000$ by less than 20% (presented in Paper I).

With these improvements in combination with the higher quality fourth flight data,

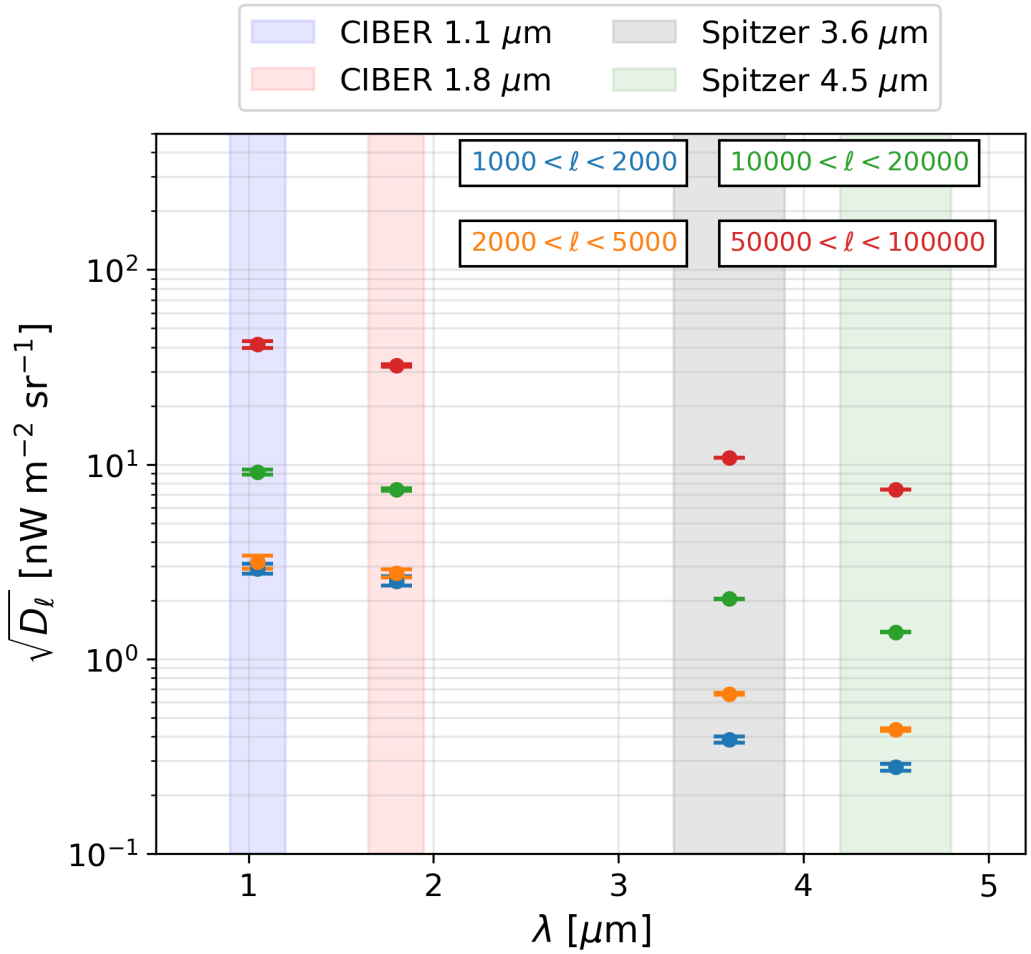

Figure 3.31: Spectrum of surface brightness fluctuations from *CIBER* and *Spitzer* auto-spectra, for a range of broad bandpowers.

we measure large-angle ($\theta > 5'$) NIR fluctuations with three to four times higher sensitivity at 1.1 μm and 1.8 μm. A novel result of this work is that DGL as traced by existing maps does not explain the bulk of large-scale fluctuations in the *CIBER* maps for scales $\ell < 2000$. Through cross-correlations with SFD maps with and without CIB deprojection, we determine that scattered DGL constitutes $\leq 10\%$ of the total fluctuation power on scales $\ell < 2000$. We also measure *CIBER* $\times$ *Spitzer* cross-correlations at high significance down to $\ell = 1000$ suggesting common sky fluctuations between instruments. While our measurements for $\ell < 1000$ are inconclusive, characterizing the SDWFS mosaics in greater detail is a topic of active research and may lead to improved measurements through characterization of the mosaic transfer function and mitigation of large-angle systematics.

In this work we probe mask-signal correlations empirically by varying the size of masks around point sources. We find perturbing the source masking radius function by ±25% about our fiducial case has a negligible impact on the large-scale clustering, however reverting to the masking function used in Z14 has a more significant impact on scales $1000 < \ell < 10000$. In this limit, interpretation of IHL models that describe the extended light profile of galaxies will require a careful treatment of such correlations, since many galaxies containing IHL may be masked.

One lever arm for deciphering properties of the EBL is characterization of its redshift distribution. Consider a cross-correlation between an intensity map I and a galaxy catalog with linear galaxy bias b_g and redshift distribution $p(z)$. The cross-power spectrum will be proportional to $b_g b_I \frac{dI}{dz} p(z)$, where b_I is the halo intensity bias and dI/dz is the intensity as a function of redshift. With knowledge of $p(z)$ and b_g from galaxy surveys and clustering redshifts *a priori*, it is possible to estimate the correlated intensity contribution $b_I \frac{dI}{dz}$ [e.g., 61, 62]. A lack of detection could imply that the bulk of the signal seen by *CIBER* comes from something more local (e.g., within the solar system), or from the more distant universe where spectroscopic catalogs are not available. External galaxy catalogs from optical and infrared surveys probe different redshift kernels and can be cross-correlated with *CIBER* imaging data to understand the galaxy populations that may correlate with the diffuse large-scale fluctuation power.

Modeling the intra-halo light contributions to NIR clustering is an open research topic, however hydrodynamical simulations suggest a number of dependencies between baryon spread and IHL fraction with the dark matter distribution and astrophysical processes (e.g., AGN and SN feedback). As an example, [132] decomposes the disk, bulge and IHL components from the EAGLE simulations [133] and finds strong dependence with galaxy morphology, suggesting a path to targeted cross-correlations that utilizes high-resolution imaging data in combination with intensity maps.

While our analysis uses data with an effective total exposure time of 220 seconds across the five science fields, existing and near-future experiments will map out the NIR EBL over the full sky and with significantly improved spectral coverage and resolving power. *CIBER*-2, the second generation of *CIBER*, has three H2RG detectors and six windowpane filters for imaging at $0.5-2.5$ μm [101, 102]. *CIBER*-2 has flown three times from White Sands, New Mexico in June 2021, June 2023 and May 2024. Along with increased sensitivity, *CIBER*-2 maps enable multi-

wavelength cross-correlations that can help further disentangle the colors of local and extragalactic astrophysical components. Furthermore, *CIBER*-2's coverage of the COSMOS field in the third flight will enable a host of galaxy cross-correlations with extremely deep NIR catalogs.

The Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (*SPHEREx*) is the next NASA Medium Class Explorer mission which is planned for launch in early 2025 [41]. *SPHEREx* will conduct a two-year, all-sky survey in 102 bands spanning 0.75-5 μm, dramatically increasing the volume of data available for intensity mapping [83]. The regions with highest surface brightness sensitivity will be two 100 deg^2 regions covering the North and South Ecliptic Poles. While the effective area covered by the NEP is $\sim$ 5 times that of our *CIBER* 4th flight analysis, observations with *SPHEREx* will be orders of magnitude more sensitive due to higher instrument sensitivity, lower ZL levels and a daily cadence over the poles throughout the nominal two-year survey. This cadence will enable a host of cross-epoch consistency tests. The multi-wavelength dataset will help study and disentangle the various astrophysical components that compose the NIR EBL. Intensity maps over the full sky will have the sensitivity to test for additional foregrounds that are correlated with the galactic plane, the solar system, etc. Both external and *SPHEREx*-internal galaxy redshift catalogs [97] will enable tomographic analyses of CNIB anisotropies [46].

.1 Validation of CIBER auto spectra with field differences

To test the consistency of our auto-spectrum results, we calculate field differenced power spectra as done in Z14. In a field difference, any correlated signals between fields should cancel out while fluctuation power unique to the fields adds linearly. For this reason, field differences are effective for mitigating flat field errors, which couple primarily to the mean sky brightness in each field.

We perform field differences on the two Boötes fields from the fourth flight as these have mean sky brightnesses with $\Delta I/I \sim 5\%$, corresponding to a relative power spectrum bias $\propto (\Delta I/I)^2$. We use a union mask of the individual field masks down to $J < 17.5$ for 1.1 μm and $H < 17.0$ for 1.8 μm, leading to unmasked fractions of $\sim 50\%$ and $\sim 40\%$, respectively. Through the field difference we null the leading order flat field bias (which couples to the bypass a flat field correction entirely and apply a mode coupling correction that depends only on the union mask. We divide the difference spectra by two in order to compare with per-field auto spectra. We

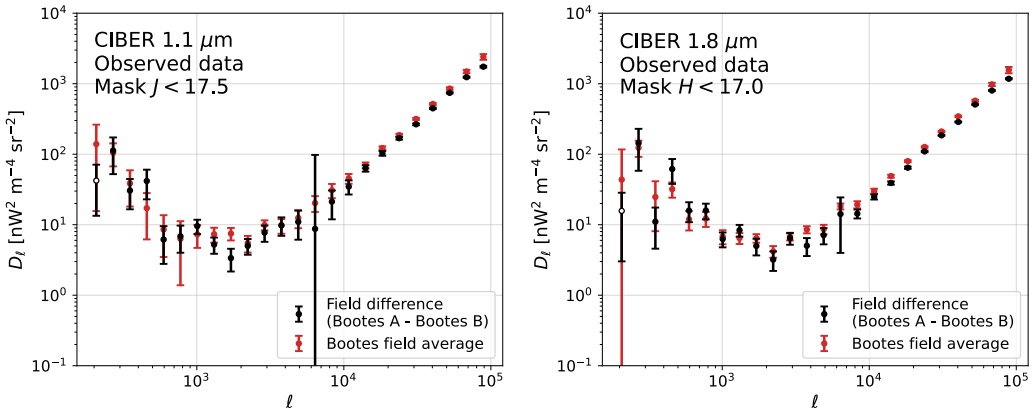

Figure .32: Comparison of auto power spectrum estimates from the two Boötes fields using our fiducial method (red) and difference spectra of the Boötes fields, for which FF errors are mitigated (black). There is broad consistency between both estimates on large and small angular scales, validating our treatment of FF errors.

show the results of this comparison in Fig. .32. While there is some difference in small scale power at the $\sim 10\%$ level, the two sets of power spectra are consistent with one another for scales $\ell < 10000$. Due to the more aggressive union mask, mode coupling on intermediate scales in the field difference is more severe, leading to degraded sensitivity and highly correlated bandpowers.

.2 Choice of image filtering

In this section we compare read noise models with dark data for two choices of pre-processing. The first set of read noise power spectra on the left of Fig. .33 show results from dark data differences derived using with a global (masked) mean subtraction. The right-hand set of power spectra differ in that each quadrant is filtered separately, leading to a more aggressive filter on the largest scales (largely the first five bandpowers).

By choosing to pre-process data quadrants separately, we accomplish two things. First, we reduce the overall variance introduced by the read noise, which otherwise can propagate as errors through the full pipeline, amplified through FF errors, mode coupling induced by the mask, noise bias subtraction, and potentially other effects. In addition, we reduce the discrepancy between our read noise model and the real data, in the sense that our model reproduces the mean and dispersion of the true read noise through Gaussian realizations. Due to the paucity of dark data representative of the flight electrical environment, uncertainty on the variance across realizations per bandpower has a non-negligible statistical uncertainty of $1/\sqrt{5}$ and $1/\sqrt{6}$ for 1.1

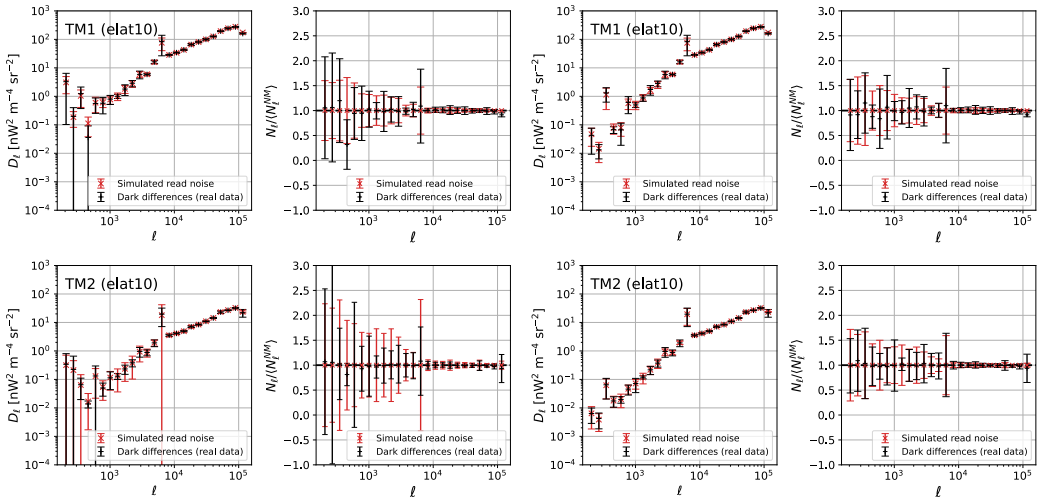

Figure .33: Validation of read noise models for *CIBER* imagers for different filtering choices. The right set of power spectra correspond to those used in the analysis (same as Figure 3.7). Left: comparison of the mean unweighted one-dimensional power spectrum derived from exposure differences taken before flight (black) and from the distribution of power spectra from read noise model realizations (red). Right: Dispersion for the two sets of power spectra, normalized to the mean power spectrum of the simulated noise.

μm and 1.8 μm, so a conservative choice is to slightly overpredict the variance, at the risk of additional uncertainties propagated to the power spectrum. When using the global mean subtraction the read noise model struggles to reproduce the variance on large scales, which could be due to effects of sample variance or non-stationary noise. In contrast, our model matches and/or overproduces variance relative to the dark data, with differences in standard deviation between 10-30%.

.3 Comparison of SDWFS mosaics

To assess the consistency of our results to different mosaicing algorithms, we compute the mean SDWFS auto- and cross-power spectra in the two CIBER Boötes fields using both the SDWFS fiducial mosaics (using the MOPEX algorithm) and those used in [105] with the self-calibration algorithm. We plot the ratio of power spectra for different band combinations in Fig. .34. While both mosaics yield similar results on scales $\theta < 5'$, there are considerable differences on larger scales of up to 50% in some bandpowers. This comparison highlights the importance of mosaicing choices when reconstructing modes much larger than the IRAC field of view, however without a more detailed characterization of mosaicing transfer functions it is difficult to interpret inconsistencies between methods and whether biases

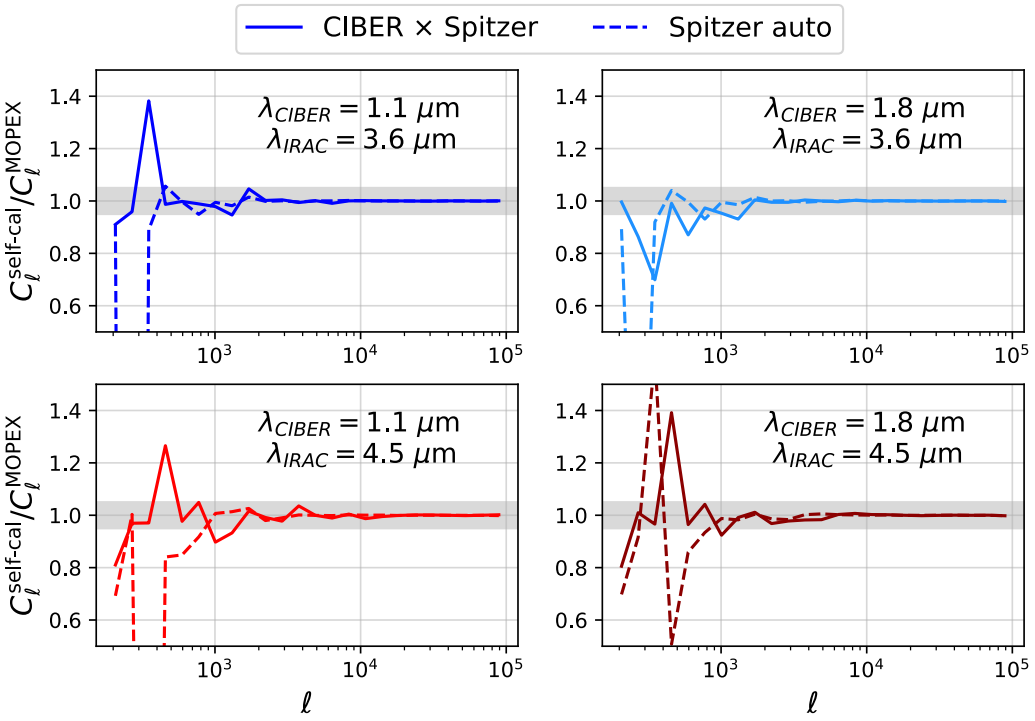

Figure .34: Ratio of recovered *Spitzer* auto-spectra (dashed) and CIBER × *Spitzer* cross-spectra (solid) using the self-calibration and MOPEX mosaicing algorithms. The grey band indicates ±5% deviations in recovered power.

exist in both clustering measurements.

References

[104] T. R. L. et al. In: *ApJL* 927.1, L8 (Mar. 2022), p. L8. DOI: 10.3847/2041-8213/ac573d. arXiv: 2202.04273 [astro-ph.GA].

[106] T. M. et al. In: *ApJ* 742.2, 124 (Dec. 2011), p. 124. DOI: 10.1088/0004-637X/742/2/124. arXiv: 1010.0491 [astro-ph.CO].

[120] R. M. S. et al. In: *High Energy, Optical, and Infrared Detectors for Astronomy III*. Ed. by David A. Dorn and Andrew D. Holland. Vol. 7021. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2008, 70210J, 70210J. DOI: 10.1117/12.789372.

[126] T. A. et al. In: *ApJ* 806.1, 69 (June 2015), p. 69. DOI: 10.1088/0004-637X/806/1/69. arXiv: 1503.04926 [astro-ph.GA].

[130] G. B. B. et al. In: *Astronomical Data Analysis Software and Systems (ADASS) XIII*. Ed. by Francois Ochsenbein, Mark G. Allen, and Daniel Egret. Vol. 314. Astronomical Society of the Pacific Conference Series. July 2004, p. 593.

[131] R. G. A., D. J. F., and S. H. M. In: *ApJ* 536.1 (June 2000), pp. 500–512. DOI: 10.1086/308923. arXiv: astro-ph/0002258 [astro-ph].

[133] J. S. et al. In: *MNRAS* 446.1 (Jan. 2015), pp. 521–554. DOI: 10.1093/mnras/stu2058. arXiv: 1407.7040 [astro-ph.GA].

Chapter 4

THE UNIVERSE SPHEREX WILL SEE: EMPIRICAL GALAXY SIMULATIONS AND REDSHIFT PREDICTIONS

Abstract: We simulate galaxy properties and redshift estimation for SPHEREx, the next NASA Medium Class Explorer. To make robust models of the galaxy population and test spectro-photometric redshift performance for SPHEREx, we develop a set of synthetic spectral energy distributions based on detailed fits to COSMOS2020 photometry spanning $0.1-8$ μm. Given that SPHEREx obtains low-resolution spectra, emission lines will be important for some fraction of galaxies. Here we expand on previous work, using better photometry and photometric redshifts from COSMOS2020, and tight empirical relations to predict robust emission line strengths and ratios. A second galaxy catalog derived from the GAMA survey is generated to ensure the bright ($m_{AB} < 18$ in the i-band) sample is representative over larger areas. Using template fitting to estimate photometric continuum redshifts, we forecast recovery of 19 million galaxies over 30000 deg^2 with redshifts better than $\sigma_z < 0.003(1+z)$, 445 million with $\sigma_z < 0.1(1+z)$ and 810 million with $\sigma_z < 0.2(1+z)$. We also find through idealized tests that emission line information from spectrally dithered flux measurements can yield redshifts with accuracy beyond that implied by the naive SPHEREx channel resolution, motivating the development of a hybrid continuum-line redshift estimation approach.

4.1 Introduction

The Spectro-Photometer for the History of the Universe, Epoch of Reionization, and Ices Explorer (SPHEREx) is the next NASA Medium Class Explorer mission which is planned for launch in early 2025. Using a wide-field 20 cm diameter telescope with an instantaneous field of view of $3.5^\circ \times 11.3^\circ$ for each of two 1×3 detector mosaics, SPHEREx will conduct the first all-sky spectrophotometric survey in the near infrared (NIR) at wavelengths between 0.75 μm and 5 μm through four consecutive surveys over the nominal two year mission.

SPHEREx and other cosmology missions generally need to build and test methodologies using the best possible simulated data. SPHEREx aims to constrain the large-scale distribution of galaxies in order to put constraints on primordial non-Gaussianity [41]. This measurement will need excellent redshifts for hundreds of

millions of galaxies, with robust control and understanding of statistical and systematic errors. In the case of SPHEREx, this is particularly challenging as the properties galaxies will display at infrared wavelengths ($0.75 - 5$ μm) with low-resolution ($R \sim 35 - 130$) spectroscopy are currently poorly constrained beyond the galaxies measured through dedicated NIR spectroscopy, e.g., MOSDEF [134], *Akari* [135], NIRSpec [136], etc.

[137] illustrated the use of synthetic SPHEREx spectrophotometry to measure redshifts of a large sample of bright galaxies to high accuracy ($\gtrsim 10^7$ galaxies over the full sky with redshift accuracy $\sigma_z < 0.003(1+z)$, and many more with accuracy at the $\sim 1 - 10\%$ level). These simulations were performed on model spectra inferred from the Cosmological Evolution Survey field [COSMOS; 138], for which the complexity of the galaxy population is well constrained through deep, 30-band photometry spanning $0.1 - 8$ μm.

Given the spectral resolution and infrared coverage of SPHEREx, emission line galaxies (ELGs) are an important population to consider. Nebular emission lines, typically rest-frame optical and UV (Hα, [OII], [OIII], Lyα), are the targets of numerous existing and upcoming surveys such as the extended Baryon Acoustic Oscillations Spectroscopic Survey [eBOSS; 139], the *Euclid* Wide and Deep surveys [using slitless spectroscopy, 36], the Dark Energy Spectroscopic Instrument ELG survey [DESI; 140], the *Roman* High Latitude Spectroscopic Survey [HLSS; 141], the Prime Focus Spectrograph Galaxy Evolution Survey [PFS; 142], the Physics of the Accelerating Universe Survey [PAUS; 143, 144], and the Javalambre-Physics of the Accelerating Universe Astrophysical Survey [J-PAS; 145], among others. These surveys promise to deliver emission line measurements for tens of millions of galaxies, increasing the size of existing samples by an order of magnitude.

Properly identified emission lines serve as anchors for precise redshift measurements, making ELGs the target of modern large scale structure studies probing cosmic expansion and acceleration [140, 38]. Emission line strengths and ratios are also commonly used as observational proxies for intrinsic galaxy properties such as the star formation rate (SFR). Upcoming ELG surveys will chart galaxy evolution and formation in unprecedented detail, enabling studies of galaxy properties both across a broad range of cosmic history and as a function of local environment [146, 147].

SPHEREx is unique in its potential to deliver precise redshifts using low-resolution spectroscopy. Both active and quiescent galaxies can yield high quality redshifts

through accurate modeling of their continuua [137, 148]. SPHEREx will help in obtaining precise redshifts for a large number of sources by extending source flux measurements into the NIR. In the case where emission lines are used to improve redshifts for continuum-selected galaxies, there may be a large number of sources where proper modeling of low-significance emission lines is important. This places requirements on the accuracy of continuum modeling, however correlations between lines and continuua measured by SPHEREx can potentially break degeneracies of single-line identifications that plague existing spectroscopic surveys [149, 150, 151]. Because SPHEREx is an imaging survey, no pre-selection is required to isolate ELG targets. This distinguishes SPHEREx from fiber-based spectroscopic measurements that are optimized with pre-selected targets. In this sense SPHEREx's survey strategy reduces the impact of target selection on ELG samples and opens up the prospect of blind searches for emission lines [e.g., 152]. Blind searches are possible with IFU surveys such as with HETDEX [153] or *Euclid* grism observations, however SPHEREx can uniquely do so over nearly the entire sky.

The goal of this work is to create and use simulated galaxies to demonstrate the accurate measurement of redshifts using SPHEREx data. Our current best constraints on the distribution of galaxy spectra over the SPHEREx bandpass come from COSMOS2020 photometry. However, the resolution is poorly matched to SPHEREx because COSMOS photometry cannot precisely constrain the emission line properties of sources over a range of redshifts. To tackle this issue, we use empirical [154, hereafter B14] and model-based COSMOS templates to fit the COSMOS2020 photometry, which yields a realistic distribution of SED continua. To model emission lines, we turn to tight empirical relations for how line strengths and ratios scale with redshift, stellar mass, and spectral type. After developing the line prediction method, we test it through comparisons at the population level (line luminosity functions, line ratio distributions, line equivalent widths) and with individual source comparisons using spectroscopic catalogs covering the COSMOS field. The catalog we develop is therefore a faithful representation of the galaxy population as constrained by COSMOS2020 [155] and numerous published studies on emission lines.

Because SPHEREx is an all-sky mission, we complement the COSMOS catalog with wider survey data from the Galaxy and Mass Assembly survey [GAMA, 156] and associated multi-wavelength photometry to ensure that the distribution of bright galaxies in our simulated sample is representative of the full sky (our COSMOS

catalog only covers 1.27 $\deg^2 = 3.9 \times 10^{-4}$ sr, while the GAMA footprint is 217 $\deg^2$).

Our procedure for painting emission lines onto continuum estimates from template fits, implemented in the tool Conditional LIne Painting on Synthetic Spectra (`CLIPonSS`), is described in §4.2, and the synthetic line catalog is validated against several measurements in the literature in §4.3. The details of converting these SEDs into synthetic SPHEREx observations are presented in §4.4, where we also consider the coverage of SPHEREx sources by external catalogs. In §4.5 we forecast recovery of galaxy continuum redshifts by running the photometric redshift template fitting code from [137] on synthetic photometry, also showing initial demonstrations of redshift estimation that utilize low-resolution, spectrally dithered line flux measurements.

All apparent magnitudes are specified in the AB magnitude system [157].

4.2 Synthetic spectral library

Our approach for predicting galaxy emission-line properties is to combine observations from the COSMOS survey with empirical models. The COSMOS photometry has sufficient depth and wavelength coverage to constrain galaxy spectral energy distributions (SEDs) over the SPHEREx bandpass. Our model then predicts emission lines expected for galaxies based on their stellar masses, redshifts, and best-fit spectral templates. Predicting emission line strengths is feasible because of the empirical observed correlations between the lines and other galaxy properties arising from the evolution of the mass-metallicity relation [158], the galaxy main sequence [159, 160, 161] and the global star formation evolution [162, 163].

The procedure for assigning emission lines to galaxy continuua is illustrated in Fig. 4.1. Unlike the COSMOS templates, which lack emission lines, many [154] templates have lines which we rescale when painting new lines onto the synthetic SEDs. In this way, the line strengths in our empirical model are calibrated to observed galaxies at low redshift ($z < 0.1$).

Multi-wavelength photometry

COSMOS2020

The galaxy simulations rely on multi-wavelength photometry from COSMOS2020, which comprises two catalogs with 30-band photometry spanning $0.1 - 8\ \mu$m [155]. In particular we use the COSMOS2020 catalog derived from `The Farmer`, a profile-

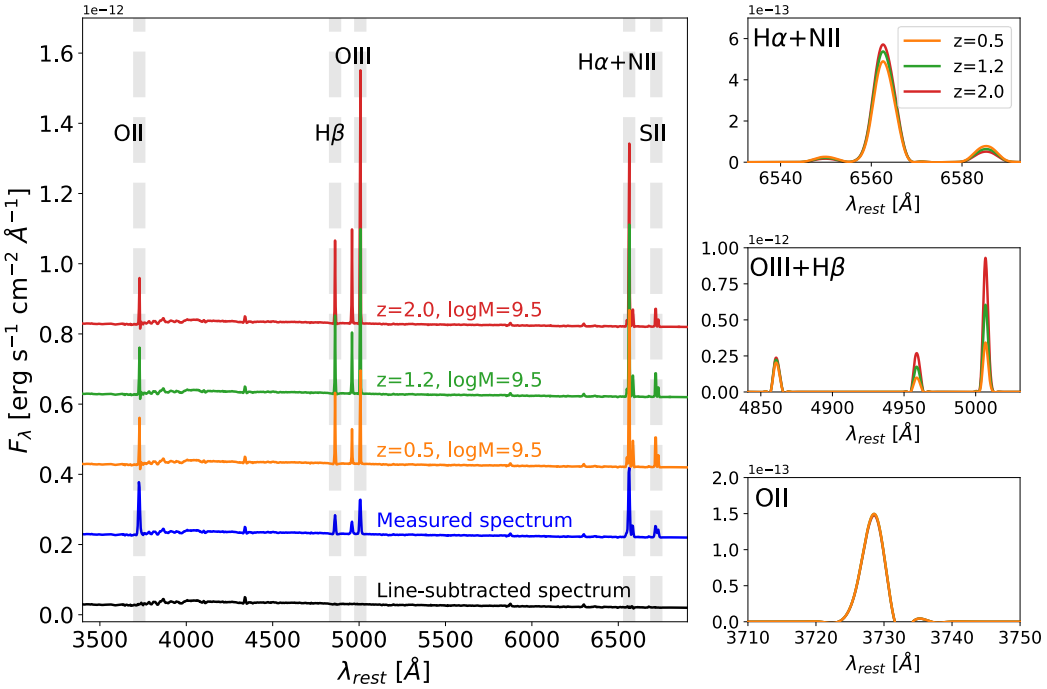

Figure 4.1: Track of synthetic galaxy SEDs as a function of redshift demonstrating our model procedure. For a given source, we start with the best-fit measured spectrum (blue) fit as a template to broad band photometry. We then subtract the existing lines (black), and insert new lines according to our empirical prescription. This is shown for three redshifts: $z = 0.5$ (orange); 1.2 (green); and 2.0 (red), for a fixed stellar mass $\log(M_\star/M_\odot)$= 9.5. Spectra are shifted vertically for visual purposes. The right hand panels show the three lines/line complexes at higher resolution for the three redshifts. [OII] is modeled with a single Gaussian rather than as a doublet.

fitting tool for multi-wavelength photometry, along with associated photometric redshifts.

We select on classified galaxies with $i < 25$ and flux measurements in at least one NIR band (UltraVISTA J and H bands). We further select on sources with consistent redshift measurements from both sets of COSMOS2020 photometry ($\hat{\sigma_z} < 0.2(1+z)$, $|\hat{z}_{\rm Classic} - \hat{z}_{\rm Farmer}|/(1+\hat{z}_{\rm Farmer}) < 0.15$), selecting galaxies as classified by `LePhare` [see §5.1 in 155]. The majority of sources with this selection have accurate redshifts ($\sigma_z < 0.02(1+z)$) and sufficiently precise photometry to estimate stellar masses. In total our selection yields a sample of 166,014 galaxies covering an effective area of 1.27 deg^2.

Bright, low-redshift galaxies from the GAMA survey

There are limitations in building a representative synthetic catalog from COSMOS2020 alone. Given the size of the COSMOS field, the diversity of spectra will be underestimated due to cosmic variance. Likewise, the limited volume probed by COSMOS suggests a limited low redshift sample, especially for massive galaxies. The COSMOS2020 galaxy catalog is limited to sources with $i > 18$, with any sources brighter than this classified as stars.

For these reasons we supplement the COSMOS2020 catalog using a combination of spectroscopic measurements from the Galaxy and Mass Assembly (GAMA) survey [156] and corresponding twelve-band photometry ranging from the far UV to the infrared. We select sources with $i < 18$ and a designated science class (SC) of 8. This is the selection for the primary science catalog used by GAMA – the catalog has spectroscopic completeness of 98 per cent down to $r < 19.5$ [164]. In total we obtain 44,135 sources across four fields with effective areas of 55, 57, 57 and 48 square degrees. We fit the same template library described in §4.2 to the available broad-band photometry with fixed redshifts.

Galaxy Template SEDs

In total we utilize 160 templates to fit observed galaxy photometry (§4.2) and then generate synthetic SEDs and SPHEREx spectrophotometry (§4.4). The two sets of templates described in this section complement each other in terms of reproducing observed colors and galaxy types.

B14 templates

The collection of 129 measured galaxy SEDs from [154] comprise a broad range of galaxy archetypes from different stages of evolution and environment. The SEDs are constructed from a combination of optical [165, 166] and infrared spectroscopy from the *Spitzer* Space Telescope Infrared Spectrograph (IRS) for $5.3 - 38$ μm [167, 168] and *Akari* [135] Infrared Camera (IRC) spectroscopy for $2.5 - 5$ μm (when available), spanning beyond the wavelength range of SPHEREx observations. Regions of the SEDs without coverage are interpolated using model spectra fit using the MAGPHYS model [169], which utilizes a stellar population synthesis (SPS) model (derived from the same set of BC03 templates described in the next sub-section) and a self-consistent prescription for dust emission, absorption and polycyclic aromatic hydrocarbon (PAH) emission in the infrared. While the measured spectra have

little to no coverage in the near infrared, the sections of interpolated spectra are calibrated against existing broadband photometry covering the range from 2MASS [89], *Spitzer* and the Wide-field Infrared Space Explorer [WISE; 170]. The `B14` galaxy SEDs constitute a diverse template basis for fitting galaxy photometry and reproducing observed galaxy colors. Many of the galaxies in the `B14` sample have well measured optical emission lines, which we use to calibrate our line model locally (i.e., $z \sim 0$) before extrapolating to higher redshifts.

COSMOS templates

Thirty one of the templates are SEDs used in [171], which include templates from [172] and twelve model-based templates made from [173] (`BC03` templates, hereafter), which are generated from SPS models along the starburst track and for passive elliptical galaxies. These templates were initially used in [155] to fit the photometric redshifts of COSMOS2020 galaxies. They complement the `B14` templates, as the `B14` templates come from low-redshift galaxies that may not be representative of higher redshift populations.

Template fits to multi-band photometry

The templates described in the previous section are fit to multi-wavelength photometry from the COSMOS and GAMA extragalactic surveys. We use the SED fitter `Fitcat`, used previously in [137], to derive continuum models and galaxy properties. These properties include stellar mass $\log(M_\star/M_\odot)$, dust extinction $\mathrm{E}(B-V)$, dust law and index of the best-fit galaxy template. Note that the photometric redshifts from COSMOS2020 are taken as fixed within `Fitcat`.

Figure 4.2 shows SED fits to COSMOS2020 photometry for six example galaxies. As can be seen, the SED templates are able to capture the properties of both star-forming and quiescent galaxies in our sample. The median reduced chi-squared of the fits is $\chi^2_{red} = 2.2$, with less than 1% of objects having $\chi^2_{red} > 10$. The higher than expected median χ^2 is driven in part by the inability of the model templates to capture the diversity of galaxy properties, in particular at long wavelengths where emission from PAHs is difficult to model. In addition, the flux uncertainties for bright sources may tend to be underestimated [155] and certain correlated uncertainties across wavelengths are neglected. A large fraction of catalog sources are best fit by a small subset (< 10) of the 160 templates, where the “best-fit template" refers to the template corresponding to the smallest model χ^2 compared with the data.

The fraction of sources best fit by the COSMOS templates increases from thirty per cent at $i = 18$ to forty percent at $i = 25$. Likewise, for the GAMA sample, the fraction of sources best fit by COSMOS templates ranges from 35% to 55% across $13 < i < 18$. These trends are consistent with our expectations, as the `B14` templates are calibrated to (local) bright objects.

While there are challenges in the template fitting as discussed above, errors in the reconstructed SEDs are small relative to SPHEREx sensitivity, such that the collection of best fit SEDs provides a realistic testbed for continuum redshift estimation and the resulting redshift forecasts (see §4.5). While we utilize redshift measurements directly from the COSMOS collaboration, poor reconstructions in the SED fitting may lead to biases on the best fit $E(B-V)$ or galaxy type. This potentially affects the distribution of predicted emission line strengths, motivating a number of validation tests on our synthetic line catalog to assess inconsistencies with existing emission line measurements.

Emission line model

The method used to assign emission lines to template continua that are fit to COSMOS2020 and GAMA sources consists of:

- Inferred Hα and [OII] equivalent widths for each galaxy based on its best-fit spectral template and/or UV continuum;
- Observed scaling relations of [NII]/Hα and [OIII]/Hβ with redshift and stellar mass; and
- The (redshift-dependent) nebular extinction correction from [174].

With these measured empirical relations we derive realistic emission lines expected for the COSMOS galaxies, which are then painted onto the best-fit template continua.

Our catalog is similar in qualities to the synthetic Emission Line COSMOS catalog [EL-COSMOS, 174], which is derived from COSMOS2015 photometry [148]. Emission lines in that catalog are modeled though a prescription for Hβ based on estimated star formation rates and assumed metallicities $Z > 0.2Z_{\odot}$ to obtain line ratios (which are effectively fixed aside from dust attenuation). The predictions in this work are similar to those in [174], however our model differs by 1) using a combination of empirically measured and model-based templates, rather than a large

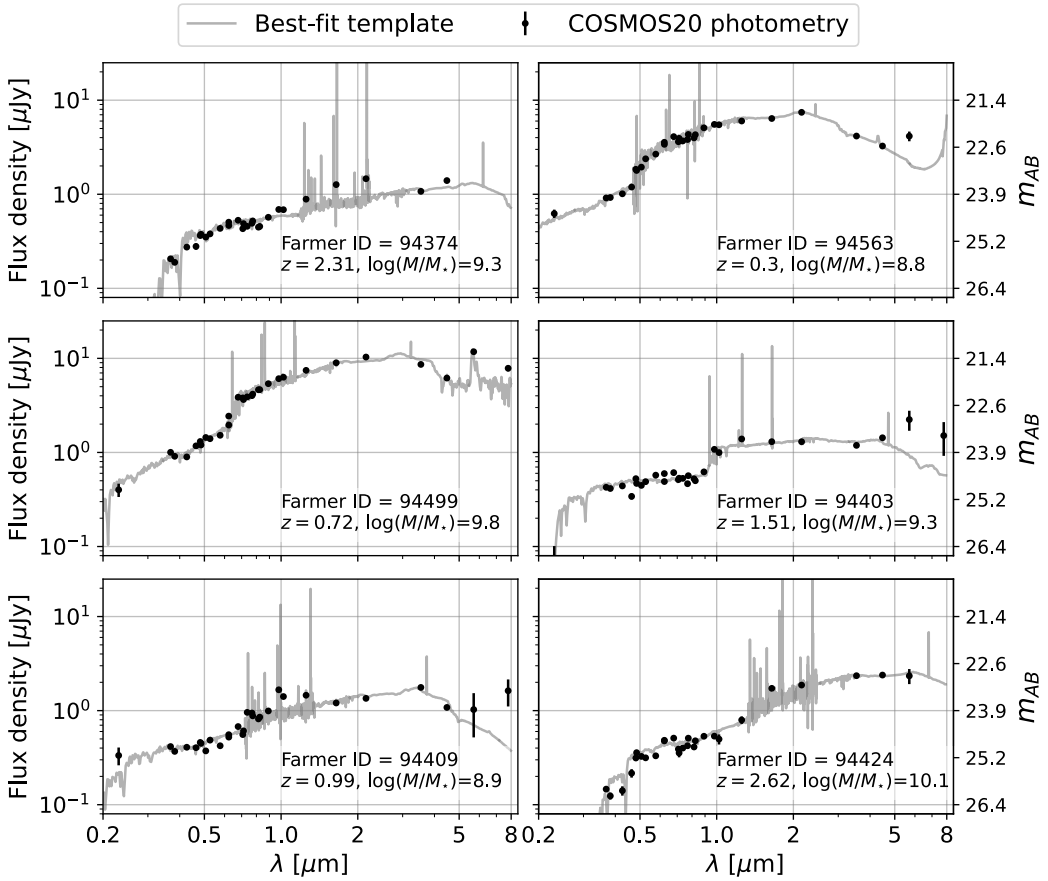

Figure 4.2: A selection of six spectral energy distribution (SED) fits to multi-band COSMOS2020 photometry for galaxies of various redshifts and stellar masses, indexed by Tractor IDs from the COSMOS2020 `Farmer` catalog. In some cases the COSMOS2020 photometry are sensitive to strong emission lines, for example in the top and bottom left panels.

grid of continuum models, and 2) combining empirical scaling relations for lines with locally calibrated ($z \sim 0$) emission line equivalent width (EW) measurements.

The rest-frame optical lines in the SPHEREx wavelength range include: Balmer series lines Hα and Hβ; singly- and doubly-ionized oxygen [OII] [λ3728] and [OIII] [λ5007]; nitrogen [NII] [λ6584]; and the sulfur doublet SII [$\lambda\lambda$6716,6731]. Given SPHEREx NIR spectral coverage, we also include the less probed Paschen-α line at $\lambda_{rest} = 1.87$ μm. Mid-infrared emission from PAHs will also be present, in particular the rest frame 3.3μm bump will be detectable for redshifts $z < 0.5$. As we lack a realistic model of PAH emission in galaxies, we choose to leave any observed PAH emission from the `B14` templates in our synthetic SEDs.

Line	$\lambda_{\rm air}$ [Å]	SPHEREx coverage
Hα	6562.8	$0.15 < z < 6.6$
[NII]	6583.5, 6548.0	same as Hα
Hβ	4861.4	$z > 0.55$
[OIII]	5006.8, 4958.9	$z > 0.5$
[OII]	3728.8	$z > 1.0$
[SII]	6716.4, 6730.8	$z > 0.1$
Paschen-α	18750.9	$0 < z < 1.6$

Table 4.1: Emission lines modeled in this work with rest-frame wavelengths in air. The third column indicates the redshift coverage for each observed line given the full SPHEREx bandpass.

We follow a procedure similar to [175] to extract emission line properties from the set of `B14` templates. In regions of the observed SEDs where relevant SPHEREx emission lines are present, we fit a simple line + continuum model directly to the spectra. For each template, a small region centered on each line (widths varying from $\pm 50 - 100$ Å for rest frame optical lines) is extracted and fit using a variable number of Gaussians depending on the number of spectral features, while the continuum is modeled locally with an offset and slope. From these models we estimate the flux and local continuum level associated with each line, from which we compute the equivalent width. We skip this step for any galaxies classified as passive using the Baldwin, Phillips & Terlevich diagram [BPT diagram; 176] in `B14`.

Predicting emission line strengths

Our procedure for generating sets of emission lines relies on tight empirical relations of lines and line ratios as a function of redshift and stellar mass. All of the empirical relations used in this section are obtained from measurements that have been corrected for dust extinction. Once the intrinsic line fluxes are computed, we then apply stellar and nebular extinction corrections to each source.

To predict Hα line fluxes, we use direct measurements of Hα from the `B14` templates (described in §2.1) to determine each "local" (low-redshift) equivalent width. For the set of active COSMOS templates, which do not include emission lines, we compute the Hα equivalent width using scaling relations between the mean UV luminosity $L_{\rm UV}$ and the star formation rate (SFR)

$$\mathrm{SFR}(M_\odot \mathrm{yr}^{-1}) = 1.4 \times 10^{-28} L_{\rm UV} \quad (\mathrm{erg\ s^{-1} Hz^{-1}}), \tag{4.1}$$

which can then be used to predict the Hα line luminosity $L_{\mathrm{H}\alpha}$ [177];

$$\mathrm{SFR}(M_{\odot}\mathrm{yr}^{-1}) = 7.9 \times 10^{-42} L_{\mathrm{H}\alpha} \quad (\mathrm{erg\ s}^{-1}). \tag{4.2}$$

For `B14` templates with detectable Hα, the rest-frame equivalent width of Hα, denoted EW(Hα), is computed and scaled to higher redshifts using the relation of Fig. 6 from [178] between ΔEW(Hα) and $\Delta \log$ ([OIII]/Hβ), i.e., the deviations of equivalent width and line flux ratio from $z = 0$. We model the redshift dependence of $\log$ ([OIII]/Hβ) using Eq. 1 from [179]

$$\log\left(\frac{[\mathrm{OIII}]}{\mathrm{H}\beta}\right) = \frac{0.61}{\log\left([\mathrm{NII}]/\mathrm{H}\alpha\right) - \gamma} + \eta, \tag{4.3}$$

where $\gamma = 0.02 + 0.1833z$ and $\eta = 1.2 + 0.03z$. We calculate the flux ratio [NII]/Hα as a function of $\log(M_{\star}/M_{\odot})$ and redshift using an interpolation of Table 1 from [180], which approximates the stellar mass vs. gas phase metallicity relation as a function of redshift. The line ratio NII[λ6585]/NII[λ6549] = 3 is applied to each [NII] doublet.

Once the Hα-[NII] complex is calculated, we apply the intrinsic ratios Hα/Hβ=2.86 and P$\alpha/H\alpha = 0.123$ to obtain Hβ and Pα line fluxes. We then use the extrapolated log[OIII]/Hβ from (4.3) and OIII[λ5007]/OIII[λ4959] = 3 to obtain the two OIII line fluxes. The sulfur doublet [SII] is calculated using a best fit parabola to the local relation from SDSS DR12 of the O3S2 BPT diagram [see Fig. 6 of 181]

$$y = -0.44 - 0.3x - 0.66x^2, \tag{4.4}$$

where y =log([SII]/Hα) and x =log([OIII]/Hβ).

We follow a similar procedure to predict [OII] as for Hα. When available, we use [OII] equivalent widths measured directly from the `B14` templates. For COSMOS templates we use the mean SFR-L$_{[OII]}$ calibration from [182], which assumes a fixed (de-reddened) [OII]/Hα ratio, based on measurements from the Near Field Galaxy Survey (NFGS)

$$\mathrm{SFR}[\mathrm{M}_{\odot}\mathrm{yr}^{-1}] = (6.58 \pm 1.65) \times 10^{-42} \mathrm{L}_{[OII]}. \tag{4.5}$$

Dust extinction

When fitting the set of galaxy templates to photometry of COSMOS2020 and GAMA sources, we apply dust attenuation using a grid of extinction curves, namely

those derived in [183, 184, 185, 186, 187]. Although the template fits constrain each galaxy's stellar extinction well, it is known from measurements of the Hα/Hβ Balmer decrement that the dust extinction in nebular regions differs from that of the stellar continuum and tends to be more pronounced at lower redshifts [188, 189, 190, 191, 192, 193]. To account for this differential nebular attenuation we scale the `Fitcat`-derived stellar extinction E$(B-V)$ by the redshift-dependent differential extinction $f(z)$ derived in [174], which is parameterized by a linear function capped at unity

$$f(z) = \frac{E(B-V)_{star}}{E(B-V)_{neb}} = \min(1, 0.44 + 0.2z). \tag{4.6}$$

For all sources we apply nebular attenuation assuming the extinction curve from [184].

4.3 Line model validation

To test the fidelity of the empirical line model, we make a number of comparisons to existing line measurements. These include population-level comparisons of line equivalent widths (Hα+[NII]), luminosity functions (Hα, [OII] and [OIII]) and line ratios as a function of stellar mass and redshift. Where external measurements in the COSMOS field are available, we also make direct, cross-matched comparisons of emission line fluxes. While we have attempted to match the selections of studies in the literature where possible, the collection of measurements we compare against use several methods to select ELG targets (e.g., cuts on broad band colors, narrow band excesses, grism line detections). Furthermore, the methods used to estimate and make corrections (for internal extinction, incompleteness) to line luminosity functions vary and may have residual errors. These choices, along with the selections imposed by our COSMOS2020 catalog, can complicate tests of consistency between our predictions and external measurements, however in general we expect such effects to be most significant at sensitivities beyond SPHEREx survey depths.

H-alpha+[NII] Line equivalent widths

We begin by comparing the predicted equivalent width of the Hα + [NII] complex as a function of stellar mass with measurements from [194] using the VIMOS VLT Deep Survey at $0.2 < z < 0.4$ [VVDS; 195], 3D-HST ($0.8 < z < 1.5$) and [196] ($2.0 < z < 2.6$). The evolution of EW(Hα) with redshift is often measured as an observational proxy for the specific star formation rate (sSFR)-redshift relation [197]. The comparison between our catalog and existing measurements is shown in Fig. 4.3. To compare with the actively star-forming samples (SF), we select all

sources with EW(Hα+[NII])$>$ 3 Å, with the caveat that the samples we compare with come from surveys with varying selections and sensitivities. Nonetheless, the average equivalent widths within each stellar mass bin from our model are in close agreement with measurements.

It is understood that more massive galaxies undergo less vigorous star formation [198, 199], leading to smaller Hα equivalent widths (at fixed redshift). This trend is captured by previous measurements and by our synthetic line catalog. Our catalog also captures the shift to higher equivalent widths at higher redshifts [178]. For galaxies with $10 \leq \log(M_\star/M_\odot) \leq 10.5$, the mean line equivalent width increases from $\langle$EW(Hα+[NII])$\rangle$ = 70 Å for $z = 0.2 - 0.4$ up to $\sim$ 200 Å for $z = 2.0 - 2.6$. For the largest mass bin considered ($11 \leq \log(M_\star/M_\odot) \leq 11.5$) the mean equivalent width goes from $\sim$ 15 Å to nearly 100 Å over the same redshift range. For all redshift bins, the average EWs in the lower two stellar mass bins are in close agreement with measurements. In the highest mass bin, our average EWs are consistently lower. This may reflect mismatch in the selection of "star-forming" (SF) galaxies; our classification is determined at the level of the template SED library, while in [194] all $> 3\sigma$ detections are labeled as star-forming, which depends on the noise properties of each survey's observations. The high mass comparison is most sensitive to such selection biases and is exacerbated by the small sample statistics of the measured high-mass samples ($\sim$ 10 detections per redshift bin). With a larger emission line sample it should be possible to discern between this and potential deficiencies in our model for high-mass galaxies.

Line luminosity functions

From the catalog of line fluxes and redshifts we compute line luminosity functions (LFs). For each LF we select all catalog sources with estimated redshifts within 0.05 of $\overline{z}$, where $\overline{z}$ denotes the central redshift of the externally measured LFs. The width of the redshift bins is chosen to be small enough that redshift evolution is negligible within the bins but large enough to obtain good sample statistics. We use the total comoving volume within each redshift slice (i.e., within $[\overline{z} - 0.05, \overline{z} + 0.05]$ with $A_{eff} = 1.27$ deg^2) to normalize the LFs. The synthetic LFs presented are not corrected for the Eddington biases sourced by flux uncertainties in the photometric catalogs [c.f. 174], and the LFs include the effects of intrinsic dust attenuation. For each line and redshift, we compute the 2σ line luminosity threshold corresponding to SPHEREx's observed flux sensitivity, discussed in more detail in §4.4. These are shown to highlight the portions of each LF that drive our

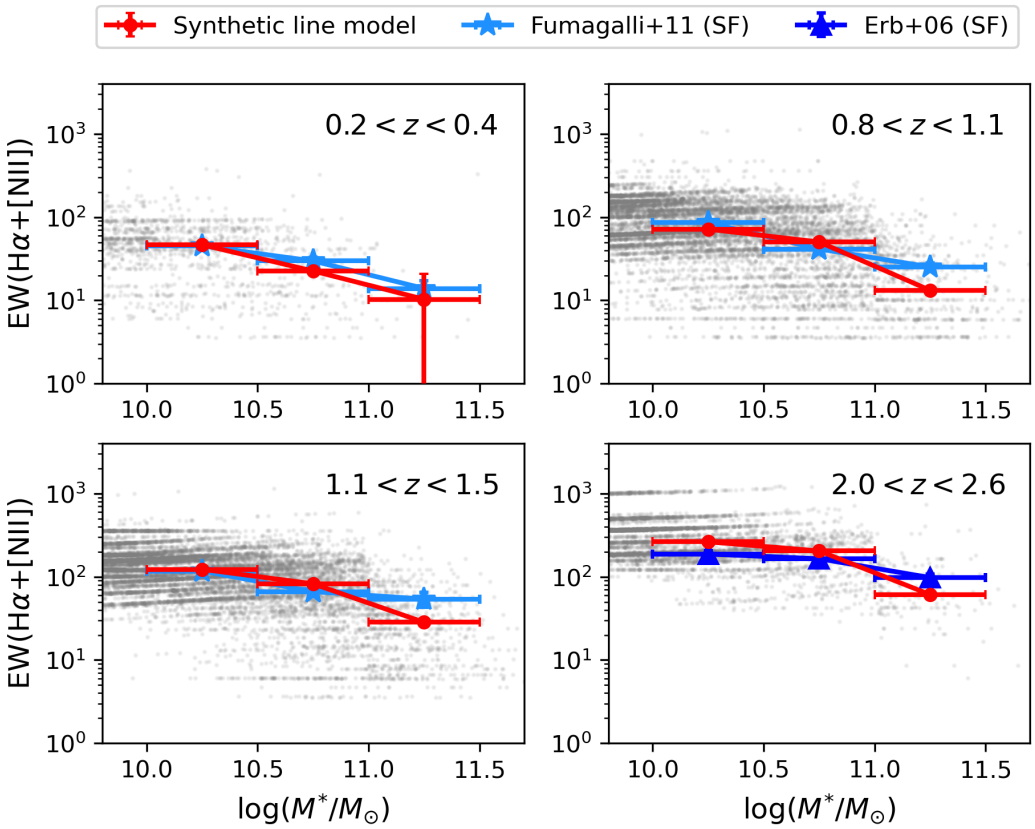

Figure 4.3: Predicted Hα + [NII] equivalent widths of our synthetic line catalog as a function of stellar mass and redshift (grey points), with the binned average of the synthetic catalog plotted in red. These relations agree with previous spectroscopic measurements in the literature presented in [194] (light blue) from VVDS ($0.2 < z < 0.4$), 3D-HST ($0.8 < z < 1.5$) and [196] (dark blue, $2.0 < z < 2.6$).

emission line predictions later in §4.4 and §4.5. These limits support the fact that faint-end incompleteness effects in the synthetic catalog do not have a large impact on the SPHEREx predictions in this work.

H-alpha

We evaluate the synthetic Hα LF for redshift bins centered at $\overline{z} = 0.4, 0.84, 1.47,$ and 2.23. Figure 4.4 shows our derived Hα LFs compared with measurements and best-fit Schechter functions from [200], in which Hα measurements are corrected for [NII] contamination following the relation from [201]. On the bright end, our model is in close agreement with measurements in all redshift bins for $L \geq L_*$. Our LFs tend to fall steeply at fainter luminosities, underestimating ϕ for $L < L_*$. Unlike the LFs from [200], our derived LFs are uncorrected for luminosity incompleteness. The

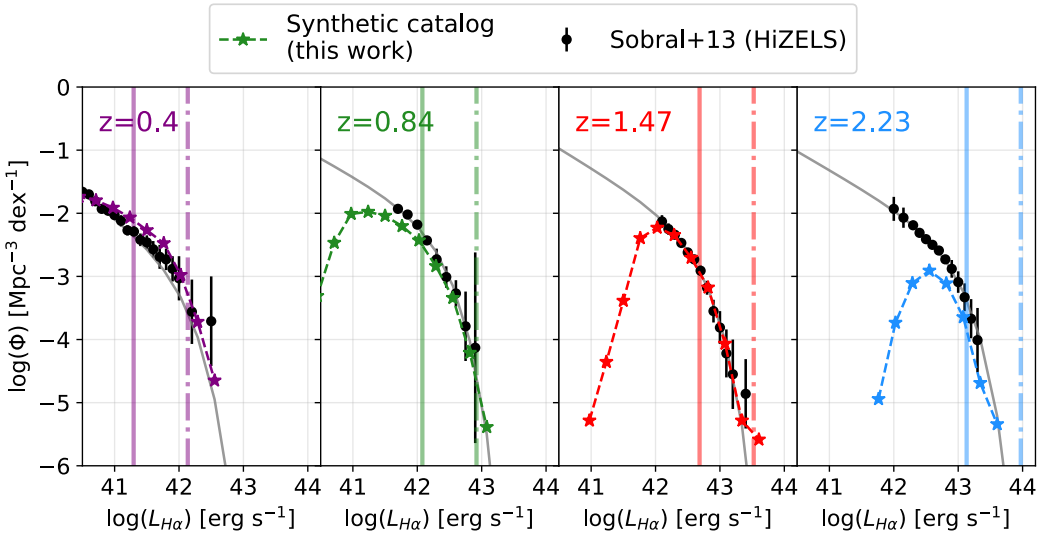

Figure 4.4: Hα line luminosity function evaluated at $z = 0.4$ (purple), 0.84 (green), 1.47 (red) and 2.23 (blue), comparing best-fit Schechter functions from [200] with empirically drawn fluxes from our model. The vertical lines indicate the estimated SPHEREx 2σ line luminosity limits at fixed redshift (see §4.4), for full-sky (dash-dot) and deep field (solid) coverage. The bright end LFs from our synthetic line catalog, which will drive the SPHEREx full sky emission line detections, are consistent with those from HiZELS within measurement uncertainties. The synthetic line LFs shown have not been corrected for catalog incompleteness, which explains the roll off in the synthetic LFs at higher redshift/lower $L_{H\alpha}$.

underprediction of ϕ is explained primarily by our selection on the COSMOS2020 catalog, rather than by the line flux modelling. Note that this behavior is common to all lines in our catalog. As the COSMOS catalog goes considerably deeper than SPHEREx photometry, this faint-source incompleteness starts to set in after the SPHEREx faint-source limit is reached, and as such is acceptable for modeling SPHEREx observations. Our line luminosities have a lower limit that is fainter than measurements in the [200] sample (with the exception of the $z = 2.23$ bin), which we understand as reflecting the fainter population probed by the COSMOS2020 catalog through broad-band continuum sensitivity.

[OII] doublet

The [OII] line will fall in SPHEREx's bandpass for galaxies with $z > 1$. While only the brightest [OII] lines will be detectable at SPHEREx full-sky depth (see Section 4.4), there are near-future telescopes that will measure the doublet through optical spectroscopy at lower redshift, for example *Euclid* [36] and *Roman* [141]. For this

reason we validate [OII] across a range of redshifts.

To validate [OII] for lower redshifts we compare our synthetic LFs with measurements from the FORS2 instrument at the Very Large Telescope (VLT) and the SDSS-III/BOSS spectrograph, along with measurements from GAMA, zCOSMOS and VVDS [202]. We plot LFs derived from our line catalog calculated in six redshift bins spanning $0 < z < 1.3$ alongside these measurements in Fig. 4.5. We also include LF measurements from the HETDEX pilot survey for $0 < z < 0.4$ [203] and from the Deep Extragalactic Evolutionary Probe 2 (DEEP2) galaxy survey for $0.7 < z < 1.3$ [204]. For redshift bins centered on $z = 0.15$ and $z = 0.3$, our LFs agree down to $L \sim 10^{40}$ erg s^{-1}. For $z > 0.5$ the bright end LF is consistent with [204], though both are higher than measurements from [202] between $L \sim 10^{41} - 10^{42}$ erg s^{-1}.

In Fig. 4.6 we compare [OII] LFs with measured constraints from the HiZELS survey [205]. The bright-end LFs broadly agree with one another aside from in the highest redshift bins, where the synthetic COSMOS2020 catalog contains few sources. In general these results indicate that [OII] is captured well at the population level.

[OIII]+H-beta

The [OIII] + Hβ complex will be detectable by SPHEREx for galaxies with $z > 0.55$ and can potentially aid redshift measurements if measured in addition to Hα and/or [OII]. We show line LFs for [OIII] and [OIII]+Hβ in Fig. 4.7. Our predicted bright end LFs are higher on average than those derived from the HiZELS survey [205] in all redshift bins. They are also higher than those from grism measurements of $z \sim 2$ galaxies using HST WFC3 [206]. In the lower two redshift bins our bright end predictions more closely agree with [207] and a recent study using $1.2 < z < 1.9$ emission line galaxies identified on the 3D-HST grism [208]. Given the disagreement in measurements between [205] and [207], it is difficult to evaluate the significance of the model disagreement with [205].

Paschen-alpha

The broad spectral coverage and resolving power offered by SPHEREx in the near-infrared enable detection of the 1.87 μm Paschen-α line. Figure 4.8 shows LF predictions for Paschen-α in four redshift bins between $0 < z < 1.6$, taking the sum of COSMOS2020- and GAMA-derived LFs for $z = 0.2$ and $z = 0.4$. The bright end LF evolves mildly for redshifts $z > 0.5$.

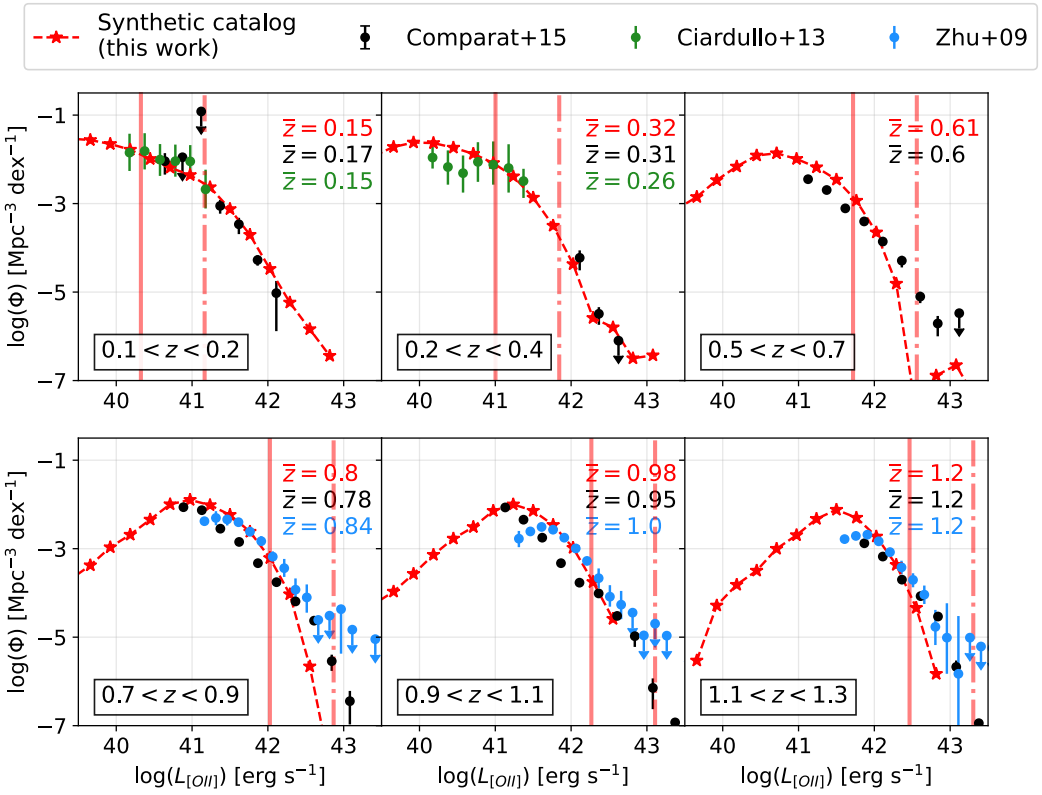

Figure 4.5: [OII] luminosity function evaluated in six redshift bins between z = 0.1 and z = 1.3. Line fluxes from our model are compared with a combination of emission line measurements from [202] (black points), [204] (blue) and [203] (green). As in Fig. 4.4 the synthetic line LFs shown have not been corrected for catalog incompleteness. While for z = 0.15 and z = 0.32 our LFs are consistent with measurements down to $L_{[OII]} \sim 10^{40}$, our line catalog at higher redshifts tends to underpredict the bright end LF relative to [204] and [202].

Direct line flux comparison in the COSMOS field

We complement validation of our line model at the population level with direct comparisons to existing line flux measurements. One-to-one comparisons of cross-matched sources in the COSMOS field allow us to quantify any consistent biases as a function of line flux while controlling for the properties of the galaxies (assuming they are well constrained by one or both surveys).

The Hα sample from [200] overlaps partially with the COSMOS field. The measured and predicted line fluxes are plotted in Fig. 4.9, in which no extinction correction is applied. Positional cross matches are performed for galaxies in the same redshift bins as Fig. 4.4, however there are many sources for which the estimated redshift from [155] differs by more than $|\Delta z| = 0.2$ from the nominal redshift bin (indicated

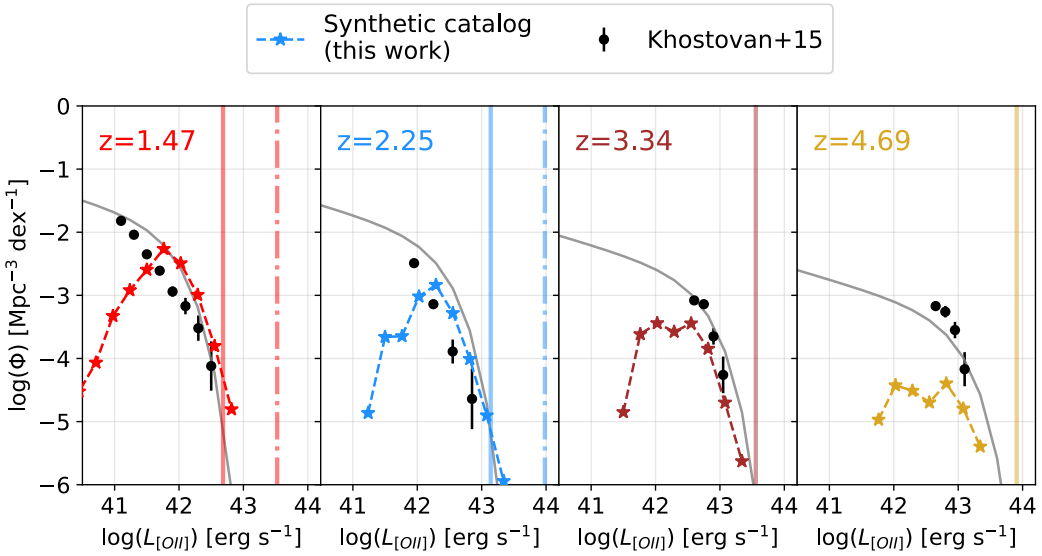

Figure 4.6: [OII] luminosity functions evaluated from our synthetic line catalog at $\bar{z} = 1.47$ (red), 2.25 (blue), 3.34 (brown) and 4.69 (yellow), compared with measurements from HiZELS [205, black points] and their best-fit Schechter functions (grey lines). The incompleteness in the synthetic catalog becomes more pronounced at high redshifts, especially in the $z = 4.69$ bin.

by open circles). These discrepancies are most pronounced for fainter line fluxes, and could be caused by outliers in the COSMOS2020 catalog. The synthetic line fluxes are relatively unbiased for $F_{H\alpha} > 10^{-16}$ erg cm^{-2} s^{-1} and at low redshift, with a more prominent positive bias for higher redshift bins.

We compare our model fluxes with two other sets of emission line measurements in the COSMOS field, namely zCOSMOS-Bright [174, 209] and 3D-HST [210, 211]. These surveys have obtained spectroscopic redshifts in the COSMOS field that are magnitude-limited to $i_{AB} < 22.5$. From zCOSMOS and 3D-HST, Hα is measured for redshifts $z \leq 0.46$ and $0.67 \leq z \leq 1.6$, respectively, while for [OII] they cover $0.47 \leq z \leq 1.57$ and $1.95 \leq z \leq 3.56$. [OIII] is measured in the redshift ranges $0.11 \leq z \leq 0.92$ and $1.19 \leq z \leq 2.39$, respectively. The line fluxes from zCOSMOS are aperture corrected following the procedure from [212], which uses the measured sizes of the galaxies to estimate the fraction of total flux falling within the one arcsecond slit. [174] estimates the line flux completenesses of zCOSMOS and 3D-HST to be $\log(F_{line}/\text{erg cm}^{-2}\text{ s}^{-1}) = -15.8$ and $\log(F_{line}/\text{erg cm}^{-2}\text{ s}^{-1}) = -16.5$.

Figure 4.10 shows the ratio of predicted and observed fluxes for Hα, [OII] and

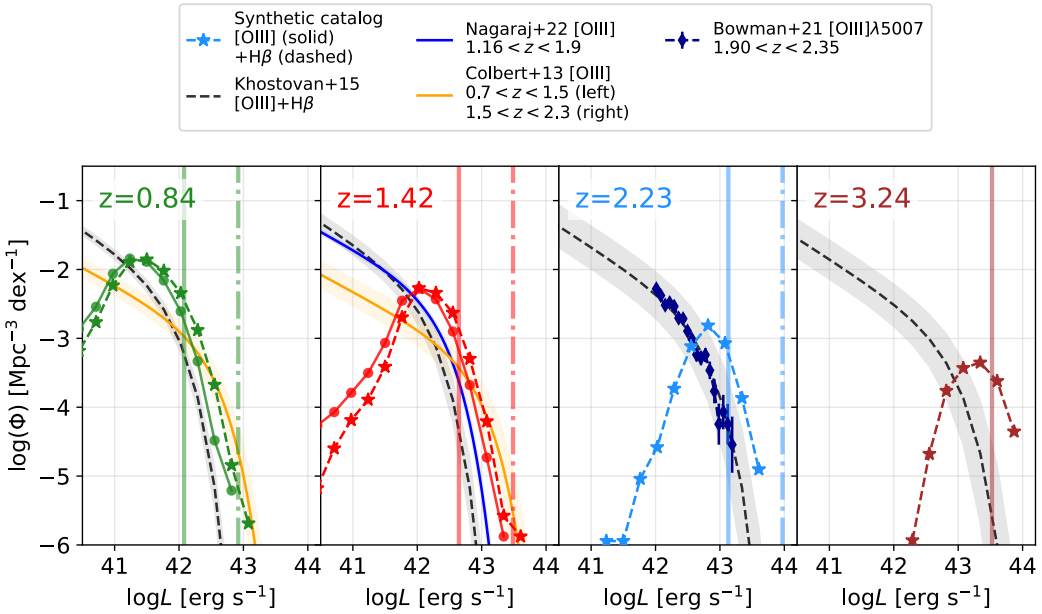

Figure 4.7: [OIII] (solid) and [OIII]+Hβ (dashed) luminosity function evaluated from our synthetic line catalog at in four redshift bins spanning $z = 0.84$ (left, green) and $z = 3.24$ (right, brown). The black line and grey shaded regions show LF constraints from [205]. Similar constraints for [OIII] are shown from [207] (yellow) and [208] (blue), while measurements from [206] (blue diamonds) are included for our $z = 2.23$ bin. While at high redshifts ($z = 2.23$, $z = 3.24$) our synthetic line catalog strongly overpredicts the bright end [OIII]+Hβ LF relative to [205], overprediction in the two lower redshift bins is difficult to interpret given the inconsistency across existing measurements.

[OIII], with the mean and scatter (binned by log-flux) plotted in black. There is strong agreement with zCOSMOS-measured fluxes for all three lines, though for [OIII] our model fluxes overestimates fluxes with $F_{[OIII]} < 2 \times 10^{-16}$ erg cm^{-2} s^{-1}. This bias is consistent with trends seen in [174] and our direct comparisons with HiZELS (see Fig. 4.9. For the 3D-HST sample, which contains more high redshift sources, our model Hα fluxes in close agreement with measured fluxes on average. However, for [OII] and [OIII] our model fluxes tend to underestimate measured fluxes above $F \sim 1 \times 10^{-16}$ erg cm^{-2} s^{-1}.

We consider several potential explanations for why our model [OII] and [OIII] fluxes are low relative to the observed sample:

- The predicted quiescent fraction may be too high for this selection of sources when in reality they are star forming. The fraction of cross-matched zCOS-

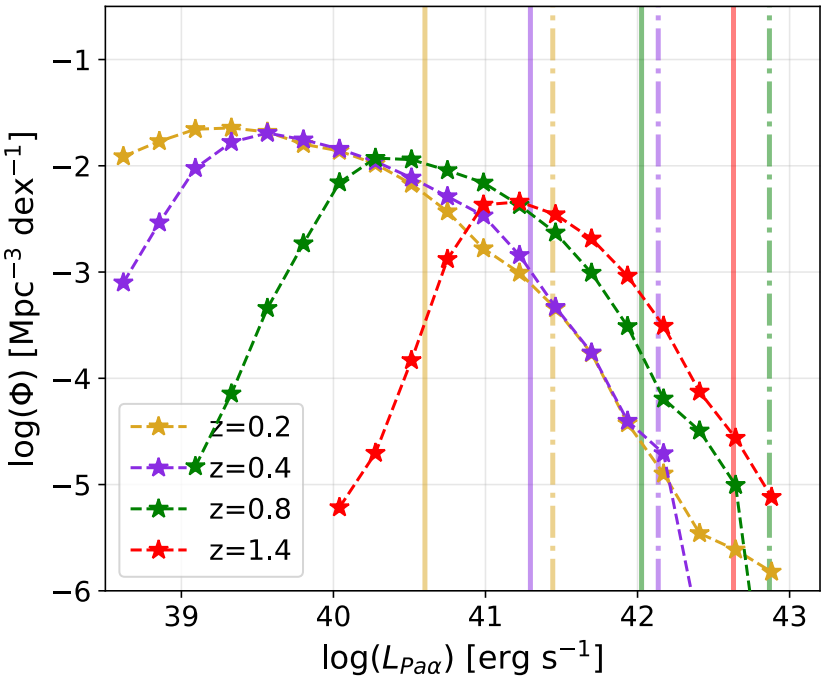

Figure 4.8: Paschen-α line luminosity function predicted from the synthetic catalog in four redshift bins spanning $0 < z < 1.6$, i.e., the redshift coverage for Paschen-α by SPHEREx.

MOS sources above the completeness limit (dash-dotted lines) labeled as passive by our model (i.e., $F_{line} = 0$) is 3.0%, 2.0% and 1.3% for Hα, [OII] and [OIII], respectively, and for the 3D-HST sample the corresponding fractions are 4.6%, 9.4% and 10.5%.

- Incorrect redshift assignments could bias the fluxes of the sources low. Only a handful of COSMOS2020 sources have discrepant redshifts when compared against spec-zs from either zCOSMOS or 3D-HST. Removing sources with $\delta z/(1+z) > 0.1$ from the comparison does not ameliorate the discrepancies seen in [OII] for the $1.95 \leq z \leq 3.56$ sample. [211] quotes a redshift scatter of $\sim 10\%$ for $z \sim 2$ galaxies when compared against duplicate measurements and follow-up observations from the MOSDEF survey [134] for $JH_{IR} < 24$, which is the case for the galaxies in question with $F_{line}^{3DHST} > 10^{-16}$ erg cm^{-2} s^{-1}.

- Likewise, the discrepancy could be present if the 3D-HST fluxes for these objects are biased high. There is evidence for a positive bias in [OII] and [OIII] on average for 3D-HST fluxes for $F_{line}^{3DHST} \lesssim 10^{-16}$ erg cm^{-2} s^{-1} however the

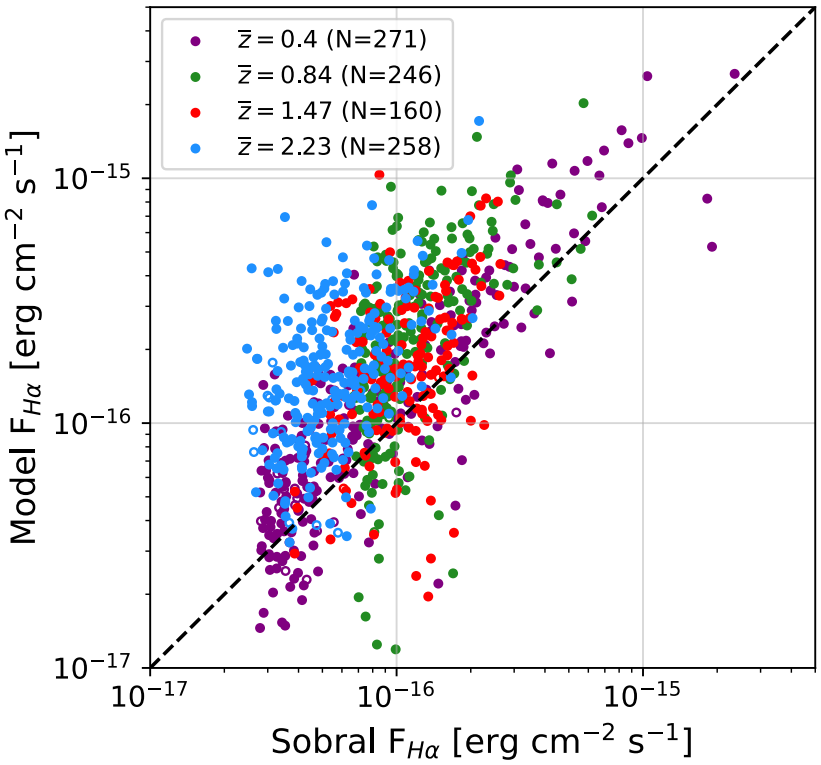

Figure 4.9: Direct Hα line fluxes compared with measurements from [200] for bins with $\bar{z}$ = 0.4 (purple), 0.84 (green), 1.47 (red) and 2.23 (blue). Open circles indicate galaxies where the COSMOS2020 photometric redshift differs from that of the nominal [200] redshift bin by more than 0.2.

paucity of direct comparisons prohibits us from assessing whether this fully explains the observed differences.

- An incomplete model of nebular attenuation may impact our [OII] and [OIII] predictions, in particular at higher redshifts.
- If there is a non-negligible AGN fraction in the high-redshift sample it may explain the higher line fluxes compared to our model fluxes. Because the WFC3 G141 grism has moderate resolution (R = 130) it is not possible to resolve the Hα+[NII] complex, meaning we cannot distinguish between star forming galaxies and AGN through the BPT diagram (e.g., Fig. 4.12). AGN have been identified and removed in the COSMOS2020 catalog using morphological and SED criteria, however it is likely there is additional AGN contamination in particular at the higher redshifts considered.

At the population level, our line model captures the scatter in line strengths seen

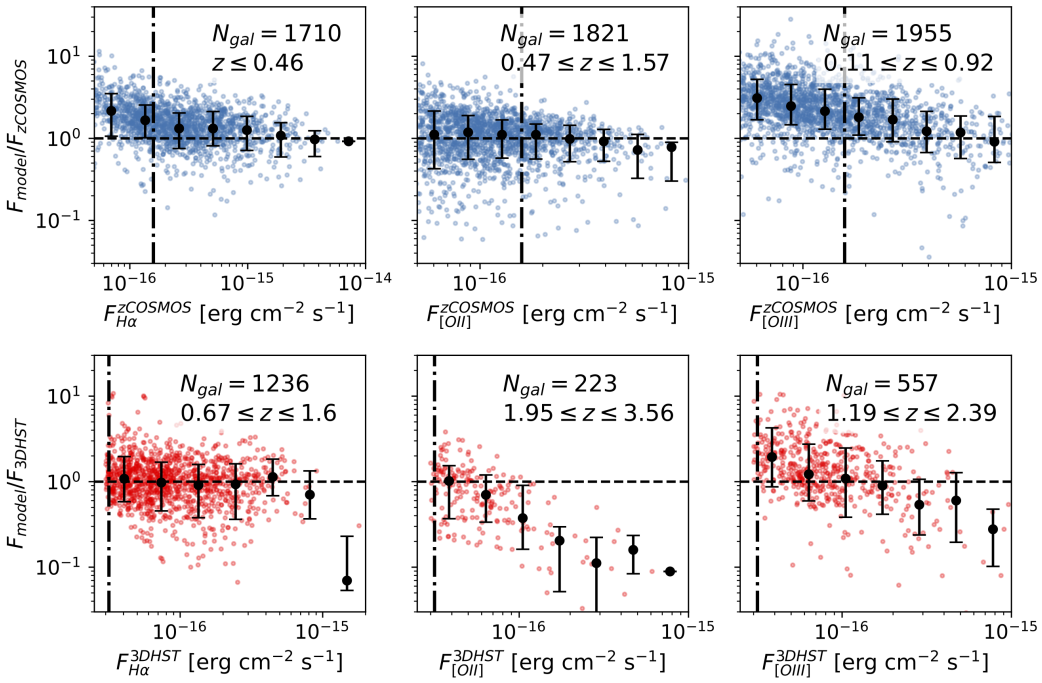

Figure 4.10: One-to-one line comparisons of Hα (left), [OII] (middle) and [OIII] (right) between model predicted fluxes and direct flux measurements from zCOSMOS (top row) and 3D-HST (bottom row). The black errorbars show the mean and 68 percentile range of the colored points in bins of equally spaced log-fluxes. The grey dash-dot lines indicate the completeness limits of the respective surveys as estimated in [174] (see Fig. 2 from that work). With the exception of [OII] in $1.95 \leq z \leq 3.56$ (bottom middle panel), for which the predicted line fluxes are signifiantly lower than directly measured fluxes above $F_{[OII]} \gtrsim 10^{-16}$ erg cm^{-2} s^{-1}(discussed in the text), our synthetic catalog yields fluxes consistent with measurements above the completeness limits of each survey on average.

in the two measured line catalogs. Figure 4.11 compares the distribution of line fluxes from our model with those from zCOSMOS and 3D-HST as a function of stellar mass. For the low redshift sample we place an additional cut on $i < 22.5$ to match the zCOSMOS selection and find that our model fluxes adequately cover the range of measured fluxes. When compared against 3D-HST, our model fluxes match both the core and tails of the measured flux distributions. While we use mean trends to predict line strengths and ratios, our synthetic line catalog are conditioned on estimates of redshift, stellar mass and dust attenuation, all of which contribute to the observed flux scatter. Reproducing this scatter in our redshift predictions is important for obtaining realistic forecasts on the number of SPHEREx sources with line detections (see §4.4).

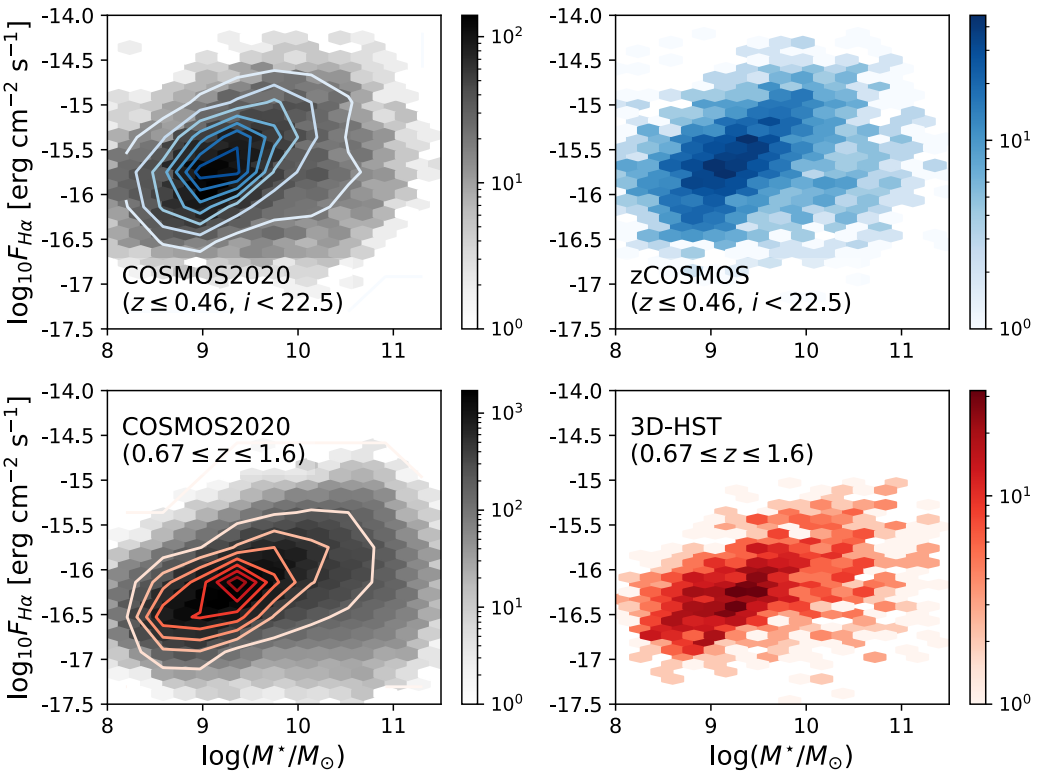

Figure 4.11: Log-density plot of Hα line fluxes and stellar masses, for synthetic fluxes from our COSMOS2020 catalog (black) and for observed samples from zCOSMOS (blue) and 3D-HST (red). The two columns show the COSMOS2020 catalog with redshift cuts matching the Hα ranges of zCOSMOS and 3D-HST. Density contours for the spectroscopic catalogs are overlaid on the synthetic flux distributions for comparison. This comparison lends confidence that our synthetic line catalog spans a representative range in $F_{H\alpha}$ and $\log(M_\star/M_\odot)$ and that our model adequately captures scatter in $F_{H\alpha}$.

Line ratios/trends

Evolution of the line ratios as a function of redshift can be seen by placing galaxy line ratios on the O3N2 BPT diagram. This is shown in Fig. 4.12, with synthetic line ratios color-coded by redshift. While the population of low-redshift galaxies resides in a locus centered on the relation from [178] (red, dashed), moving to higher redshift has the effect of shifting this locus upward on the BPT diagram. This trend is consistent with measurements from MOSFIRE [213] of a sample of 251 galaxies between $z = 1.5$ and $z = 3.5$ ($\langle z \rangle \sim 2.3$).

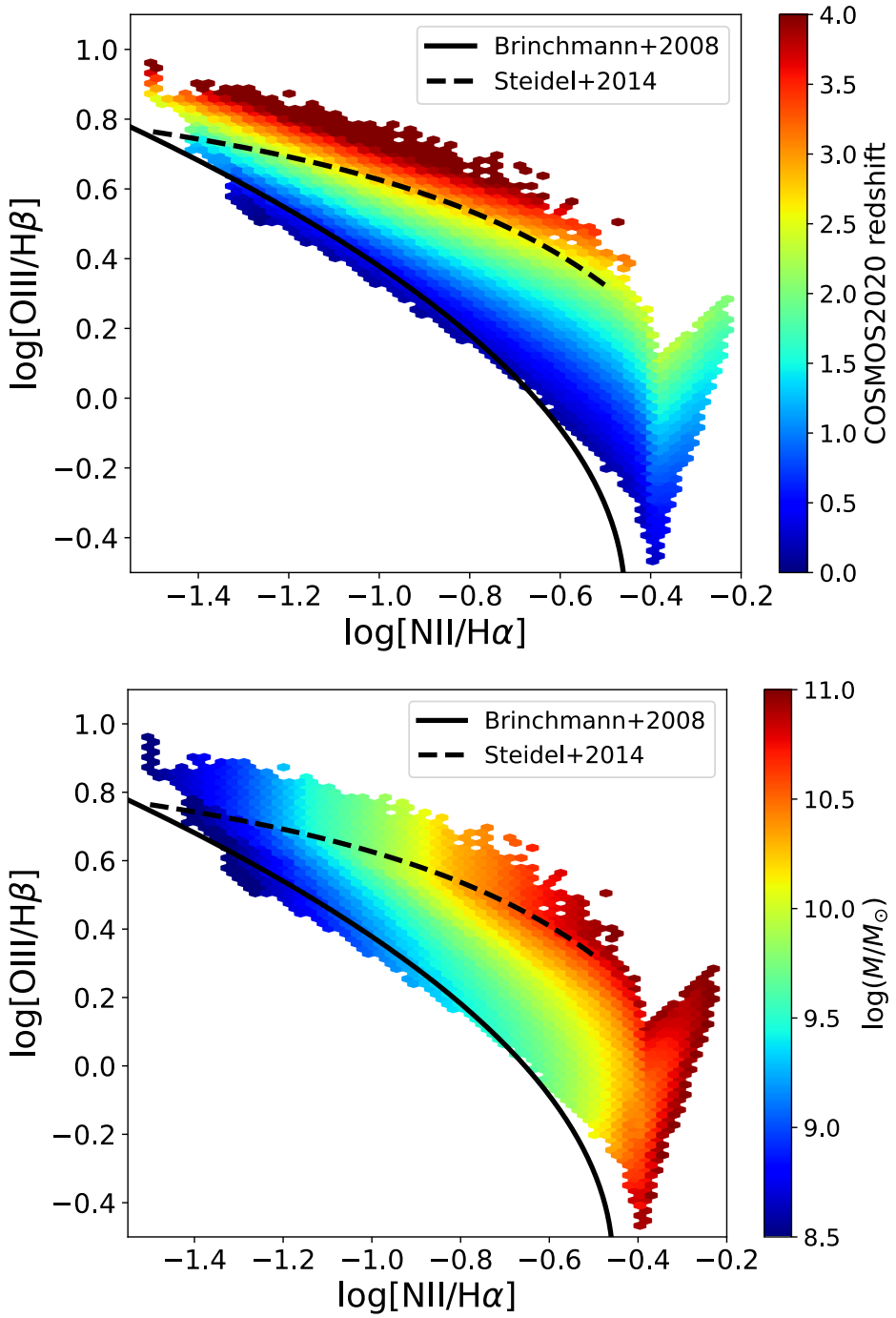

Figure 4.12: O3N2 BPT diagram with line ratios from the synthetic line catalog, plotted against measured trends of local ($z \sim 0$, [178]) and higher redshift ($\langle z \rangle = 2.3$, [213]) galaxies. Plots show line ratios of the same sources, colored by redshift (top) and stellar mass (bottom). The synthetic catalog matches trends in the BPT diagram measured by both [178] and [213].

Quiescent fraction

The fraction of passive galaxies is important to quantify as a check that our model does not overproduce star-forming galaxies with strong emission lines. Using our synthetic line catalog, we find that 7% of objects are either best fit by passive galaxy templates (see §4.2) or have Hα equivalent width less than 5 Å. At low redshifts ($z < 1$), the quiescent fractions of our synthetic catalog for mass bins $\log(M_\star/M_\odot) \in (10, 10.5]$, $\log(M_\star/M_\odot) \in (10.5, 11.0]$ and $\log(M_\star/M_\odot) \in (11, 11.5]$ are 20%, 34% and 44%, respectively. We also compute quiescent fractions the more standard separation using the UVJ diagram [214, 215]. We use the rest frame $(NUV - r)$ vs. $(r - J)$ selection from [216] ($NUVrJ$ in shorthand):

$$(NUV - r) > 3(r - J) + 1 \text{ and } (NUV - r) > 3.1. \tag{4.7}$$

The absolute magnitudes are calculated in the `Farmer` catalog from [155] using the best-fit photo-z solutions. Using this classification, we find the same selections of galaxies have quiescent fractions of 29%, 41% and 53%, respectively. While our quiescent galaxy classification using EW(Hα) is more conservative than using $NUVrJ$, the trends of both classifications match the expectation of an increasing fraction of quiescent galaxies with higher stellar mass. Our $NUVrJ$ classification yields quiescent fractions consistent with other studies using COSMOS2020 [c.f. Fig. 9 of 217]. These quiescent fractions are lower than those determined from measurements using UltraVISTA and 3D-HST which range from 25% up to 70% across the same range of $\log(M_\star/M_\odot)$ [e.g., Fig. 2 of 218].

4.4 Predicting SPHEREx photometry

We now use the empirical model detailed above to generate synthetic SPHEREx spectrophotometry with realistic colors. In this section we describe the unique spectral scan strategy employed by SPHEREx and detail the noise properties of the full-sky and deep surveys.

SPHEREx

SPHEREx uses six HAWAII-2RG (H2RG) detector arrays arranged in two mosaics, separated by a dichroic beam splitter that allows the focal plane to be simultaneously imaged [details on the instrument configuration can be found in 219]. A set of linear variable filters (LVFs) sit above the focal planes, which function as bandpass filters whose central wavelength varies linearly with detector position. The spectrum for a source can thus be obtained by modulating its position across the field of view in a series of exposures.

Over its nominal two-year mission, SPHEREx will complete four full-sky surveys, where each survey comprises measurements in 102 spectral channels on each sky position. Observations are shifted by half a spectral channel between the first/third and second/fourth surveys to Nyquist sample the response function. Each spectral channel is defined in steps of $\Delta\lambda = \lambda/R$ across each of the six detectors (detectors in this context are also referred to as "bands"). The resolution of each LVF is fixed between $R = 35 - 130$, where

$$R = \frac{\lambda_c T(\lambda_c)}{\int T(\lambda) d\lambda}, \tag{4.8}$$

λ_c is the central wavelength and $T(\lambda)$ is the filter transmission which varies continuously as a function of detector position. This will be done through a scan strategy that involves a combination of large and small slews as the spacecraft follows a low-earth sun-synchronous orbit, observing near great circles which precess over six months. As a result of this observing strategy, SPHEREx will scan the northern and southern ecliptic caps (NEP and SEP, respectively) with much higher cadence, leading to an total area of 200 deg^2 with $\sim$ 200 measurements per spectral channel after two years. The two survey depths are distinguished as "shallow" (or "full-sky") and "deep" throughout this work.

Figure 4.13 shows coverage maps of both NEP and SEP for the simulated deep field survey. The presence of the Small Magellanic Cloud near the SEP motivates an avoidance strategy for SPHEREx which leads to slightly shallower coverage compared with NEP, along with some asymmetric structure. In these regions SPHEREx obtains considerably more measurements than of the full sky. However in the deep field regions the number of complete spectra (defined as having one measurement per channel) varies considerably as a function of ecliptic latitude, from roughly fifty complete spectra per line of sight in the outskirts of the field up to over four hundred spectra ($\sim$ 40000 dithered measurements per source) in the deepest parts of the NEP field. This will yield galaxy spectro-images that are oversampled both in the spatial and spectral domains.

We compute observed fluxes by integrating the set of noiseless SEDs over our nominal set of 102 SPHEREx channel bandpasses. Once this is done we then add observational noise consistent with SPHEREx's expected sensitivity. While the Zodiacal light (scattered sunlight and thermal emission by interplanetary dust grains; ZL) varies in intensity as a function of celestial position and time, for simplicity we use the conservative maximum expected value (MEV) estimates for

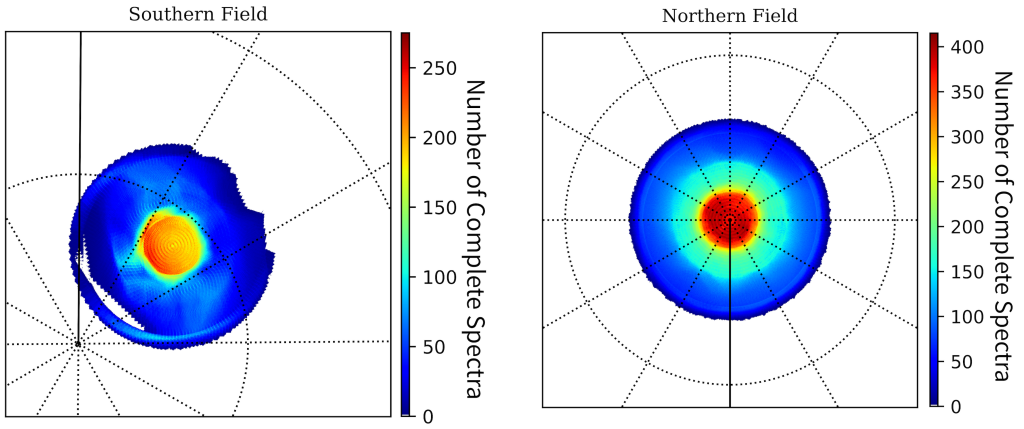

Figure 4.13: Coverage maps for the SPHEREx deep fields, located near the south (left) and north (right) ecliptic poles. The southern field is slightly offset from the SEP in order to avoid contamination from the Small Magellanic Cloud (SMC). Each of the two coverage maps shown covers $\sim$ 100 deg^2, with the central red core of each map spanning a diameter of 3.5 degrees.

point source sensitivity in each channel. These estimates assume a sky-averaged ZL surface brightness level informed by measurements from DIRBE [220]. Many of the galaxies at full-sky sensitivity will have fluxes that are photon noise-limited (or limited by confusion noise), however we also add Poisson noise which primarily impacts the brightest galaxies. We plot the SPHEREx point source flux sensitivities used throughout this work in Fig. 4.14. At full-sky depth, the average MEV 5σ per channel point source depths varies from 19.3 at 0.75 μm to 19.7 at 3.8 μm (bands 1-4), with reduced sensitivity for $\lambda > 3.8$ μm (bands 5 and 6). The deep fields will push roughly two magnitudes deeper in point source sensitivity, with a dependence on celestial position (see Fig. 4.13).

The synthetic COSMOS catalog used in this work goes several magnitudes deeper than the SPHEREx full-sky sensitivity. This means the catalog can be used to make predictions for the majority of sources for which SPHEREx can measure redshifts. The catalog can also be used to simulate realistic distributions of fainter sources that contribute in the form of confusion noise, however we do not simulate source confusion in this work.

Figure 4.15 highlights two sample galaxies from the COSMOS2020 sample with simulated photometry; in the upper panel is a massive, quiescent galaxy at low redshift ($z = 0.25$), while the lower panel shows an ELG observed at $z = 1.14$. Like in the example shown, many quiescent galaxies have well-resolved, rest-frame

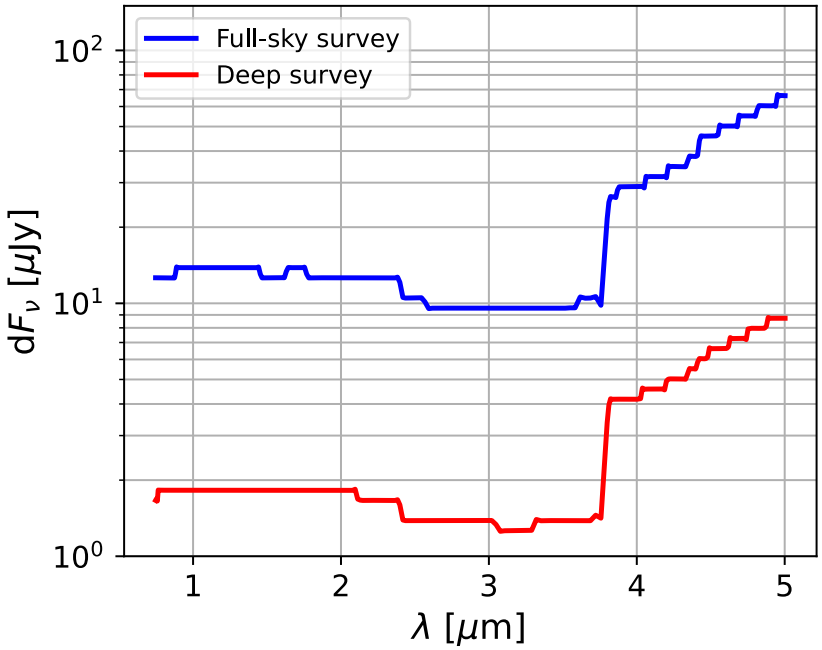

Figure 4.14: 1σ MEV point source sensitivity, per spectral channel. The curves represent the average sensitivity for the SPHEREx full-sky (blue) and deep (red) field depths after the nominal two-year mission. These correspond to per-channel 5σ sensitivities of $m_{AB} = 19.7 - 17.5$ at full-sky depth and $m_{AB} = 22.0 - 19.8$ for the deep survey, with reduced sensitivity at long wavelengths due to its diffraction-limited PSF.

1.6μm bumps driven by a minimum in H^- opacity along with PAH features at longer wavelengths. These continuum features can help derive precise redshifts for luminous red galaxies [221, 222]. In contrast, star-forming galaxies are typically less massive but contain several emission lines/line complexes which are detectable by SPHEREx, namely Hα+[NII], [OIII]+Hβ and Paschen-α.

Comparison with existing/near-future surveys

Ancillary catalog data from surveys in the optical and the infrared are important for defining SPHEREx's "reference catalog", the list of sources SPHEREx will measure the spectra for using forced photometry. The instrumental point spread function (PSF) for SPHEREx varies as a function of wavelength, with a core PSF full width at half maximum (FWHM) ranging from 2″ at short wavelengths to a diffraction-limited 7″ at longer wavelengths. This does not include additional spread from pointing jitter, however this is expected to be controlled at the $\sigma_{RMS} \lesssim 1''$ level. The SPHEREx PSF will in general be undersampled due to the relatively large 6.2″

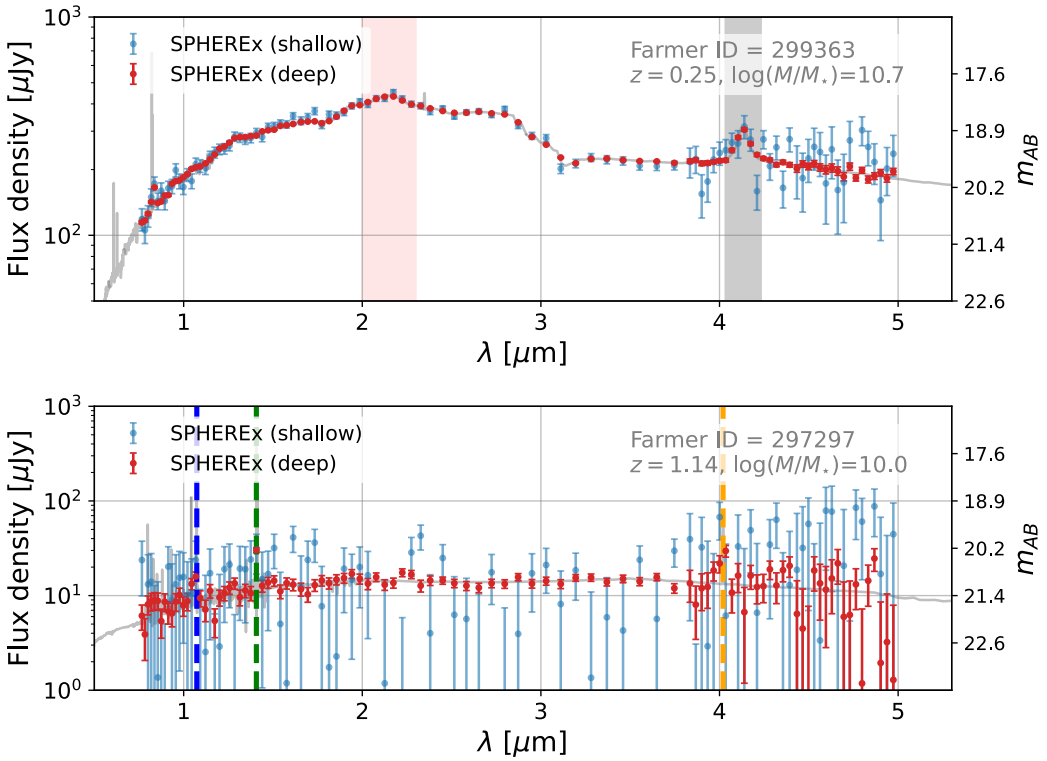

Figure 4.15: Mock photometry of two galaxies: a massive, quiescent galaxy at $z = 0.25$ (top) and an active emission line galaxy at $z = 1.14$ (bottom). Simulated flux measurement uncertainties are based on the current SPHEREx all-sky MEV point source sensitivity for 102 spectral channels. The grey lines show the underlying SEDs generated through our emprirical model, while black points show existing COSMOS photometry of the two sources. Both galaxies have photometry simulated at full-sky (blue) and deep (red) survey depths. The spectrum in the top panel shows signatures for both the rest-frame 1.6 μm bump [red shaded region; 222] and PAH emission in the black shaded region. In the bottom panel, the galaxy is detected at low significance at full sky depth while deep field photometry of the same source is sensitive to Hα+[NII] ($\lambda_{obs} \sim 1.4$ μm, green dashed line), [OIII]+Hβ ($\lambda_{obs} \sim 1.4$ μm, blue) in the rest frame optical and Paschen$-\alpha$ in the near infrared ($\lambda_{obs} \sim 4.0$ μm, orange).

$\times 6.2''$ pixel size, which makes deblending adjacent sources difficult. This motivates the use of forced photometry at the locations of reference catalog sources, and is effective when the positional errors from the reference catalog are small compared to the SPHEREx pixel scale (generally the case for the ancillary catalogs mentioned in this section). [223] demonstrate PSF estimation and photometry in this limit on mock SPHEREx exposures. Redshifts derived from these sources will form the basis for downstream cosmology measurements.

At optical wavelengths, several broad-band photometric surveys have been undertaken over large portions of the sky. The PanSTARRS 3π survey has a nominal 5σ depth of $i = 23.1$ [90]. The DESI Legacy Imaging Surveys combines optical data from three separate surveys (the Beijing-Arizona Sky Survey (BASS), the Dark Energy Camera Legacy Survey (DeCALS) and the Mayall z-band Legacy Survey (MzLS)), covering $\sim 14,000$ deg^2 and reaching median grz 5σ depths of $m_{AB} = 24.0$, 23.4 and 22.8, respectively [224]. Looking ahead, the *Rubin* observatory LSST will obtain optical photometry in six bands ($ugrizy$) across 18000 deg^2 in the southern sky [37]. Catalogs from the first data release are expected to reach a 5σ co-added depth of $i \sim 25$ after one year of observations, while after ten years the depth is predicted to be $i \sim 26.8$. These optical catalogs will resolve sources with considerably finer angular resolution than SPHEREx.

In the infrared, catalogs from full sky WISE imaging detect large numbers of extragalactic sources in its two broad bands centered at 3.4 and 4.5 μm (W1 and W2, respectively). The infrared catalog from 5-year WISE co-added images reaches depths of $m_{AB} \sim 20.7$ and 19.9 in W1 and W2 band, respectively [225], which is slightly deeper than the typical SPHEREx single-channel full-sky sensitivity. By the time of SPHEREx's launch WISE will have catalogs derived from eight years of imaging. When we cross-match WISE sources against known galaxy positions in the COSMOS field, we find a positional accuracy ranging from $0.2''$ on the bright end up to $0.5 - 1''$ for $W1 \sim 20 - 21$.

Predictions from synthetic photometry

The set of high resolution model SEDs are convolved with LSST i-band and WISE W1 filters to obtain predicted magnitudes. These can then be compared with the nominal external survey depths to quantify the coverage afforded by the reference catalog for measurable SPHEREx sources. Figure 4.16 shows the distribution of SPHEREx sources and synthetic *Rubin*/WISE counterparts. For $i < 18$ we include

a subset of galaxies from our GAMA catalog matching the effective area of the COSMOS catalog.

SPHEREx will be effective at probing the population of infrared bright, optically faint galaxies. We inspect the population of sources in our synthetic COSMOS catalog with $24 < i < 25$ and SPHEREx $20 < m_{[2.2\mu m]} < 21$, which are measurable at SPHEREx full-sky depths. These are shown in Fig. 4.17. As implied by our selection, these sources are very red ($i - [2.2\ \mu\mathrm{m}] > 3$) and are typically massive ($\log(M_\star/M_\odot) > 10.5$), quiescent (EW(H$\alpha$) < 50 Å) galaxies with redshifts $z \sim 1 - 2$. Higher-redshift populations like this are important for SPHEREx cosmology measurements, which rely on precise measurement of long wavelength modes to constrain local primordial non-Gaussianity [pNG; 39].

From estimated survey depths of the ancillary catalogs, the redshift distribution of sources selected by each catalog can be computed. This is shown in Fig. 4.18 for two cuts $K < 20$ and $K < 22$. For $K < 20$, the optical and infrared external catalogs are complete out to $z \sim 1$. For redshifts $z > 1$, completeness using PanSTARRS and DECaLS falls considerably, while the WISE-selected catalog remains largely complete. For the $K < 22$ cut, incompleteness of the external catalogs is more severe. The union of optical/infrared catalogs (e.g., WISE+DECaLS) complements each individual catalog (low redshifts covered by optical, higher redshifts covered by infrared), however no combination of these catalogs is complete down to SPHEREx's deep field sensitivity across the 200 deg^2 covering the ecliptic poles.

Another consideration is that the source density of the varies significantly between the optical and infrared catalogs that comprise the reference sample. This is seen clearly in Fig. 4.19, in which synthetic LSST i-band and WISE 3.4 μm densities are plotted for a range of magnitudes near each catalog's expected limiting depth. As one approaches $W1 = 21$ in the WISE catalog, the source density approaches 0.06 pixel^{-1}, or roughly one source per 16 SPHEREx pixels. In contrast, the LSST source density for all sources with $i < 25$ is ~ 0.4 pixel^{-1}, i.e., one source per 2.5 SPHEREx pixels. This has implications for strategies that utilize deeper reference catalogs to define SPHEREx targets in forced photometry, and will require some parsimony in the effective number of sources that are fit simultaneously.

Prevalence of detectable emission lines in SPHEREx observations

Quantifying the fraction of galaxies where emission lines are present is important in forecasting redshift constraints. It has been shown using COSMOS 30-band

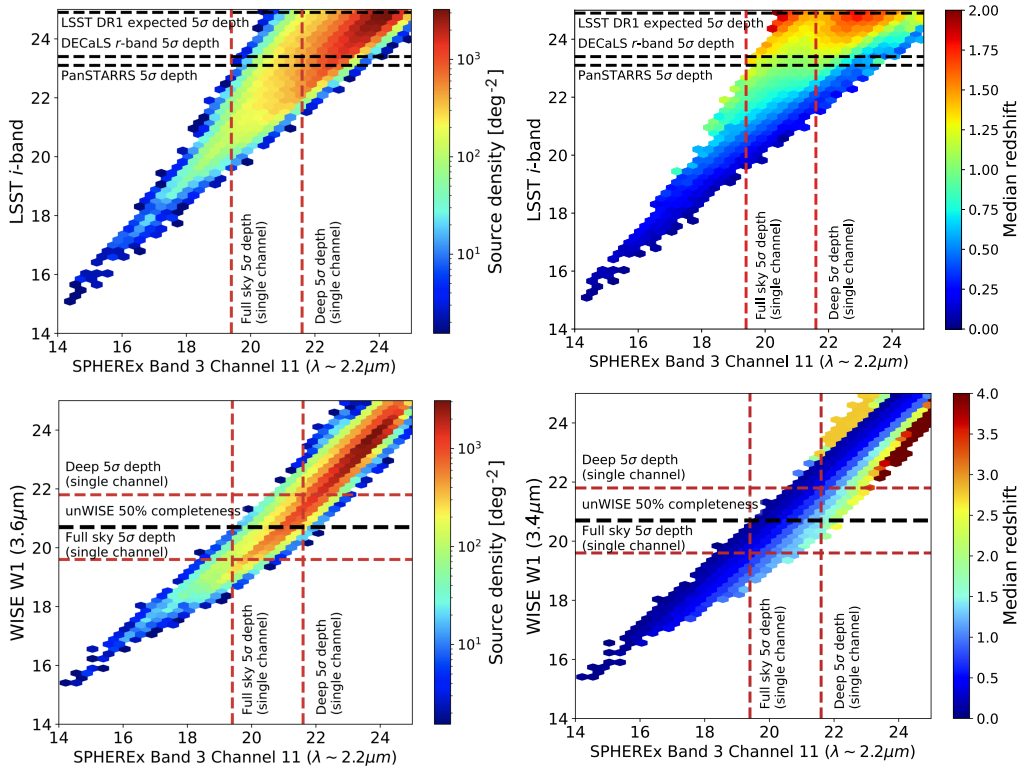

Figure 4.16: Magnitude-magnitude diagrams of simulated COSMOS2020 SEDs, for SPHEREx 2.2 μm magnitudes (x-axis) and LSST i-band (top) and WISE W1 (bottom). Also shown are expected 5σ point source sensitivities for the three photometric surveys. The SPHEREx MEV 5σ single channel depths are indicated by the red dashed lines for 2.2 μm (vertical) and 3.4 μm (horizontal). The SPHEREx deep field depths do not include confusion noise. While WISE, PanSTARRS and DECaLS provide reasonably complete coverage of detectable SPHEREx sources at SPHEREx full-sky depth, deeper catalogs from, e.g., *Rubin* imaging will be required to pre-select SPHEREx sources down to 5σ per-channel sensitivity in the deep fields.

photometry that accounting for emission lines in photo-z measurements treatment can increase redshift accuracy by over a factor of two [171], however it is an open question what impact emission lines will have for the SPHEREx sample. The number of detected lines and their equivalent widths will determine the redshift precision of the ELG sample. In addition, accounting for emission lines, for example Hα, can improve SFR and dust opacity estimates for many galaxies [226].

To compute SPHEREx line flux sensitivity we start by assuming the total flux from an emission line falls into an individual channel, i.e., $\Delta\lambda \gg \sigma_{line}$, where $\Delta\lambda$ is the channel width. For a given flux F_{line} in erg cm^{-2} s^{-1}, the flux density averaged

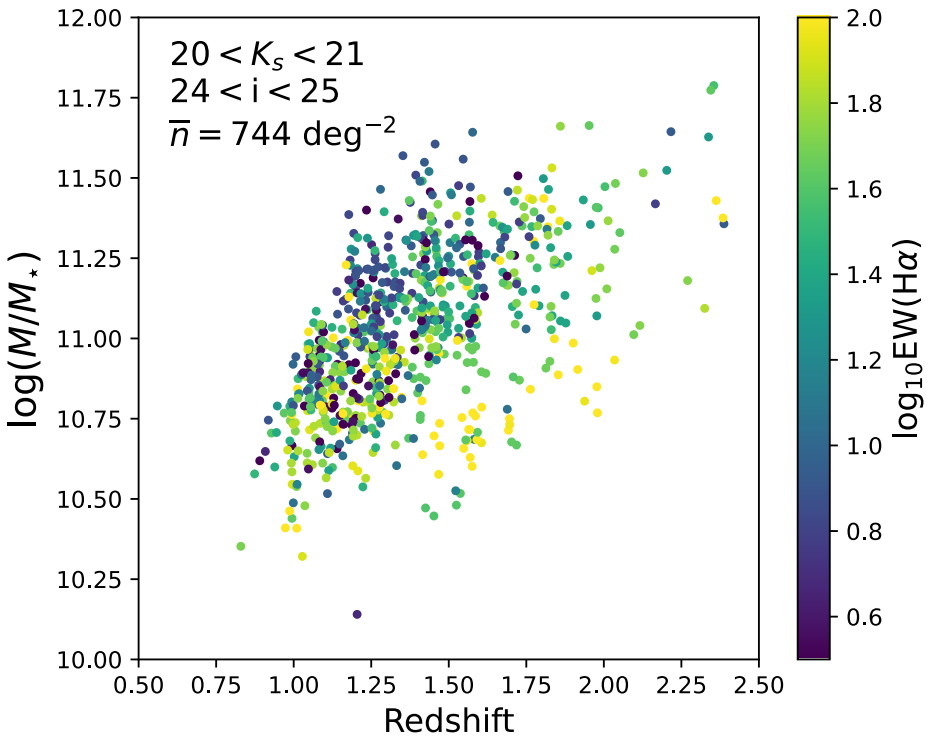

Figure 4.17: Distribution of galaxy redshifts and stellar masses after applying a cut on i- and K-band magnitudes. The points are colored by their Hα (log-) equivalent widths. This red, massive, intermediate redshift ($1 \leq z \leq 2$) galaxy population is an important target for SPHEREx clustering studies that seek to constrain primordial non-Gaussianity.

across channel i is given by

$$\Delta S_{\lambda_i}^{line}[\mu Jy] = \frac{10^{29}\lambda_i^2}{c} \times \frac{F_{line}}{\Delta\lambda_i}, \tag{4.9}$$

where λ_i is the central wavelength of the bandpass in Å and $c = 3.0\times10^{18}$ Å s^{-1}. This expression allows us to compute the significance of detecting a signal from the line in the presence of noise. We compute this at both full-sky and deep survey depths over the full set of 102 nominal channels, assuming $R = [41, 41, 41, 35, 110, 130]$ across the six SPHEREx bands, where each band corresponds to a separate detector [227]. This calculation does not take into account the spectral dithering with which SPHEREx will sample emission lines, nor the details of line-continuum separation, which may introduce additional errors and covariances.

The resulting flux sensitivities are shown in Fig. 4.20. F_{line}^{lim} depends on a combination of the wavelength dependent point source sensitivity and the spectral resolution

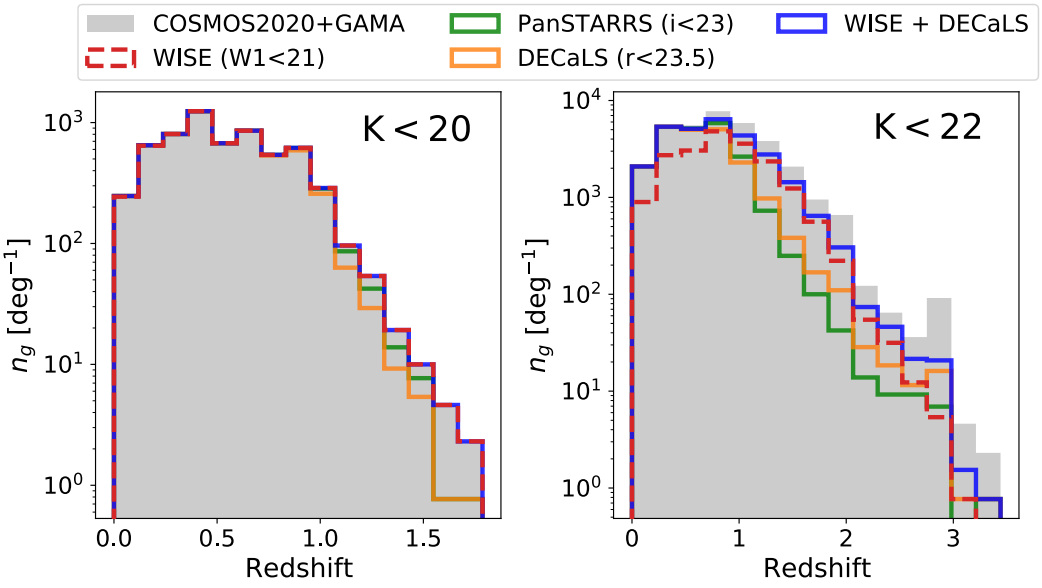

Figure 4.18: Comparison of the redshift distribution of our synthetic COSMOS2020+GAMA catalog (grey), along with that obtained using PanSTARRS (green), DECaLS (orange) and WISE (red) reference catalogs.

across the six LVFs. These sensitivity estimates exclude effects of confusion noise. Despite the larger instrumental noise expected at longer wavelengths, the spectral resolution is > 3 times higher than at short wavelengths, leading to a minimum in line sensitivity around 4 μm that coincides with the minimum in Zodiacal light intensity. At full-sky depth, the 3σ line flux sensitivity ranges from 3×10^{-15} erg cm^{-2} s^{-1} at $\lambda \sim 1$ μm down to a minimum of 6×10^{-16} erg cm^{-2} s^{-1} at $\lambda \sim 4$ μm. At the same reference wavelengths the 3σ deep survey sensitivities range from $4{\times}10^{-16}$ erg cm^{-2} s^{-1} down to $\sim 9 \times 10^{-17}$ erg cm^{-2} s^{-1}. This approaches the line sensitivity expected for the *Euclid* and *Roman* grisms and should complement these surveys through coverage beyond 2 μm.

Using these sensitivity estimates we can predict, given a catalog of emission line fluxes and redshifts, how many sources are detectable by SPHEREx. This is done for Hα+[NII], [OIII]+Hβ, [OII] and Paschen-α. The number of $> 2\sigma$ line detections at both full-sky and deep survey depth are shown as a function of redshift in Fig. 4.21. Combining line detections from both COSMOS2020 and GAMA catalogs (weighted by their effective areas), we predict full-sky number densities of $\overline{n} \sim 770$ (260) deg^{-2} for Hα detectable at 2σ (3σ). The next most prevalent line is Paschen-α, with $\overline{n} \sim 250$ (140) deg^{-2}. Despite the small intrinsic line flux ratio $P\alpha/H\alpha = 0.12$ (for case B recombination), Paschen-α is more immune to dust extinction than

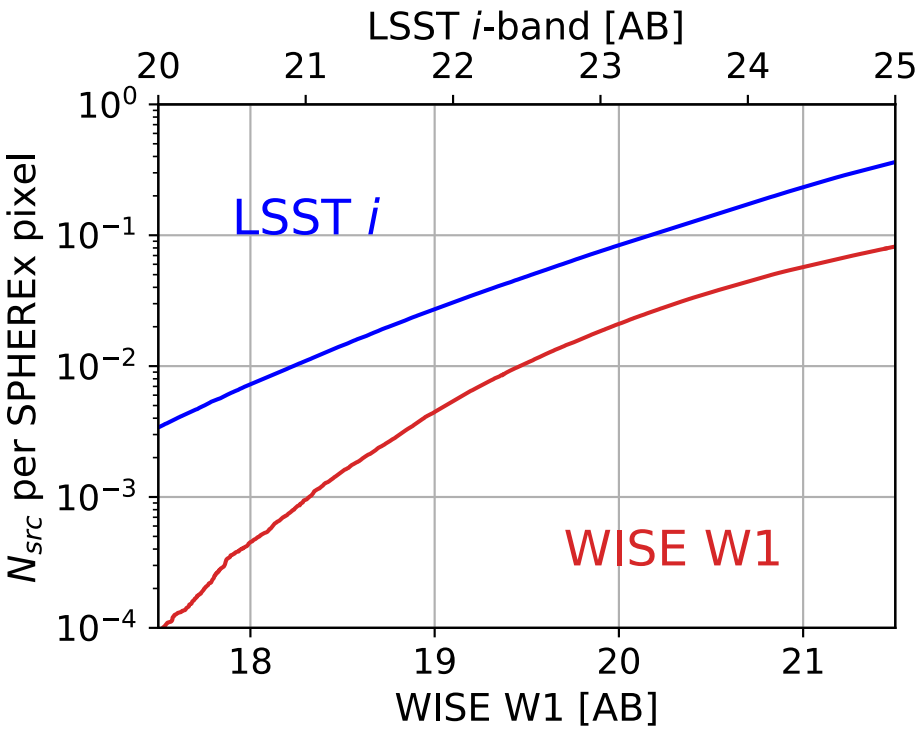

Figure 4.19: Source density as a function of reference catalog depth, selecting on LSST i-band (blue, top axis) and WISE W1 at 3.4 μm (red, bottom axis). These number densities are calculated assuming the COSMOS2020 footprint of 1.27 deg^2 and assuming $6.2'' \times 6.2''$ pixels. While selection on WISE W1-detected sources corresponds to one source per $\sim$ 10 SPHEREx pixels, a blind selection on deeper optical catalogs (with $i_{lim} = 24 - 25$) corresponds to one source per 2.5 SPHEREx pixels, i.e., much further into the confusion-dominated regime. This highlights the importance of reference catalog pre-selections for SPHEREx in the limit where much deeper catalogs with high surface density are available.

Hα and is observed at longer wavelengths where SPHEREx has better line flux sensitivity. SPHEREx is less sensitive to rest-frame optical lines blueward of Hα, due to a combination of intrinsic line ratios, more severe dust attenuation and poorer sensitivity in the blue end of the SPHEREx bandpass ($\lambda < 2\,\mu$m). At full-sky depth, the [OIII]+Hβ complex is detectable at $> 2\sigma$ for 110 COSMOS2020 catalog sources (85 deg^{-2}), however this drops to only 15 sources detected at $> 3\sigma$. The situation is worse for [OII], for which no lines are predicted to be detectable at full-sky depth. [OII] line emission is expected to be primarily detected by SPHEREx in the deep fields, for individual sources or in aggregate through line intensity mapping [e.g., 228].

Detecting multiple lines simultaneously with SPHEREx will enable precise and

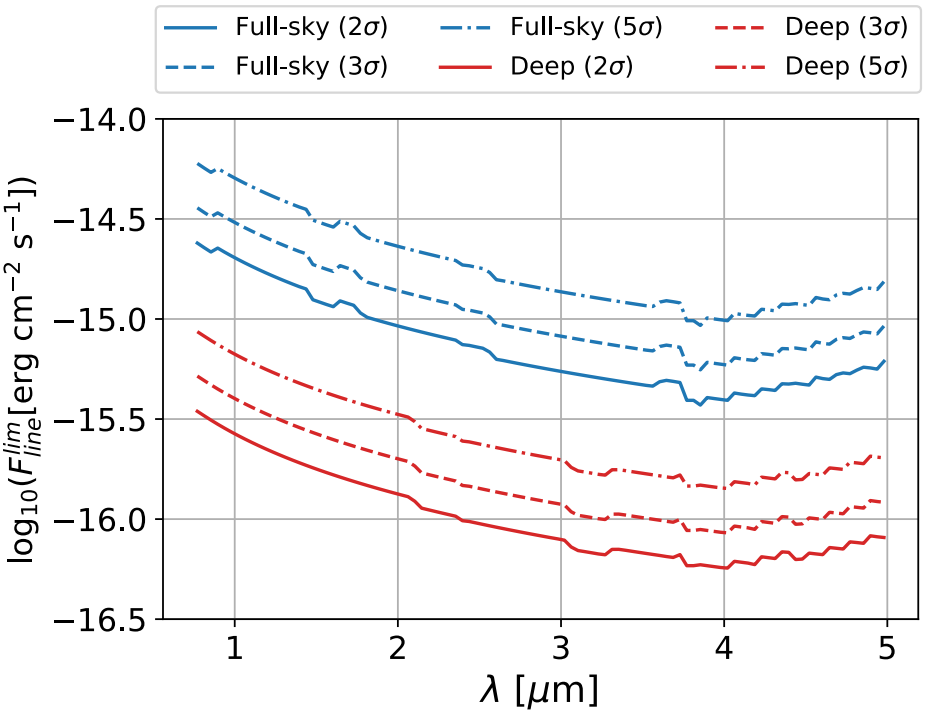

Figure 4.20: Line flux sensitivity as a function of observed wavelength, for SPHEREx full-sky (blue) and deep (red) surveys. The sensitivities are calculated for 2σ (solid), 3σ (dashed) or 5σ (dash-dot) detections, ignoring errors from line-continuum separation and other confusion noise. While SPHEREx's line flux sensitivity at full-sky depth is relatively shallow compared to existing spectroscopic surveys, the deep field sensitivities approach those expected from *Euclid* and *Roman* with complementary, longer-wavelength coverage.

robust redshift measurements across a broad range of distances. At full-sky depth, our synthetic catalog predicts 12% and 15% of Hα lines will have a Paschen-α counterpart in which both lines are detected at $> 2\sigma$ and 3σ, respectively. The fraction of Hα detections with [OIII]+Hβ are 11% and 5% with the same criteria, though most [OIII]+Hβ detections should have a Hα detection (red dashed histogram in top panel of Fig. 4.21). Simultaneous detection of Hα, Paschen-α and [OIII] is rare due to limited redshift overlap.

Our deep field predictions paint a much richer picture for the putative emission line sample near the ecliptic poles, with multiple-line detections extending from $z = 0.17$ to $z = 4$ and beyond. Because SPHEREx will probe the bright end of each line LF (which varies strongly with luminosity), the predicted number densities are highly sensitive to observing depth. Indeed, while the mean sensitivity of the deep survey

is $\sim 10\times$ that of the full-sky, the predicted number densities are larger by factors ranging from $O(10)$ up to $O(1000)$. We summarize the implied ELG number densities at full-sky and deep survey depths in Table 4.2. The number density estimates for the deep fields may be conservative in the sense that our initial cut on $i < 25$ has a larger impact on the deep field forecasts. Due to this and our incomplete knowledge of higher-redshift line populations, we caution over-interpretation of the deep field predictions.

4.5 Redshift recovery

We test redshift recovery using the photometric redshift estimation code implemented in [137], which is similar in spirit to the widely used template fitting code `LePhare` [229, 230]. The code performs a χ^2 minimization across a pre-specified grid of models, which we construct from the same underlying set of 160 templates used in §4.2 to generate our synthetic observations. For each template, we deploy a grid of models with $E(B-V) = 0-1$ in steps of $\delta E(B-V) = 0.05$ for three dust extinction laws (Prevot, Calzetti and Allen) with redshifts spanning $z = 0-3$ with $\delta z = 0.002$. We assume flat priors over these parameters and the set of templates. In §4.5 we test reducing the set of templates used in redshift estimation as a measure of robustness for our results.

While sufficient for broad band photometric redshift measurements, the emission line model implementations of these codes have shortcomings with stronger implications for intermediate spectral resolution SPHEREx measurements. As a result we choose in this work to assess the redshift information from continua and lines separately, which can be combined in a hybrid line-continuum redshift estimation approach that will be the subject of a future publication.

To emulate the selection of SPHEREx target galaxies, we evaluate redshift recovery for galaxies pre-selected using optical and infrared ancillary catalogs. In particular, we select any galaxies detected by DECaLS ($g < 24.0$, $r < 23.4$ or $z < 22.8$) or WISE ($W1 < 20.5$ or $W2 < 20.5$), which constitutes fifty four thousand galaxies of the full 160K simulation catalog. We include synthetic DECaLS grz and WISE $W1/W2$ photometry with representative noise in the fits, adding a 1% noise floor to capture additional photometric errors in the external catalogs.

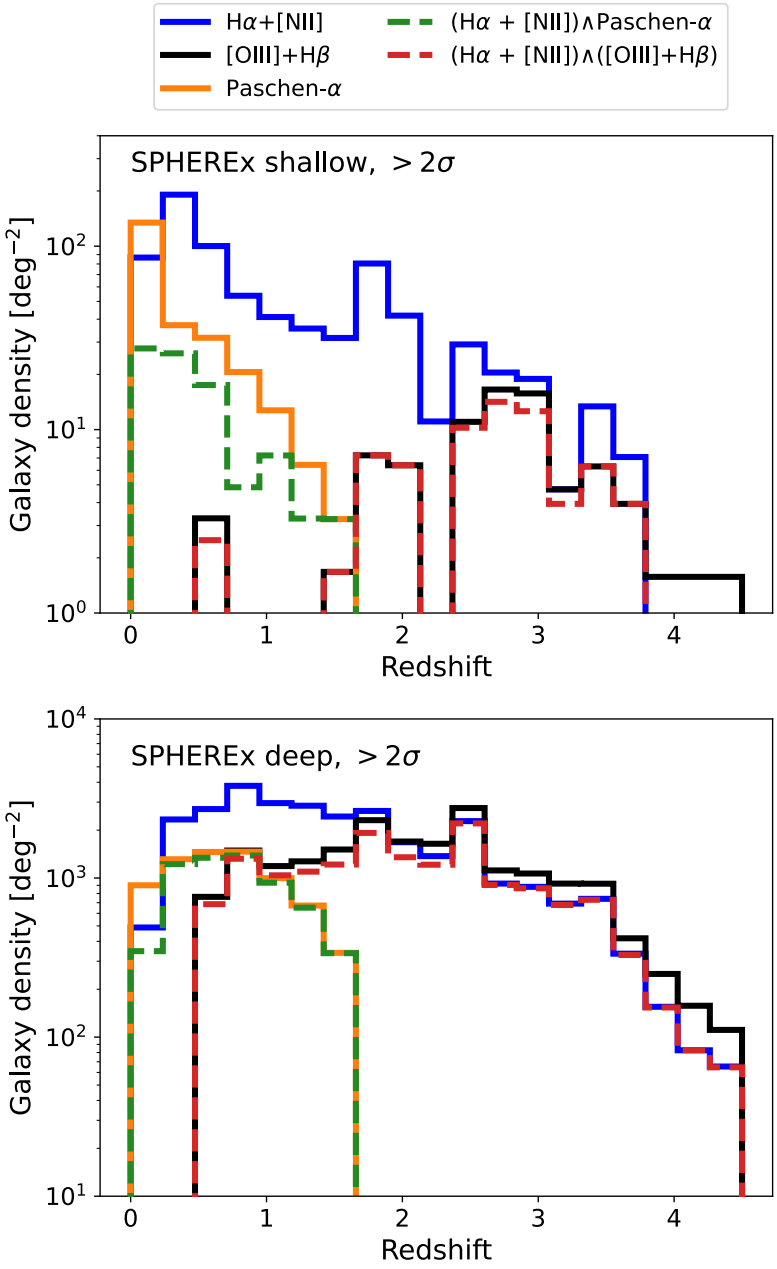

Figure 4.21: Redshift distribution of galaxies with detectable lines (above 2σ significance) at full-sky (top) and deep (bottom) survey depth. This is shown for individual lines/line complexes (blue, green, orange) as well as for simultaneous line detections satisfying $> 2\sigma$ in both lines. At full-sky depth Hα and Paschen-α are predicted to constitute the majority of emission line detections, while in the deep fields [OIII]+Hβ (along with Hα) will be detectable with considerably higher number density out to $z = 3$ and potentially beyond to higher redshifts.

Line(s)	2σ (full sky) [deg^{-2}]	3σ (full sky) [deg^{-2}]	4σ (full sky) [deg^{-2}]	5σ (full sky) [deg^{-2}]	3σ (deep) [deg^{-2}]	5σ (deep) [deg^{-2}]
Hα + [NII]	770	260	115	65	17800	8000
Pα	250	140	100	75	3800	1600
(Hα + [NII]) $\wedge$ Pα	90	40	20	10	3100	1100
[OIII] + Hβ	85	12	< 10	< 10	10500	3600
(Hα + [NII]) $\wedge$ ([OIII] + Hβ)	74	12	< 10	< 10	7900	2600
[OII]	< 1	< 1	< 1	< 1	30	< 10

Table 4.2: Predicted number densities for galaxies with lines detectable through SPHEREx spectrophotometry. Number densities for combinations of lines correspond to sources where both lines/line complexes are detected above the noted significance.

Continuum redshift results

COSMOS2020

We calculate the mean χ^2 statistic of the fits to be 101.7, corresponding to a reduced chi-squared of $\chi^2_{red} \approx 0.99$ assuming four model parameters (redshift, $E(B-V)$, template scale, template index) and 107 total bands (SPHEREx + external). We confirm that the distribution of best fits follows a χ^2 distribution, with few galaxies having $\chi^2_{red} > 1.5$. Given the use of the same galaxy templates used to fit the photometry as used to generate the SEDs, this level of agreement indicates that our fits are well behaved.

We plot a random selection of redshift PDFs with increasing σ_z in Figure 4.22. Each redshift point estimate is computed from the first moment of the redshift $p(z)$, which roughly coincides with the maximum *a posteriori* estimate for unimodal distributions. It can be seen that many of the redshift PDF estimates have non-Gaussian structure, including heavy tails and often more than one local χ^2 minimum. This is to be expected given the complexity of the template set, for which several templates may be degenerate, along with other parameters in the model space. Motivated by this, we include two-sided redshift uncertainty estimates derived from the highest (posterior) density interval (HPDI), defined as the shortest interval on a posterior density for some given confidence level. In some cases where the uncertainty is comparable to the redshift step size ($\sigma_z \sim \delta z = 0.002$, e.g., top left panel of Fig. 4.22), discretization effects may impact the redshift estimates, suggesting that for the highest accuracy sample some refinement of $p(z)$ will be necessary.

Using these estimates we plot redshift error distributions for the COSMOS sample in Fig. 4.23 relative to the true redshifts of the samples. These are binned by reported fractional redshift uncertainties, $\hat{\sigma}_{z/(1+\hat{z})}$, which can be compared with the true errors to assess the fidelity of the redshift estimates. For this comparison we compute redshift uncertainties corresponding to one-half the width of the 68% credible interval (i.e., $\hat{\sigma}_z^{68}$). To quantify redshift errors in each bin, we calculate the normalized median absolute deviation (NMAD), a measure of dispersion that is robust to outliers. We also quantify the outlier fraction η, which is defined as the fraction of 3σ outliers given $\hat{\sigma}_z$ and the true error. These results demonstrate that SPHEREx will measure a wide range of high- and low-accuracy galaxy redshifts. In this test configuration the reported uncertainties closely match the true errors (which can also be evaluated in terms of the z-score distribution, see Appendix .2). The

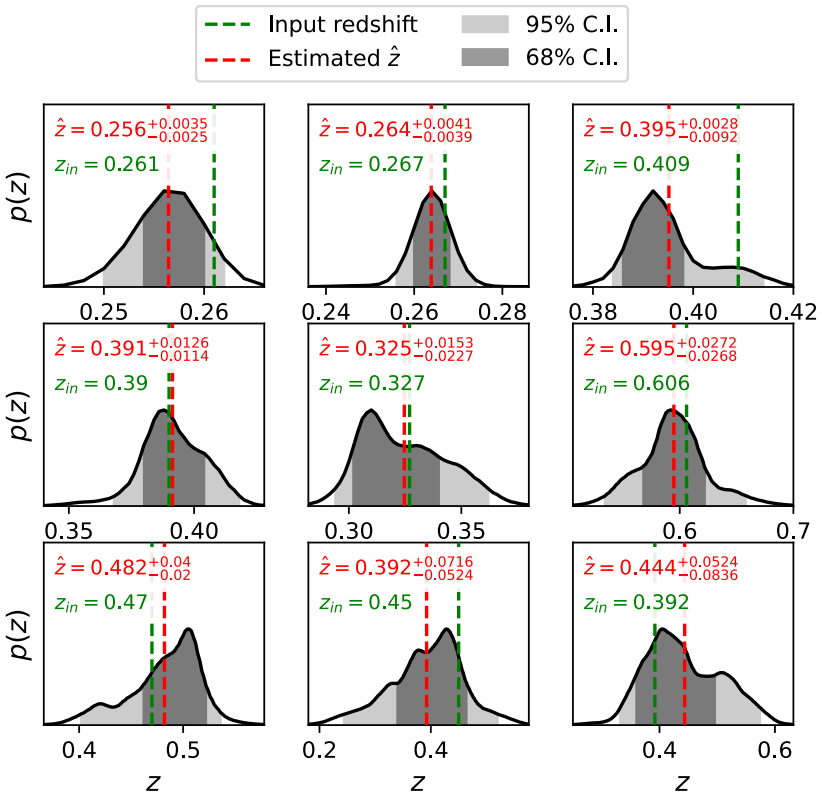

Figure 4.22: Normalized redshift probability density functions (PDFs) for a collection of examples from the COSMOS2020 sample. For each source, the true redshift, the reported redshift estimate and 68%/95% credible intervals are indicated.

3σ outlier fraction remains at the few percent level for all redshift uncertainty bins. We note that there is a mild negative bias that becomes larger for the lower-accuracy samples, which upon further inspection is largely driven by the quiescent galaxy samples. This motivates further investigation into the parameter degeneracies within our template fitting, which may lead to multi-modal redshift solutions for some fits.

To further understand the redshift results we also plot the input and recovered redshifts for the COSMOS2020 sample as a function of $W1$ magnitude in Fig. 4.24. Within each $W1$ bin there is a wider distribution of redshift uncertainties. Nonetheless there is a clear trend between the redshift accuracy and $W1$, along with for the mean bias and outlier fraction. Redshift measurements in the bin $20.5 < W1 < 21.5$, for which long-wavelength SPHEREx data are largely uninformative, show a clear bias toward lower redshift values which warrants further investigation.

By combining the recovered number densities from COSMOS2020 and GAMA (see Appendix .1 for redshift recovery results of the GAMA sample) we can forecast the

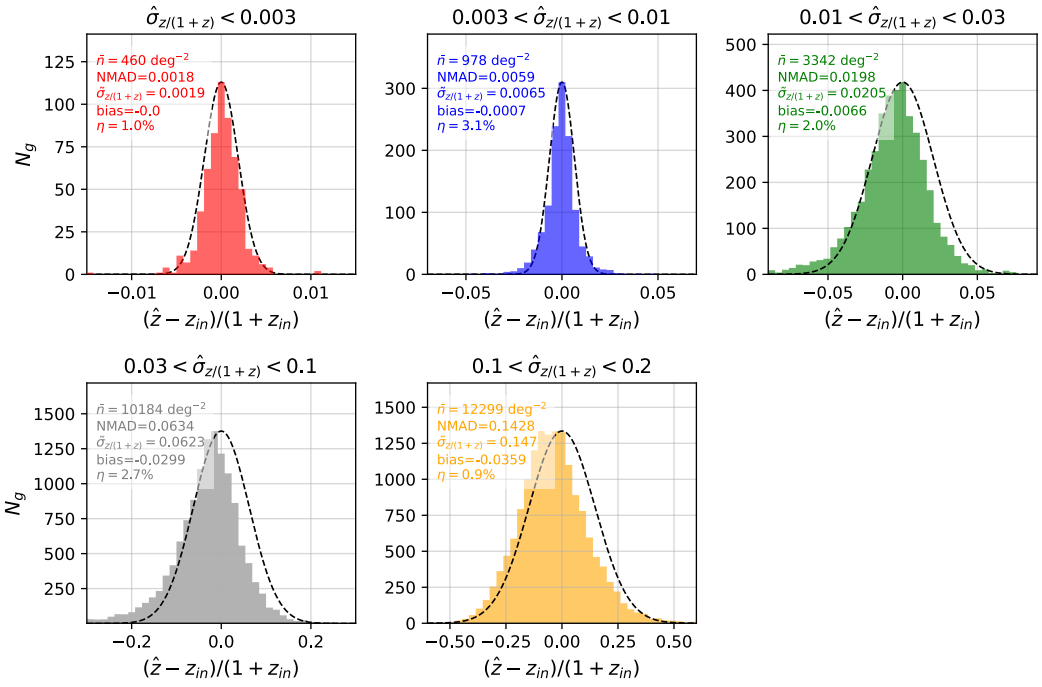

Figure 4.23: Redshift error distributions for COSMOS sources, binned by the reported redshift uncertainty $\hat{\sigma}_{z/(1+z)}$. The dashed Gaussian curves have widths that correspond to the median reported uncertainties within each redshift uncertainty bin, $\tilde{\sigma}_{z/(1+z)}$ and are normalized to the peaks of the histograms. For each sample we report $\tilde{\sigma}_{z/(1+z)}$, the mean redshift bias, the normalized median absolute deviation (NMAD) and the 3σ outlier fraction η. These results do not account for effects of source confusion. While there is a mild redshift bias in the lower accuracy samples which warrants further investigation, these results (combined with those from our supplementary GAMA sample) suggest SPHEREx can acquire high- and low-accuracy redshift samples with comfortable margin relative to the nominal SPHEREx cosmology science requirements.

number of galaxy redshifts accessible to the SPHEREx full-sky cosmology sample. Assuming an effective area of $3{\times}10^4$ deg^2 and removing 3σ outliers, our results imply a sample of 19 million galaxies with $\delta z < 0.003(1{+}z)$, which primarily occupy redshifts $z \lesssim 1$. There are many more intermediate and low-accuracy redshifts; we forecast a sample of 445 million galaxies with $\delta z < 0.1(1 + z)$, and this grows to 810 million galaxies with $\delta z < 0.2(1 + z)$. While the loss of information on f_{NL} from galaxies becomes more significant beyond $\sigma(z) = 0.1(1{+}z)$ [231], the number density of galaxies grows significantly across this range, meaning these low-accuracy galaxies do still provide useful information for pNG constraints. Other studies may also benefit from such low-accuracy samples, for example cross correlations with

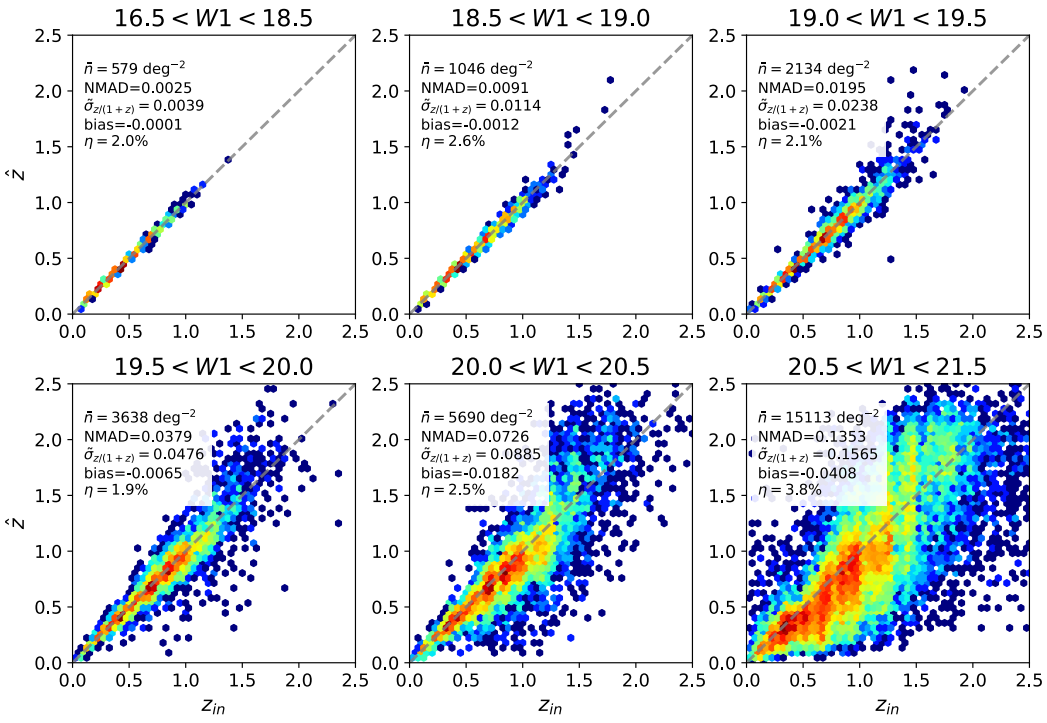

Figure 4.24: Recovered redshifts for synthetic SPHEREx photometry derived from the COSMOS2020 catalog from sources selected in the $W1$ band. For each magnitude bin the galaxy density $\overline{n}$, normalized median absolute deviation (NMAD), median redshift uncertainty $\bar{\sigma}_{z/(1+z)}$, mean bias and 3σ outlier fraction η are included. The colormap corresponds to the (log-) number density of galaxies in each bin.

CMB lensing [232, 233].

Sensitivity to choice of template library

Thus far we have assumed the same set of templates used to fit the redshifts as were used to generate galaxy SEDs. However in practice a reduced set of templates may suffice, both for recovering reliable redshifts and for computational performance. To test the effects of different template sets on the recovered redshifts, we perform similar fits using the empirical Brown templates (129 in total) and the model-based COSMOS templates (31 in total) separately. These results are shown in Figure 4.25. Looking at sources with $\hat{\sigma}_{z/(1+z)} < 0.1$ (black points), both the Brown-only and COSMOS-only template sets perform reasonably well compared to the full template set, albeit with slightly higher outlier fractions and NMAD. Interestingly, the mean bias from using the 31 model-based templates is much smaller than the other two cases. In contrast, the high-accuracy results ($\hat{\sigma}_{z/(1+z)} < 0.01$, red points)

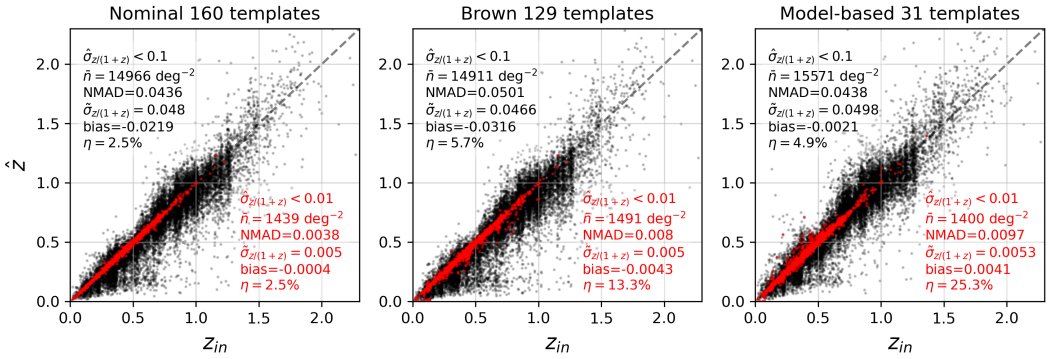

Figure 4.25: Input vs. recovered redshifts on the COSMOS2020 sample, using three different template library sets: the full 160 template set used to generate the photometry (left), the 129 empirical B14 templates (middle) and the 31 model-based templates used in [155] (right). We distinguish high- and low-accuracy (red, black) samples whose redshift recovery statistics are noted as well. These results suggest, perhaps unsurprisingly, that the high-accuracy SPHEREx galaxy sample is more sensitive to template incompleteness in redshift determination than the lower-accuracy sample.

are more sensitive to template coverage. Compared to the full template set results, those using B14 templates alone have a 5× higher outlier fraction, a 10× larger bias, and a NMAD that is much larger than the reported uncertainty. The high-accuracy results degrade further when using just the 31 model-based templates, with a 3σ outlier fraction that rises to 25.3% and a NMAD that is nearly twice as large as the reported uncertainties. These results confirm our intuition that high SNR fits are much more sensitive to template coverage than the lower-accuracy samples.

Redshifts from spectrally dithered emission line measurements

Our line detection and redshift results thus far have utilized photometry in 102 homogenized spectral bandpasses, corresponding to the 102 SPHEREx "channels". However in practice, each observed SPHEREx source will have flux measurements sampled at a unique set of sub-channel positions which roughly Nyquist sample the spectral response function (see §4.4). This additional complexity comes with an opportunity. In this section we demonstrate that by modeling emission lines with the native flux measurements it is possible to go beyond the naive redshift accuracy implied by the per-channel resolution.

To illustrate the potential of SPHEREx's low-resolution spectroscopy we simulate spectrally dithered line flux measurements consistent with the nominal full-sky survey strategy. We focus in this work on the Hα+[NII] complex and Paschen-

α, using line fluxes from our GAMA catalog. For each source we simulate four measurements per channel, and assume the filters are separated at twice the channel resolution. We consider an idealized setup in which the continuum is perfectly subtracted and the continuum measurements constrains the positions of emission lines with a redshift accuracy $\lesssim 10\%$, i.e., the position of each lines are known to within a few SPHEREx channels. Within this range we assume a uniform prior over redshift.

We employ a χ^2 minimization to fit the flux measurements from one or several lines, evaluated over a grid of redshifts using Gaussian line profiles. To model the Hα+[NII] complex we use a fixed prior on the line ratio [NII]/Hα= 0.35, which is informed by the distribution of detectable lines at full-sky depth. At each redshift, we marginalize over the amplitude of the line(s) (denoted $\{A_i\}$) fit to the collection of flux measurements (denoted $\vec{F}$) to obtain the conditional maximum *a posteriori* (MAP) estimate:

$$\{\hat{A}_i\}^{MAP} = \max\left[\ln p(\vec{F}|\{A_i\}) + \ln p(\{A_i\}|z)\right] \tag{4.10}$$

$$\approx \max\left[\ln p(\vec{F}|\{A_i\})\right]. \tag{4.11}$$

We do not impose a prior on the line amplitudes (beyond a fixed Hα/[NII] line ratio), i.e., we do not enforce positive solutions for the line amplitudes. Once the model is evaluated over the pre-determined redshift range we compute the global MAP estimate and 68% credible interval of the 1D redshift PDF $p(z|\vec{F})$.

Figure 4.26 shows the line fitting results for three $z \sim 0.2$ galaxies with varying levels of detectable Hα and Paschen-α emission. The simulated flux measurements use the nominal 102 channel filters with central wavelengths spaced at twice the channel resolution, which approximates Nyquist sampling of the spectral response function. In reality, the observations will have more dispersion in filter locations that depend on the sub-channel (pixel) positions of the sources and the overall survey strategy. The redshift estimates become more precise as the total line SNR increases, with uncertainties ranging from $\sigma_z = 0.01$ for our faint example down to $\sigma_z = 0.0015$ for the brightest example.

Figure 4.27 shows redshift errors plotted as a function of line flux for one- and two-line fits. Our redshift estimates are unbiased with redshift errors that decrease from $\delta z/(1+z) \sim 0.01$ down to sub-percent precision for brighter lines. In each case there is a flux limit below which the redshift errors effectively revert to the original redshift range considered, i.e., the lines are uninformative. These thresholds

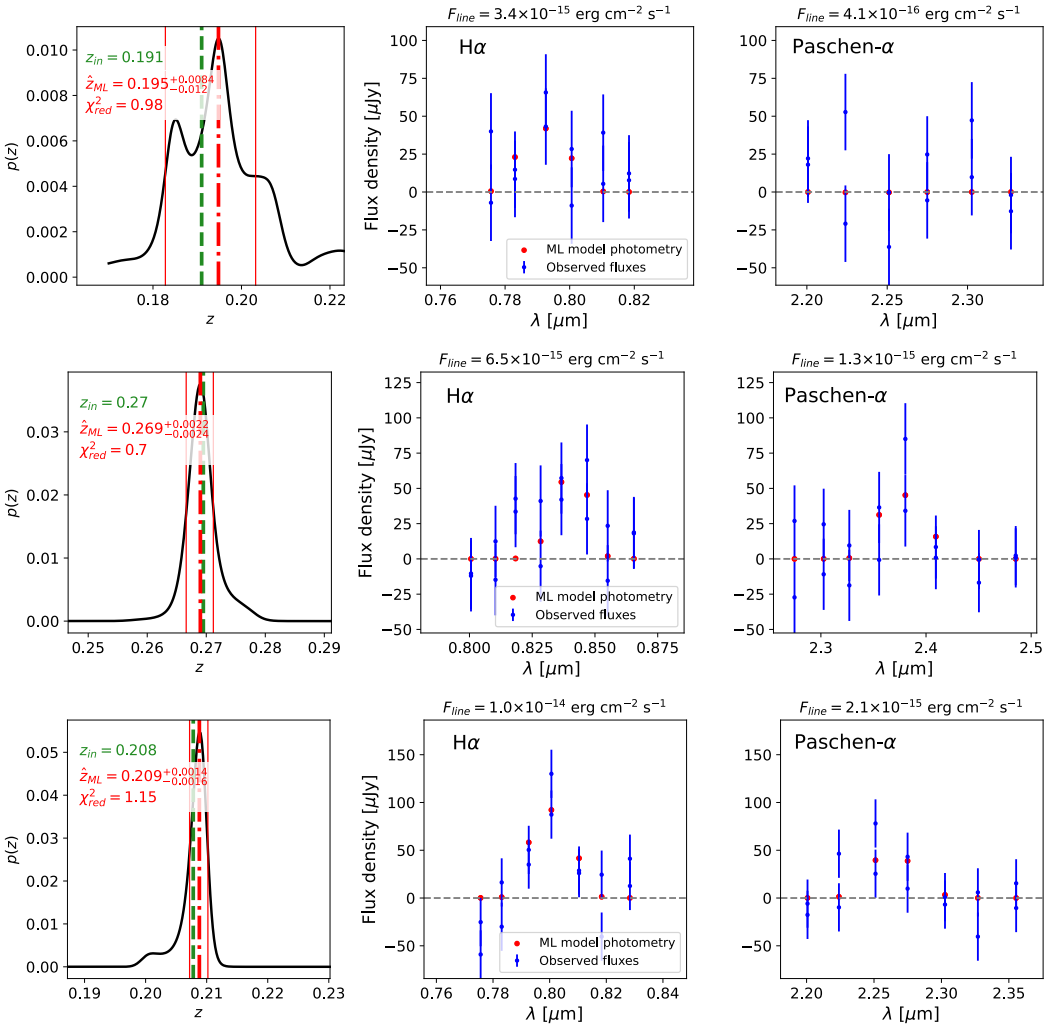

Figure 4.26: Line fits and redshift PDFs for simulated SPHEREx measurements of three isolated, low-redshift galaxies. The left column shows the derived line redshift PDFs – the green and red dashed lines indicate the input and recovered redshifts for each case, while the solid red lines bound the 68% credible interval of each PDF. The middle and right columns show the synthetic fluxes (blue) and best-fit model photometry (red) for measurements near Hα and Paschen-α, respectively.

roughly correspond to the 2σ-3σ line flux sensitivities at 0.8 and 2.0μm. Despite SPHEREx's coarse spectral resolution, in our tests we found that fitting a single Gaussian to the Hα+[NII] complex resulted in a mild bias $\delta z/(1+z) = -0.001$. This motivates the use of line models that account for the full line complex, as done for Hα+[NII] in this work, and will be relevant for other cases such as the [OIII]+Hβ complex.

To prevent spurious effects of overfitting we calculate the improvement in χ^2 from the best line model relative to the null model case (i.e., no lines) as a test statistic to place an example cut on the sources with line fits. Assuming a likelihood ratio Λ and invoking Wilks' theorem with $n_p = 2$ and 3 model parameters for the one- and two-line cases, respectively, the test statistic $-2\ln\Lambda$ should be χ^2-distributed with n_p degrees of freedom. We identify the subset of lines with best fit models above the 95th percentile of their expected χ^2 distributions (plotted in red), while those below are plotted in blue. These criteria are flexible in the sense that we can specify the desired likelihood ratio threshold, however in general the cut is effective at separating line fits with high accuracy from those unconstrained by the photometry.

Table 4.3 summarizes the redshift errors as a function of Hα and Paschen-α line fluxes after making cuts on $\Delta\chi^2$. As seen by eye in Fig. 4.27, the errors decrease monotonically with increasing line flux. The single-line redshift errors for Hα and Paschen-α are of similar size at fixed line SNR. When Hα and Paschen-α are fit together, the dispersion of redshift errors is $\sim 30\%$ smaller than that from fitting Hα alone to the same set of sources. Broadly speaking, these results demonstrate that when one or several lines are detectable and correctly identified, it is possible to recover highly accurate redshifts.

Isolating the emission line measurements from the full SED fits enables redshift estimates that are more robust to the details of the SED model, however this technique relies on some prior estimate for the redshift and thus the line center(s). The quality of the continuum redshift and errors in the continuum model near the lines will primarily impact the purity of the recovered line redshifts. Using continuum-driven line ratio priors can potentially improve redshift estimates for many ELGs, however they may not be appropriate in all cases. This procedure is being considered by the flight software pipeline team to account for as-observed spectrophotometry, however a detailed investigation of the technique and its limits are left to future work. The performance of this approach is expected to improve in the deep fields due to a larger number of flux measurements that increase line flux sensitivity, enable accurate line-

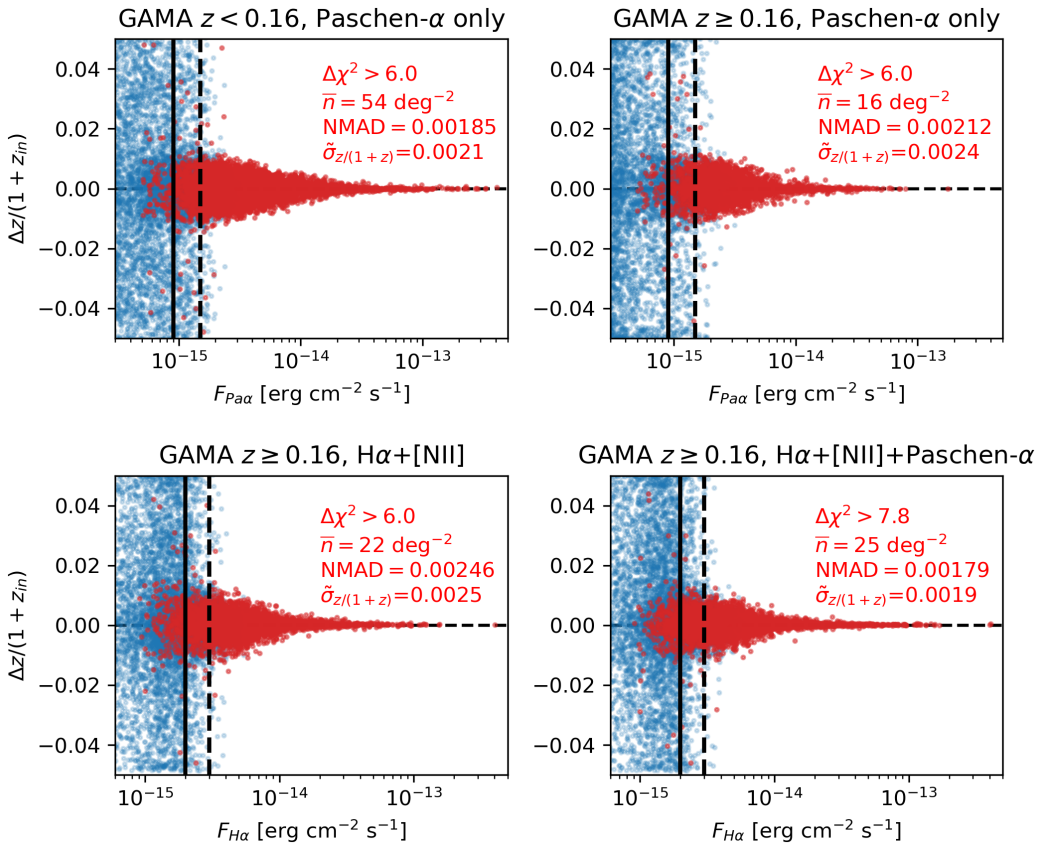

Figure 4.27: Fractional redshift errors as a function of line flux, for different one- and two-line fits. The red and blue points indicate galaxies with and without "high-quality fits", respectively, where high-quality fits have a sufficient improvement in the delta log-likelihood compared to a model with no lines ($\Delta\chi^2 > 6$ for single line (complex) fits, $\Delta\chi^2 > 7.8$ for two lines). The mean redshift uncertainty and NMAD are calculated for each high-quality fit sample. The solid and dashed black lines indicate the 2σ and 3σ line flux sensitivities from Fig. 4.20, respectively, at observed wavelengths of 0.8 μm for Hα and 2.0 μm for Paschen-α. We assume a fixed ratio [NII]/Hα when fitting the native photometry, though the true line ratios in the simulated lines vary.

$F_{H\alpha}$ [$\times 10^{-15}$ erg cm^{-2} s^{-1}]		[2, 3)	[3, 5)	[5, 7)	[7, 10)	[10, 50)	[50, 100)	[100, 1000)
NMAD ($H\alpha$ only)		0.0048	0.0041	0.0028	0.0022	0.0011	0.0003	0.0001
NMAD ($H\alpha$+Pα)		0.0049	0.0038	0.0023	0.0014	0.0005	0.0001	< 0.0001
$F_{P\alpha}$ [$\times 10^{-15}$ erg cm^{-2} s^{-1}]	[0.8, 2)	[2, 3)	[3, 5)	[5, 7)	[7, 10)	[10, 50)	[50, 100)	[100, 1000)
NMAD (Pα only)	0.0038	0.0027	0.0018	0.0012	0.0008	0.0004	0.0001	0.0001
NMAD ($H\alpha$+Pα)	0.0020	0.0014	0.0010	0.0007	0.0006	0.0003	0.0001	< 0.0001

Table 4.3: Normalized median absolute deviation (NMAD) of redshift errors $(\hat{z} - z_{in})/(1 + \hat{z})$ for the GAMA sample, separated as a function of $H\alpha$ and Pα line fluxes. These galaxies come from our quality cuts based on the line fits ($\Delta\chi^2 > 6$ for single lines, $\Delta\chi^2 > 7.8$ for two lines).

continuum separation and multi-line detections, and cover a more uniform sampling of each source's effective channel response[1].

Redshift validation

While our use of synthetic data allows us to directly quantify redshift errors, spectroscopic data from existing surveys will be used in practice to validate SPHEREx redshifts. This approach has been used to perform redshift validation for existing surveys including the Dark Energy Survey [DES; 234] and near-future surveys conducted with *Euclid* [235]. Such a procedure relies on a set of independent, high-resolution spectroscopic measurements that span the spectro-photometric color space occupied by SPHEREx galaxies, and may motivate targeted spectroscopic surveys such as the Complete Calibration of the Color-Redshift Relation survey [C3R2; 236] to fill in observational gaps.

4.6 Conclusion

In this work we have presented a set of galaxy SEDs that combine multi-wavelength fits to existing photometry with an emission line prescription rooted in empirical scaling relations. After validating that the line model is consistent with a number of existing measurements, we generate mock photometry from the synthetic SEDs observed across the SPHEREx bandpass using existing sensitivity estimates. We then demonstrate that precise, accurate redshifts can be obtained using continuum and line information from simulated full-sky depth photometry. These redshift simulations will form the basis for downstream clustering measurements through the power spectrum and bispectrum.

Due to the SPHEREx survey strategy, the observing depth in the NEP and SEP (100 deg^2 each) will be considerably deeper than that of the full sky survey, with measurements along each line of sight that are both spatially and spectrally dithered, oversampling the SPHEREx response function. Given the dearth of existing NIR spectroscopic measurements, galaxy spectra from the deep fields will be useful for refining the galaxy templates used for the full sky survey (the results of this work assume by construction that the templates are representative of the SPHEREx photometry). However, obtaining reliable spectra will rely on a proper treatment of source confusion, which will be much more pronounced. This is not addressed in

[1]The dithered measurements in each channel will in practice come from several independent transmission curves, so this is qualitatively different from oversampling a single spectral response curve.

this work as we do not directly perform forced photometry on mock observations. A dedicated study of deep field photometry and implications for calibrating the full sky survey are left to future work.

We show (under idealized conditions) that by directly fitting the spectrally dithered flux measurements and assuming a mild redshift prior derived from the galaxy continuum, it is possible to recover reliable redshifts with accuracy approaching that of higher-resolution spectroscopic surveys. More work is needed to implement this technique for more realistic cases that incorporate errors in line-continuum separation, errors due to line interlopers and confusion noise from sources along the same line of sight. Despite these additional complexities, it should be possible to obtain accurate redshifts when lines are detected at moderate significance. These results may also be improved in the limit of more dithered measurements (as in the deep fields).

The synthetic catalogs from this work do not include other objects such as stars and active galactic nuclei (AGN). As we only simulate galaxies in this work, our redshift predictions assume that source classes are properly separated. A similar empirical approach to this work may be used with a set of star and/or AGN templates [e.g., 237] to generate synthetic SEDs. While not explored in this work, SPHEREx has an advantage for star-galaxy-AGN separation because of its broad spectral coverage and intermediate spectral resolution, however external information may be needed for certain cases.

Recent studies have shown that the assumption of universality in the halo mass function is a poor description for relating the pNG bias b_ϕ to linear galaxy biases b_1 [238]. However, by relaxing the universality assumption and choosing informed galaxy sub-samples, it may be possible to improve constraints on f_{NL} beyond those currently assumed [239, 240]. The diversity of SPHEREx galaxy types provides opportunities to isolate samples with different linear biases b_1 and non-Gaussian biases b_ϕ for multi-tracer analyses using the power spectrum and potentially higher-order statistics such as the bispectrum.

Our line model comparisons with measurements in the literature are affected by uncertainties in ELG selection, flux measurements and incompleteness corrections that preclude more stringent quantitative assessments of consistency. However, tests at the distribution level (e.g., using the K-S statistic) are powerful for establishing internal consistency (or the lack thereof), for example when comparing detected ELGs at varying SPHEREx survey depth, or when comparing with larger, near-

future spectroscopic samples.

We release our full synthetic COSMOS and GAMA catalogs to the public (these along with high-resolution SEDs are available upon request from the corresponding author). While these simulations will form the basis for tests of the SPHEREx redshift pipeline, the synthetic line catalogs and mock spectra should also be useful for surveys beyond SPHEREx. Our modeling framework is implemented in the tool `CLIPonSS`. The tool is flexible and can be tailored to other use cases which require consistent modeling of line features and continuum properties. Forecasts on emission lines will also become more refined with observations from upcoming spectroscopic surveys such as DESI and PFS and at higher redshift by JWST. One avenue for future work involves characterizing the burstiness of star formation in lower mass galaxy populations using variations in line luminosity and UV continuum luminosity. These observables probe star formation on different time scales and can probe burstiness through individual galaxies and/or direct galaxy+EBL cross-correlations [241, 242].

While more work is needed to match the realism of the full SPHEREx survey and to soon process observed samples, we have demonstrated a redshift procedure which is effective and meets the SPHEREx science requirements with margin, laying the groundwork for measurements of galaxy clustering on both small and large scales.

.1 SED fits of GAMA sources and redshift recovery

In this section we include example SED fits to multiband photometry of GAMA galaxies described in §4.2. Each galaxy in the catalog we use has spectroscopic redshifts and cross-matched photometry from GALEX (FUV/NUV), the VST Kilo-Degree Survey (KiDS, $ugri$), the near-infrared VISTA VIKING survey (ZYJHK_s) and WISE all-sky infrared data (W1 and W2) [164]. A sample of our fits is shown in Fig. .1. The synthetic GAMA catalog represents a bright, low-redshift population for which SPHEREx will measure many precise redshifts. We perform template fitting on synthetic SPHEREx 102-band continuua for ten thousand galaxies of the 44K GAMA sample, also including WISE W1/W2 and DECaLS grz photometry. To account for photometric errors of bright sources we include a 1% uncertainty floor on the external photometry. In addition, given the GAMA sample is comprised of low-redshift galaxies, we deploy a finer grid from $z = 0$ to $z = 1$ with $\delta z = 0.001$. We show the results of this test in Fig. .29. The majority of sources from this sample (85%) have redshift uncertainties $\sigma_{z/(1+z)} < 0.003$, contributing an additional $\bar{n} =$

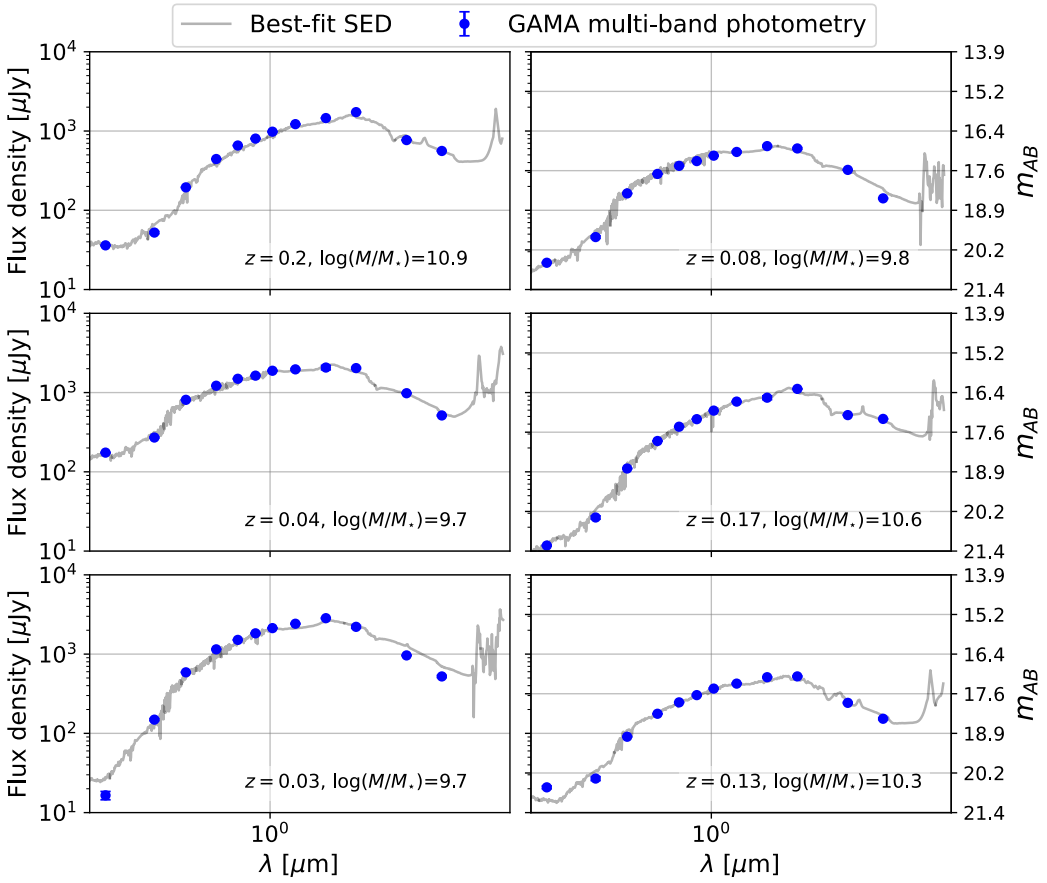

Figure .28: Six example fits to 12-band photometry of GAMA sources with $i < 18$ (blue points), with best fit SEDs shown in grey. These sources will be measured by SPHEREx at high SNR across many channels.

170 deg^{-2} to our high accuracy sample with an 3σ redshift outlier fraction of 0.5%. There is a -0.3σ bias in the sample which warrants further investigation.

.2 Validation of redshift estimates using z-scores

To test the reliability of photometric redshift uncertainties we calculate the z-score distribution for our redshift results. The z-score for an individual estimate is given by $Z = (\hat{z} - z_{in})/\hat{\sigma}_z$ (with capital Z indicating a z-score, not a redshift) and should be unit Gaussian-distributed if the uncertainties statistically match the true errors. In Figure .30 we show the z-score distributions for the nominal COSMOS2020 results in §4.5, grouped in the same redshift uncertainty bins. This distribution is qualitatively similar in shape to the redshift error distributions, with a negative tail of outliers for the medium- and low-accuracy samples. For the first three uncertainty bins, the distribution of z-scores indicates consistency between reported

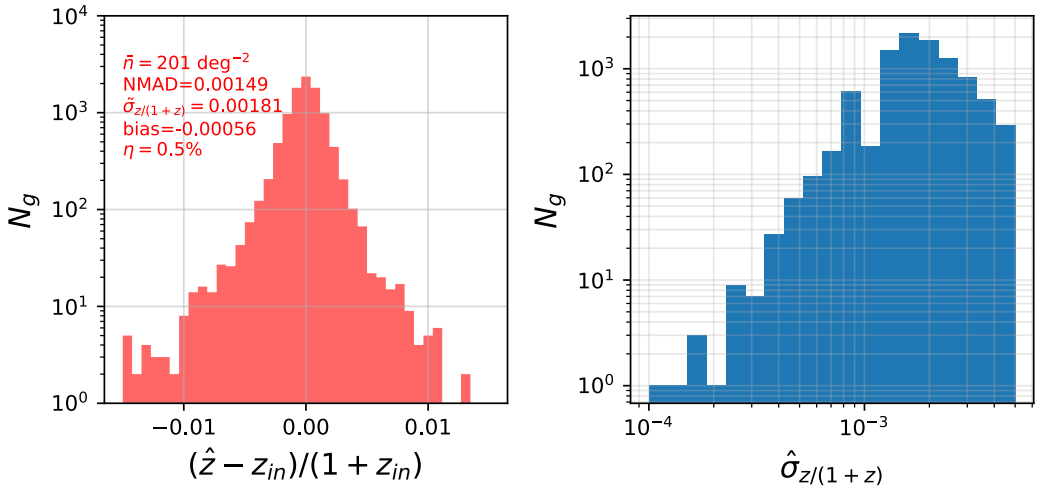

Figure .29: GAMA bright source catalog redshift results. The redshift error distribution is shown on the left. The right panel shows the distribution of reported uncertainties for the same sample. This $i < 18$ sample corresponds to $\sim$ 6 million galaxies over 30000 deg^2 for which SPHEREx will measure high-accuracy redshifts, comprising roughly one-third of the expected SPHEREx high-accuracy redshift sample.

uncertainties and true errors within 5%. For the two lowest accuracy samples, the widths of the z-score distributions suggest that the reported uncertainties are $\sim$ 10% overconfident, though this may also be driven by the larger mean biases and outlier tails. Stricter tests, such as those utilizing the probability integral transform (PIT), can be used to test the reliability of the full $p(z)$ distribution, for example near the tails of the distribution. These metrics will be important for assessing the reliable $p(z)$ information that gets passed downstream to clustering measurements.

References

[36] R. L. et al. In: *arXiv e-prints*, arXiv:1110.3193 (Oct. 2011), arXiv:1110.3193. arXiv: 1110.3193 [astro-ph.CO].

[38] Z. Z. et al. In: *arXiv e-prints* (Sept. 2021). arXiv: 2109.12216 [astro-ph.GA].

[39] N. D. et al. In: *PhRvD* 77.12, 123514 (June 2008), p. 123514. DOI: 10.1103/PhysRevD.77.123514. arXiv: 0710.4560 [astro-ph].

[41] O. D. et al. In: *arXiv e-prints*, arXiv:1412.4872 (Dec. 2014), arXiv:1412.4872. arXiv: 1412.4872 [astro-ph.CO].

[89] M. F. S. et al. In: *AJ* 131.2 (Feb. 2006), pp. 1163–1183. DOI: 10.1086/498708.

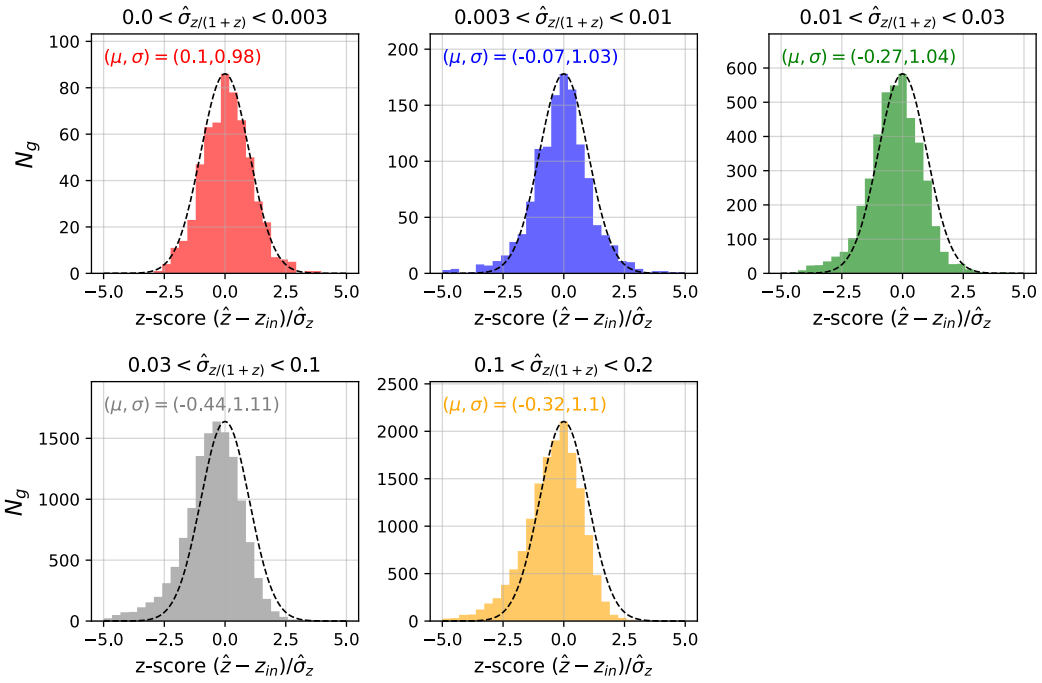

Figure .30: Distribution of z-scores for COSMOS2020 continuum redshift estimates, binned in redshift uncertainty $\hat{\sigma}_{z/(1+z)}$. The black dashed curves indicate a Gaussian with unit variance, i.e., the desired z-score distribution. The (μ, σ) labels indicate the mean and standard deviation of the z-score histograms. With the exception of a mild redshift bias which is largest in the $0.03 < \hat{\sigma}_{z/(1+z)} < 0.1$ bin (-0.44σ), these results suggest the employed template fitting approach recovers reliable redshifts with faithful uncertainty estimates.

[90] K. C. C. et al. In: *arXiv e-prints* (Dec. 2016). arXiv: 1612.05560 [astro-ph.IM].

[134] M. K. et al. In: *ApJS* 218.2, 15 (June 2015), p. 15. DOI: 10.1088/0067-0049/218/2/15. arXiv: 1412.1835 [astro-ph.GA].

[135] H. M. et al. In: *PASJ* 59 (Oct. 2007), S369–S376. DOI: 10.1093/pasj/59.sp2.S369. arXiv: 0708.1796 [astro-ph].

[136] P. J. et al. In: *A&A* 661, A80 (May 2022), A80. DOI: 10.1051/0004-6361/202142663. arXiv: 2202.03305 [astro-ph.IM].

[137] N. R. S. et al. In: *arXiv e-prints* (June 2016). arXiv: 1606.06374 [astro-ph.IM].

[138] N. S. et al. In: *ApJS* 172.1 (Sept. 2007), pp. 1–8. DOI: 10.1086/516585. arXiv: astro-ph/0612305 [astro-ph].

[139] A. R. et al. In: *MNRAS* 500.3 (Jan. 2021), pp. 3254–3274. DOI: 10.1093/mnras/staa3336. arXiv: 2007.09007 [astro-ph.CO].

[140] A. R. et al. In: *arXiv e-prints* (Aug. 2022). arXiv: 2208.08513 [astro-ph.CO].

[141] Y. W. et al. In: *ApJ* 928.1, 1 (Mar. 2022), p. 1. DOI: 10.3847/1538-4357/ac4973. arXiv: 2110.01829 [astro-ph.CO].

[142] J. G. et al. In: *arXiv e-prints* (June 2022). arXiv: 2206.14908 [astro-ph.GA].

[143] M. E. et al. In: *MNRAS* 484.3 (Apr. 2019), pp. 4200–4215. DOI: 10.1093/mnras/stz204. arXiv: 1809.04375 [astro-ph.GA].

[144] A. A. et al. In: *MNRAS* 501.4 (Mar. 2021), pp. 6103–6122. DOI: 10.1093/mnras/staa3659. arXiv: 2007.11132 [astro-ph.GA].

[145] N. B. et al. In: *ApJL* 692.1 (Feb. 2009), pp. L5–L8. DOI: 10.1088/0004-637X/692/1/L5. arXiv: 0812.3568 [astro-ph].

[146] J. P. et al. In: *ApJ* 888.2, 79 (Jan. 2020), p. 79. DOI: 10.3847/1538-4357/ab5f5c. arXiv: 1912.02261 [astro-ph.GA].

[147] N. S. et al. In: *ApJS* 206.1, 3 (May 2013), p. 3. DOI: 10.1088/0067-0049/206/1/3. arXiv: 1303.6689 [astro-ph.CO].

[148] C. L. et al. In: *ApJS* 224.2, 24 (June 2016), p. 24. DOI: 10.3847/0067-0049/224/2/24. arXiv: 1604.02350 [astro-ph.GA].

[149] E. N. K. et al. In: *ApJ* 660.1 (May 2007), pp. 62–71. DOI: 10.1086/513464. arXiv: astro-ph/0701747 [astro-ph].

[150] B. G. et al. In: *A&A* 562, A23 (Feb. 2014), A23. DOI: 10.1051/0004-6361/201322790. arXiv: 1310.1008 [astro-ph.CO].

[151] H. O. et al. In: *PASJ* 68.3, 47 (June 2016), p. 47. DOI: 10.1093/pasj/psw043. arXiv: 1504.05592 [astro-ph.GA].

[152] D. M. et al. In: *A&A* 635, A194 (Mar. 2020), A194. DOI: 10.1051/0004-6361/201937001. arXiv: 2002.00214 [astro-ph.IM].

[153] K. G. et al. In: *ApJ* 923.2, 217 (Dec. 2021), p. 217. DOI: 10.3847/1538-4357/ac2e03. arXiv: 2110.04298 [astro-ph.IM].

[154] M. J. I. B. et al. In: *ApJS* 212.2, 18 (June 2014), p. 18. DOI: 10.1088/0067-0049/212/2/18. arXiv: 1312.3029 [astro-ph.CO].

[155] J. R. W. et al. In: *ApJS* 258.1, 11 (Jan. 2022), p. 11. DOI: 10.3847/1538-4365/ac3078. arXiv: 2110.13923 [astro-ph.GA].

[156] S. P. D. et al. In: *MNRAS* 513.1 (June 2022), pp. 439–467. DOI: 10.1093/mnras/stac472. arXiv: 2203.08539 [astro-ph.GA].

[157] J. B. O. and J. E. G. In: *ApJ* 266 (Mar. 1983), pp. 713–717. DOI: 10.1086/160817.

[158] C. A. T. et al. In: *ApJ* 613.2 (Oct. 2004), pp. 898–913. DOI: 10.1086/423264. arXiv: astro-ph/0405537 [astro-ph].

[159] E. D. et al. In: *ApJ* 670.1 (Nov. 2007), pp. 156–172. DOI: 10.1086/521818. arXiv: 0705.2831 [astro-ph].

[160] K. E. W. et al. In: *ApJL* 754.2, L29 (Aug. 2012), p. L29. DOI: 10.1088/2041-8205/754/2/L29. arXiv: 1205.0547 [astro-ph.CO].

[161] J. S. S. et al. In: *ApJS* 214.2, 15 (Oct. 2014), p. 15. DOI: 10.1088/0067-0049/214/2/15. arXiv: 1405.2041 [astro-ph.GA].

[162] A. J. C. et al. In: *ApJL* 486.1 (Sept. 1997), pp. L11–L14. DOI: 10.1086/310829. arXiv: astro-ph/9706255 [astro-ph].

[163] P. M. and M. D. In: *ARA&A* 52 (Aug. 2014), pp. 415–486. DOI: 10.1146/annurev-astro-081811-125615. arXiv: 1403.0007 [astro-ph.CO].

[164] S. B. et al. In: *MNRAS* 496.3 (Aug. 2020), pp. 3235–3256. DOI: 10.1093/mnras/staa1466. arXiv: 2005.11215 [astro-ph.GA].

[165] J. M. and Jr. K. R. C. In: *ApJS* 164.1 (May 2006), pp. 81–98. DOI: 10.1086/500971. arXiv: astro-ph/0511729 [astro-ph].

[166] J. M. et al. In: *ApJS* 190.2 (Oct. 2010), pp. 233–266. DOI: 10.1088/0067-0049/190/2/233. arXiv: 1007.4547 [astro-ph.CO].

[167] M. W. W. et al. In: *ApJS* 154.1 (Sept. 2004), pp. 1–9. DOI: 10.1086/422992. arXiv: astro-ph/0406223 [astro-ph].

[168] J. R. H. et al. In: *ApJS* 154.1 (Sept. 2004), pp. 18–24. DOI: 10.1086/423134. arXiv: astro-ph/0406167 [astro-ph].

[169] E. d. C., S. C., and D. E. In: *MNRAS* 388.4 (Aug. 2008), pp. 1595–1617. DOI: 10.1111/j.1365-2966.2008.13535.x. arXiv: 0806.1020 [astro-ph].

[170] E. L. W. et al. In: *AJ* 140.6 (Dec. 2010), pp. 1868–1881. DOI: 10.1088/0004-6256/140/6/1868. arXiv: 1008.0031 [astro-ph.IM].

[171] O. I. et al. In: *ApJ* 690.2 (Jan. 2009), pp. 1236–1249. DOI: 10.1088/0004-637X/690/2/1236. arXiv: 0809.2101 [astro-ph].

[172] M. P. et al. In: *ApJ* 663.1 (July 2007), pp. 81–102. DOI: 10.1086/518113. arXiv: astro-ph/0703255 [astro-ph].

[173] G. B. and S. C. In: *MNRAS* 344.4 (Oct. 2003), pp. 1000–1028. DOI: 10.1046/j.1365-8711.2003.06897.x. arXiv: astro-ph/0309134 [astro-ph].

[174] S. S. et al. In: *MNRAS* 494.1 (May 2020), pp. 199–217. DOI: 10.1093/mnras/staa727. arXiv: 2003.06394 [astro-ph.CO].

[175] J. S. S. and D. J. E. In: *MNRAS* 469.1 (July 2017), pp. 1186–1204. DOI: 10.1093/mnras/stw1485.

[176] J. A. B., M. M. P., and R. T. In: *PASP* 93 (Feb. 1981), pp. 5–19. DOI: 10.1086/130766.

[177] Jr. K. R. C. In: *ARA&A* 36 (Jan. 1998), pp. 189–232. DOI: 10.1146/annurev.astro.36.1.189. arXiv: astro-ph/9807187 [astro-ph].

[178] J. B., M. P., and S. C. In: *MNRAS* 385.2 (Apr. 2008), pp. 769–782. DOI: 10.1111/j.1365-2966.2008.12914.x. arXiv: 0801.1678 [astro-ph].

[179] L. J. K. et al. In: *ApJL* 774.1, L10 (Sept. 2013), p. L10. DOI: 10.1088/2041-8205/774/1/L10. arXiv: 1307.0514 [astro-ph.CO].

[180] A. L. F. et al. In: *ApJ* 855.2, 132 (Mar. 2018), p. 132. DOI: 10.3847/1538-4357/aab1fc. arXiv: 1710.00834 [astro-ph.GA].

[181] D. M., A. F., and P. C. In: *ApJ* 828.1, 18 (Sept. 2016), p. 18. DOI: 10.3847/0004-637X/828/1/18. arXiv: 1605.04314 [astro-ph.GA].

[182] L. J. K., M. J. G., and R. A. J. In: *AJ* 127.4 (Apr. 2004), pp. 2002–2030. DOI: 10.1086/382723. arXiv: astro-ph/0401172 [astro-ph].

[183] M. L. P. et al. In: *A&A* 132 (Mar. 1984), pp. 389–392.

[184] D. C. et al. In: *ApJ* 533.2 (Apr. 2000), pp. 682–695. DOI: 10.1086/308692. arXiv: astro-ph/9911459 [astro-ph].

[185] M. J. S. In: *MNRAS* 187 (June 1979), pp. 785–795. DOI: 10.1093/mnras/187.4.785.

[186] D. A. A. In: *MNRAS* 174 (Feb. 1976), 29P–33. DOI: 10.1093/mnras/174.1.29P.

[187] E. L. F. and D. M. In: *ApJ* 307 (Aug. 1986), p. 286. DOI: 10.1086/164415.

[188] A. P. et al. In: *A&A* 586, A83 (Feb. 2016), A83. DOI: 10.1051/0004-6361/201526782. arXiv: 1507.00005 [astro-ph.GA].

[189] D. C., A. L. K., and T. S. B. In: *ApJ* 429 (July 1994), p. 582. DOI: 10.1086/174346.

[190] D. I. V. et al. In: *A&A* 631, A82 (Nov. 2019), A82. DOI: 10.1051/0004-6361/201936232. arXiv: 1907.02111 [astro-ph.GA].

[191] N. A. R. et al. In: *ApJ* 806.2, 259 (June 2015), p. 259. DOI: 10.1088/0004-637X/806/2/259. arXiv: 1504.02782 [astro-ph.GA].

[192] D. K. et al. In: *ApJS* 241.1, 10 (Mar. 2019), p. 10. DOI: 10.3847/1538-4365/ab06c4. arXiv: 1812.01529 [astro-ph.GA].

[193] A. L. F. et al. In: *ApJ* 884.2, 133 (Oct. 2019), p. 133. DOI: 10.3847/1538-4357/ab425b. arXiv: 1909.03076 [astro-ph.GA].

[194] M. F. et al. In: *ApJL* 757.2, L22 (Oct. 2012), p. L22. DOI: 10.1088/2041-8205/757/2/L22. arXiv: 1206.2645 [astro-ph.CO].

[195] O. L. et al. In: *A&A* 439.3 (Sept. 2005), pp. 845–862. DOI: 10.1051/0004-6361:20041960. arXiv: astro-ph/0409133 [astro-ph].

[196] D. K. E. et al. In: *ApJ* 644.2 (June 2006), pp. 813–828. DOI: 10.1086/503623. arXiv: astro-ph/0602473 [astro-ph].

[197] A. A. K. et al. In: *MNRAS* 503.4 (May 2021), pp. 5115–5133. DOI: 10.1093/mnras/stab778. arXiv: 2103.10959 [astro-ph.GA].

[198] S. J. et al. In: *ApJL* 619.2 (Feb. 2005), pp. L135–L138. DOI: 10.1086/427937. arXiv: astro-ph/0411775 [astro-ph].

[199] X. Z. Z. et al. In: *ApJL* 661.1 (May 2007), pp. L41–L44. DOI: 10.1086/518690. arXiv: astro-ph/0702208 [astro-ph].

[200] D. S. et al. In: *MNRAS* 428.2 (Jan. 2013), pp. 1128–1146. DOI: 10.1093/mnras/sts096. arXiv: 1202.3436 [astro-ph.CO].

[201] D. S. et al. In: *MNRAS* 420.3 (Mar. 2012), pp. 1926–1945. DOI: 10.1111/j.1365-2966.2011.19977.x. arXiv: 1109.1830 [astro-ph.CO].

[202] J. C. et al. In: *A&A* 575, A40 (Mar. 2015), A40. DOI: 10.1051/0004-6361/201424767. arXiv: 1408.1523 [astro-ph.GA].

[203] R. C. et al. In: *ApJ* 769.1, 83 (May 2013), p. 83. DOI: 10.1088/0004-637X/769/1/83. arXiv: 1304.5537 [astro-ph.CO].

[204] G. Z., J. M., and M. R. B. In: *ApJ* 701.1 (Aug. 2009), pp. 86–93. DOI: 10.1088/0004-637X/701/1/86. arXiv: 0811.3035 [astro-ph].

[205] A. A. K. et al. In: *Monthly Notices of the Royal Astronomical Society* 452.4 (Aug. 2015), pp. 3948–3968. ISSN: 0035-8711. DOI: 10.1093/mnras/stv1474. eprint: https://academic.oup.com/mnras/article-pdf/452/4/3948/18234696/stv1474.pdf. URL: https://doi.org/10.1093/mnras/stv1474.

[206] W. P. B. et al. In: *ApJ* 920.2, 78 (Oct. 2021), p. 78. DOI: 10.3847/1538-4357/ac1a0e. arXiv: 2108.02222 [astro-ph.GA].

[207] J. W. C. et al. In: *ApJ* 779.1, 34 (Dec. 2013), p. 34. DOI: 10.1088/0004-637X/779/1/34. arXiv: 1305.1399 [astro-ph.CO].

[208] G. N. et al. In: *arXiv e-prints* (Nov. 2022). arXiv: 2211.16650 [astro-ph.GA].

[209] S. J. L. et al. In: *ApJS* 172.1 (Sept. 2007), pp. 70–85. DOI: 10.1086/516589. arXiv: astro-ph/0612291 [astro-ph].

[210] G. B. B. et al. In: *ApJS* 200.2, 13 (June 2012), p. 13. DOI: 10.1088/0067-0049/200/2/13. arXiv: 1204.2829 [astro-ph.CO].

[211] I. G. M. et al. In: *ApJS* 225.2, 27 (Aug. 2016), p. 27. DOI: 10.3847/0067-0049/225/2/27. arXiv: 1510.02106 [astro-ph.GA].

[212] F. L. et al. In: *A&A* 495.1 (Feb. 2009), pp. 53–72. DOI: 10.1051/0004-6361:200810397. arXiv: 0811.2053 [astro-ph].

[213] C. C. S. et al. In: *The Astrophysical Journal* 795.2 (Oct. 2014), p. 165. DOI: 10.1088/0004-637x/795/2/165. URL: https://doi.org/10.1088/0004-637x/795/2/165.

[214] R. J. W. et al. In: *ApJ* 691.2 (Feb. 2009), pp. 1879–1895. DOI: 10.1088/0004-637X/691/2/1879. arXiv: 0806.0625 [astro-ph].

[215] E. D. et al. In: *ApJ* 617.2 (Dec. 2004), pp. 746–764. DOI: 10.1086/425569. arXiv: astro-ph/0409041 [astro-ph].

[216] O. I. et al. In: *A&A* 556, A55 (Aug. 2013), A55. DOI: 10.1051/0004-6361/201321100. arXiv: 1301.3157 [astro-ph.CO].

[217] J. R. W. et al. In: *arXiv e-prints* (Dec. 2022). DOI: 10.48550/arXiv.2212.02512. arXiv: 2212.02512 [astro-ph.GA].

[218] N. S. M. et al. In: *ApJL* 827.2, L25 (Aug. 2016), p. L25. DOI: 10.3847/2041-8205/827/2/L25. arXiv: 1606.04090 [astro-ph.GA].

[219] P. M. K. et al. In: *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*. Ed. by Makenzie Lystrup et al. Vol. 10698. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July 2018, 106981U, 106981U. DOI: 10.1117/12.2312860.

[220] T. K. et al. In: *ApJ* 508.1 (Nov. 1998), pp. 44–73. DOI: 10.1086/306380. arXiv: astro-ph/9806250 [astro-ph].

[221] M. S. In: *AJ* 124.6 (Dec. 2002), pp. 3050–3060. DOI: 10.1086/344682. arXiv: astro-ph/0209437 [astro-ph].

[222] C. S. and P. E. In: *PASP* 111.760 (June 1999), pp. 691–701. DOI: 10.1086/316373. arXiv: astro-ph/9903067 [astro-ph].

[223] T. S. et al. In: *ApJS* 252.2, 24 (Feb. 2021), p. 24. DOI: 10.3847/1538-4365/abcaa5. arXiv: 2102.01094 [astro-ph.IM].

[224] A. D. et al. In: *AJ* 157.5, 168 (May 2019), p. 168. DOI: 10.3847/1538-3881/ab089d. arXiv: 1804.08657 [astro-ph.IM].

[225] E. F. S., A. M. M., and G. M. G. In: *ApJS* 240.2, 30 (Feb. 2019), p. 30. DOI: 10.3847/1538-4365/aafbea. arXiv: 1901.03337 [astro-ph.IM].

[226] R. S. et al. In: *ApJ* 833.2, 254 (Dec. 2016), p. 254. DOI: 10.3847/1538-4357/833/2/254. arXiv: 1511.08808 [astro-ph.GA].

[227] S. C. et al. In: *Space Telescopes and Instrumentation 2022: Optical, Infrared, and Millimeter Wave*. Ed. by Laura E. Coyle, Shuji Matsuura, and Marshall D. Perrin. Vol. 12180. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Aug. 2022, 121804S, 121804S. DOI: 10.1117/12.2630662. arXiv: 2208.05099 [astro-ph.IM].

[228] Y. T. C., T. C. C., and J. J. B. In: *ApJ* 901.2, 142 (Oct. 2020), p. 142. DOI: 10.3847/1538-4357/abb023. arXiv: 2005.05341 [astro-ph.CO].

[229] S. A. et al. In: *MNRAS* 310.2 (Dec. 1999), pp. 540–556. DOI: 10.1046/j.1365-8711.1999.02978.x. arXiv: astro-ph/9902290 [astro-ph].

[230] O. I. et al. In: *A&A* 457.3 (Oct. 2006), pp. 841–856. DOI: 10.1051/0004-6361:20065138. arXiv: astro-ph/0603217 [astro-ph].

[231] R. d. P. and O. D. In: *PhRvD* 95.12, 123513 (June 2017), p. 123513. DOI: 10.1103/PhysRevD.95.123513. arXiv: 1412.3854 [astro-ph.CO].

[232] A. K., S. F., and M. W. In: *JCAP* 2021.12, 028 (Dec. 2021), p. 028. DOI: 10.1088/1475-7516/2021/12/028. arXiv: 2105.03421 [astro-ph.CO].

[233] G. S. F. et al. In: *arXiv e-prints* (Sept. 2023). DOI: 10.48550/arXiv.2309.05659. arXiv: 2309.05659 [astro-ph.CO].

[234] J. M. et al. In: *MNRAS* 505.3 (Aug. 2021), pp. 4249–4277. DOI: 10.1093/mnras/stab1515. arXiv: 2012.08566 [astro-ph.CO].

[235] K. N. et al. In: *A&A* 670, A149 (Feb. 2023), A149. DOI: 10.1051/0004-6361/202244795. arXiv: 2208.10503 [astro-ph.CO].

[236] D. C. M. et al. In: *ApJ* 877.2, 81 (June 2019), p. 81. DOI: 10.3847/1538-4357/ab184d. arXiv: 1904.06394 [astro-ph.GA].

[237] M. J. I. B. et al. In: *MNRAS* 489.3 (Nov. 2019), pp. 3351–3367. DOI: 10.1093/mnras/stz2324. arXiv: 1908.03720 [astro-ph.GA].

[238] A. B. et al. In: *JCAP* 2020.12, 013 (Dec. 2020), p. 013. DOI: 10.1088/1475-7516/2020/12/013. arXiv: 2006.09368 [astro-ph.CO].

[239] A. B. and E. K. In: *arXiv e-prints* (Feb. 2023). DOI: 10.48550/arXiv.2302.09066. arXiv: 2302.09066 [astro-ph.CO].

[240] J. M. S., T. P., and U. S. In: *arXiv e-prints*, arXiv:2303.08901 (Mar. 2023), arXiv:2303.08901. DOI: 10.48550/arXiv.2303.08901. arXiv: 2303.08901 [astro-ph.CO].

[241] N. E. et al. In: *ApJ* 881.1, 71 (Aug. 2019), p. 71. DOI: 10.3847/1538-4357/ab211a. arXiv: 1809.06380 [astro-ph.GA].

[242] G. S. et al. In: *MNRAS* 524.2 (Sept. 2023), pp. 2395–2406. DOI: 10.1093/mnras/stad2000. arXiv: 2305.08847 [astro-ph.GA].

Chapter 5

PCAT-DE: RECONSTRUCTING POINT-LIKE AND DIFFUSE SIGNALS IN ASTRONOMICAL IMAGES USING SPATIAL AND SPECTRAL INFORMATION

Abstract: Observational data from astronomical imaging surveys contain information about a variety of source populations and environments, and its complexity will increase substantially as telescopes become more sensitive. Even for existing observations, measuring the correlations between point-like and diffuse emission can be crucial to correctly inferring the properties of any individual component. For this task information is typically lost, either 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 point-like and diffuse signals. This work incorporates both explicit spatial templates and a set of non-parametric 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 (CIB) galaxy counts can be recovered in cases of significant contamination. In the second we show that the spatially extended thermal Sunyaev-Zel'dovich (tSZ) effect signal can be reliably measured even when it is subdominant to the point-like emission from individual galaxies.

5.1 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 or behind galactic

synchrotron [243], or stars embedded in regions of high nebulosity [244]. 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., 245], 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., 246]. 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., 247].

Many methods for separating point-like and diffuse signals exist. A review of source detection strategies is presented in M. et al. [248]; work in this field since this review includes M. et al. [249], Z. et al. [250], P. et al. [251], O. and Z. [252], R. et al. [253], L., d. G., and B. [254], C. et al. [255], L., M., and R. [256], and D. et al. [257]. When spatial and/or spectral source properties are well understood, matched filtering is an effective method of source extraction [252, 258], though optimal results only hold under strict assumptions, e.g., sources are isolated in background dominated images, with perfect knowledge of the PSF, noise model, etc. Multi-scale 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 [259, 260]. These approaches can involve several transformations of the data, meaning the quality of extraction of one component (typically the point sources) is emphasized at the cost of poor fidelity on the other components, however work has been done to improve signal reconstructions through more informed transformations and data representations [261, 262]. It has been shown that point source photometry in the presence of nebulosity can be improved considerably by learning a pixel-wise, non-stationary covariance matrix for the structured signal surrounding each source (see [263]; application to DECaPS2 survey in [264]). 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 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., $I(\theta; g(\mathcal{I})) \leq I(\theta; \mathcal{I})$ [265]. 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 datasets.

The work presented in this paper builds on *probabilistic cataloging* [PCAT, 266, 267, 268, 269], a framework that combines transdimensional inference [270] with Bayesian hierarchical modeling to sample from a *metamodel* (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, 271].

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-meter *Herschel* space observatory [272]. The different applications in this work make assumptions about the spatial and spectral behavior of the components, however 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 sub-mm wavelengths; point-like 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; separation of X-ray point sources from diffuse galaxy cluster emission, among others.

The paper is structured as follows. We begin in §2 with an introduction to probabilistic cataloging and its extension to model diffuse emission. The mock *Herschel*-SPIRE data sets are introduced in §3 and we test the performance of our implementation on reconstructing blended emission components based on a range of models and data in §4. The first application explores how well point sources and

their properties are detected/measured (§5), while the second models out the impact of point sources and diffuse emission on the thermal Sunyaev-Zel'dovich (SZ) effect (§6). We conclude in §7 with a discussion of the current `PCAT-DE` implementation and propose a number of potential applications for this formalism.

5.2 Probabilistic cataloging

As telescopes become more sensitive, source extraction becomes increasingly limited by the ability to spatially resolve overlapping sources [37]. 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 [273]. For some datasets 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[1]. For example, source blending will be a major source of systematic uncertainty in a variety of Stage-IV cosmology surveys which rely on accurate galaxy photometry [276].

These challenges motivate probabilistic cataloging. By sampling the full catalog space consistent with a given dataset, probabilistic cataloging can be used to infer both the properties of 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 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 probabilistic cataloging 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 (SDSS), probabilistic cataloging recovers sources with completeness one magnitude deeper than the crowded-field photometry tool

[1] For certain large area surveys the conventional catalog of sub-threshold point sources may not contain more information than the intensity maps themselves [274, 275].

DAOPHOT [277], along with a lower false discovery rate for brighter sources [251]. Performance by these metrics is further improved by extending the hierarchical model to multi-band data, in which case the maps are fit simultaneously [278, 279].

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:

$$\lambda_{ij}^b = \mathcal{P}^b \circledast \Big[\sum_{n=1}^{\mathrm{N_{src}}} S_n^b \delta(x_i - x_n^b, y_j - y_n^b) + A_b^{\mathrm{t}} \sum_{t=1}^{\mathrm{N}_{temp}} \mathrm{I}_b^{\mathrm{t}}(x_i, y_j) \; + \; B_{ij}^b \Big]. \tag{5.1}$$

In this equation, $\mathcal{P}^b$ is the beam function kernel which convolves the signal measured in band b by the point spread function. The signal is decomposed into a sum of point sources with flux densities $\{S_n\}_{n=1}^{\mathrm{N}_{src}}$, $\mathrm{N_{temp}}$ templates for resolved components with known position/spatial structure encoded in surface brightness templates $\{\mathrm{I}_b^t\}$ and amplitudes $\{A_b^t\}$ (see §5.6), 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, non-parametric model is used in signal reconstruction (see §5.2).

For each band b, let $\vec{d}_b$ define the data vector corresponding to the unraveled image with size $\mathrm{W}_b \times \mathrm{H}_b$ pixels and corresponding per-pixel errors given by $\vec{\sigma}_b^2$. 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 MCMC sampling, we compute log-likelihoods, turning the products over pixels and bands into sums:

$$\log \mathcal{L} \approx \sum_{b=1}^{\mathrm{B}} \sum_{i=1}^{\mathrm{W}_b} \sum_{j=1}^{\mathrm{H}_b} -\frac{(\vec{d}_b - \vec{\lambda}_b)^2}{2\vec{\sigma}_b^2}. \tag{5.2}$$

While the likelihood is in a space of fixed dimension set by the data, the space of models is transdimensional, i.e., it is the union of catalogs with varying N_{src}. Because these models reside in nested sub-spaces of one another, priors can be placed

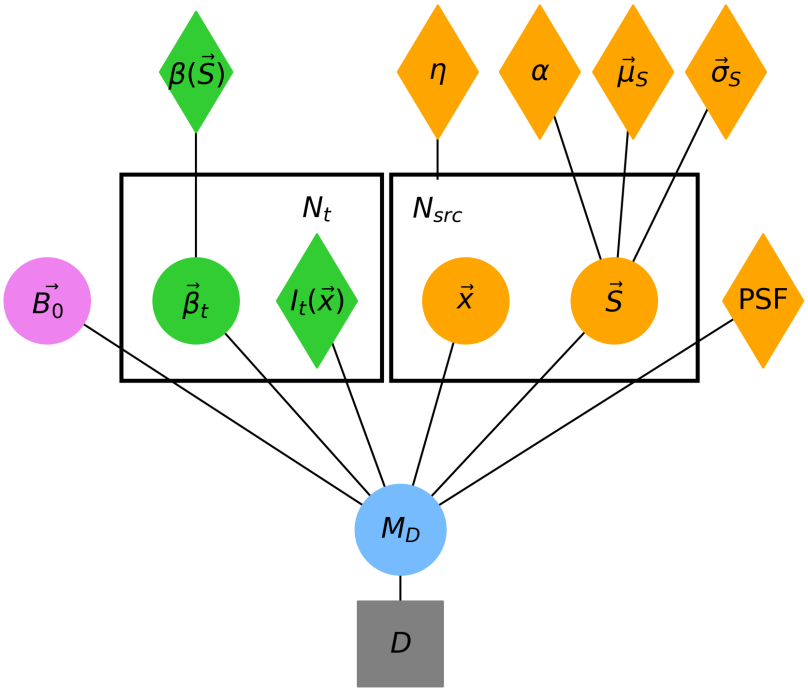

Figure 5.1: Probabilistic graphical model (PGM) for PCAT-DE. The top level of the PGM shows hyperparameters $(\alpha, \vec{\mu}_S, \vec{\sigma}_S, \eta, \beta(\vec{S}))$ that characterize priors over the point source parameters $\{\vec{x}, \vec{S}, N_{src}\}$ and diffuse component colors $\{\vec{\beta}_t\}$. These parameters are then used to generate model images M_D that are compared with the data. Diamonds and circles indicate variables that are fixed and floated, respectively.

on individual mixture components (the point sources) while defining a posterior over the full model space [251].

Figure 5.1 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 Gaussian prior on colors with mean $\vec{\mu}_S$ and width $\vec{\sigma}_S$) and diffuse components (e.g., the color of the diffuse component encoded in $\beta(\vec{S})$) 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. Note for this work that N_t is fixed, while N_{src} is floated as a free parameter.

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:

$$B_{ij}^{b} = B_{0}^{b} + \sum_{n_x=1}^{\mathrm{N_{FC}}} \sum_{n_y=1}^{\mathrm{N_{FC}}} \boldsymbol{\beta}_{n_x n_y} \cdot \boldsymbol{\mathcal{F}}_{ij}^{n_x n_y}. \tag{5.3}$$

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

$$\boldsymbol{\mathcal{F}}_{ij}^{n_x n_y} = \begin{pmatrix} \sin\left(\frac{n_x \pi x_j}{\mathrm{W}}\right) \sin\left(\frac{n_y \pi y_j}{\mathrm{H}}\right) \\ \sin\left(\frac{n_x \pi x_j}{\mathrm{W}}\right) \cos\left(\frac{n_y \pi y_j}{\mathrm{H}}\right) \\ \cos\left(\frac{n_x \pi x_j}{\mathrm{W}}\right) \sin\left(\frac{n_y \pi y_j}{\mathrm{H}}\right) \\ \cos\left(\frac{n_x \pi x_j}{\mathrm{W}}\right) \cos\left(\frac{n_y \pi y_j}{\mathrm{H}}\right) \end{pmatrix}. \tag{5.4}$$

The vector $\boldsymbol{\beta}_{n_x n_y}$ encodes the amplitudes of the Fourier components. All four components of $\boldsymbol{\mathcal{F}}_{ij}^{n_x n_y}$ are necessary in the absence of boundary conditions on the images, which in general will be arbitrary. Throughout this work we use the parameter $\mathrm{N_{FC}}$ when comparing models. The minimum angular scale captured by Fourier components can be approximated by the half-period of the highest frequency Fourier mode along each dimension[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 fluctuations on the scale of the beam, however 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 $k_\theta^{max} < k_\theta^{beam}$, where k_θ^{beam} corresponds to the angular scale of the PSF

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

full width at half maximum (FWHM). Many methods for point source detection, such as `SExtractor` [280] and `Starfinder` [281], include empirical estimates of the local sky signal surrounding each source using a pixel-wise mean or median. The Fourier component representation is flexible enough to accomplish effective sky subtraction, however 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 sampling (see Appendix .1) accelerates the burn-in phase of 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 .3 for a derivation).

5.3 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 [272]. While `PCAT-DE` has been applied to real SPIRE data in [282], controlled sets of mocks are used here in order to characterize performance of the implementation in different limits. SPIRE maps typically contain a combination of emission from 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" [283, 284], 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 is extremely steep [for differential SPIRE number counts see Figure 13 of 285], 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.

Cosmic infrared background galaxies

The majority of sources detected at far-infrared (FIR) wavelengths are $z \sim 2$ galaxies with a 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 B. et al. [286, 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 [282].

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 far-infrared [287]. While cirrus has a blue spectrum across the SPIRE bands similar to that of many observed CIB sources, cirrus contains 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 mask[3] to *Planck* observations, and sample positions uniformly across the unmasked sky. The maps at these positions are queried, re-gridded to SPIRE resolution and extrapolated to 250, 350 and 500 μm using the *Planck*-estimated parameters of a modified blackbody SED

$$S(\nu) = A \left(\frac{\nu}{\nu_0} \right)^{\beta} B_\nu(T_d) \tag{5.5}$$

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. The have dust temperatures ranging from 19-22K over $15'$ patches of the sky, while $\beta \sim 1.5 \pm 0.1$ at the same smoothing scale [288]. We define a nominal "1x-*Planck*" case as a diffuse signal whose power spectrum is parameterized by a single power-law slope:

$$P(k) = P_0(|k|/k_0)^{\gamma} \tag{5.6}$$

where $\gamma = -2.6$ [289] and P_0 is determined by the unmasked sky-averaged power spectrum. Synthetic cirrus maps are drawn as Gaussian random realizations of the 1x-*Planck* power spectrum, and progressively more severe cirrus realizations are

[3]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`.

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 are representative of the worst 50, 32, 17, and 5 per cent of the *Planck* unmasked sky.

Lastly, mock observations are generated for a range of noise levels ranging from 1 mJy beam^{-1} (confusion-dominated for 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 [290], and is dominated by thermal emission from the primary mirror.

Figure 5.2 shows a set of $10' \times 10'$ mock observations at 250 μm 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. Twenty 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 μm maps are used, however the full three-band maps are used in §5.6, where color information helps to distinguish the SZ effect from other astrophysical components.

5.4 Separating point-like and diffuse emission in astronomical images

Model priors

The priors used in this work are nearly identical to those from [282] 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 mininum flux density is imposed on the primary 250 μm band, set to $S^{250}_{min} = 3$ and $S^{250}_{min} = 5$ mJy for low/high instrument noise configurations, respectively. We find the results in this work are relatively insensitive to specific choice of $S_{\rm min}$, though for rigorous characterization of sub-threshold number counts $S_{\rm min}$ can be varied as a hyperparameter. The mean additive normalization of each SPIRE map, denoted 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 .1).

One difference made in this work is the choice of prior on the number of sources,

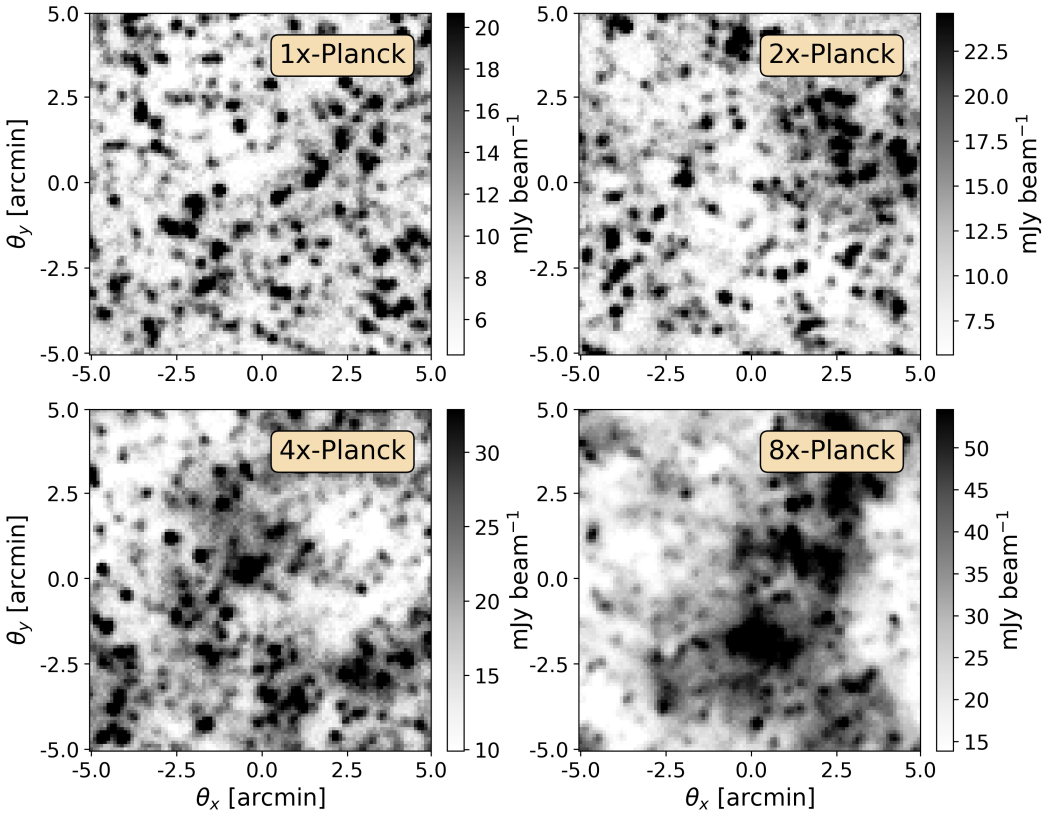

Figure 5.2: Mock realizations of CIB and cirrus at SPIRE resolution with 1 mJy beam^{-1} instrument noise. In this limit confused point sources are the dominant agent reducing our modeling precision. Different panels show levels of cirrus consistent with the median "clean sky" *Planck* sample (top left) and progressively more contaminated fields.

$\pi(\mathrm{N_{src}})$, which counteracts the effect of overfitting due to additional parameters. As explained in Section 3.1 of [279], 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 remove sources at random ("deaths"). The second pair of proposals either split individual model sources into several or merge pairs of sources. As the number of degrees of freedom approaches infinity, the expected improvement in the log-likelihood approaches 1/2 for each additional parameter [291], implying an exponential prior on the number of sources

$$\log \frac{\pi(N+1)}{\pi(N)} = -\frac{1}{2}\rho\left(2+n_f\right). \tag{5.7}$$

As the number of degrees of freedom approaches infinity, the average improvement in the log-likelihood approaches 1/2 per degree of freedom [291]. 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 d.o.f.). 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 non-asymptotic limit in Appendix .2. For our runs we use $\rho = 1.5$, which is slightly more aggressive compared to that assuming single band source number densities from [286] and $S_{min} = 3$ mJy, which suggests $\rho \sim 1.35$ (see Fig. .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.

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 residual diffuse fluctuation signal, however 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 non-trivial because the number of parameters in the Fourier component model scales as k^2_{max}, meaning a penalization based on the number of additional parameters becomes prohibitive.

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

Figure 5.3 shows the residual root-mean-square (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 component-wise 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 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 Fig. 5.3. 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., 4x- and 8x-*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 1x- and 2x-*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 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 component-wise residuals. Figure 5.4 shows the distribution of component-wise residuals over pixels for the most contaminated case (8x-*Planck*). As N_{FC} is increased from two to fifteen, the skewness in both CIB and cirrus residual is reduced considerably. The anticorrelation between CIB and cirrus residual 1-point distributions reflects 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 §5.4. This corresponds to N_{FC} = 5, 7, 11 and 15 for 1x-, 2x-, 4x- and 8x-*Planck*, respectively.

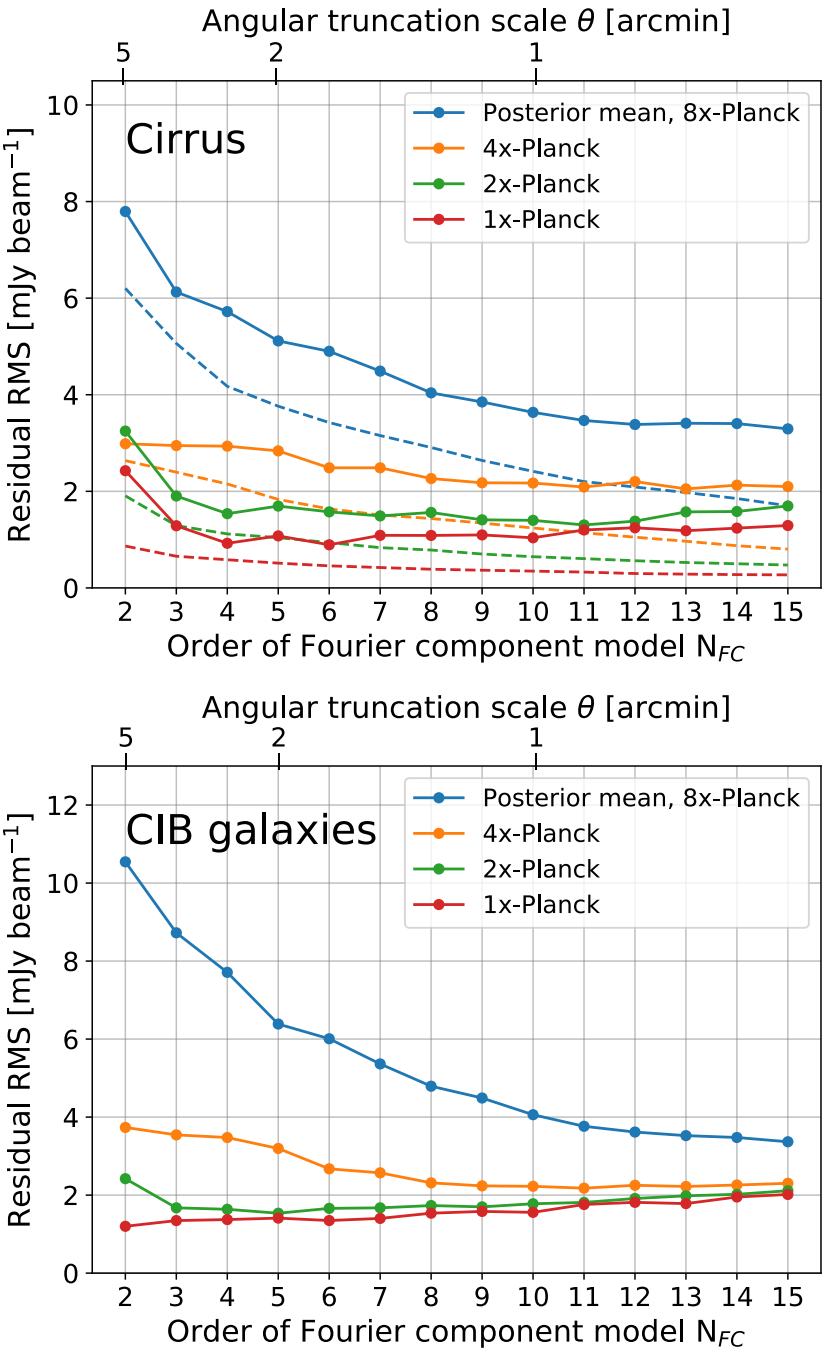

Figure 5.3: Reconstruction accuracy for diffuse cirrus (top) and CIB galaxies (bottom) in mock SPIRE observations, as a function of Fourier component truncation scale (ranging from $\theta_{FC}^{min} = 5'$ to $40''$). Different colors show how results change upon increasing the level of cirrus signal. 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.

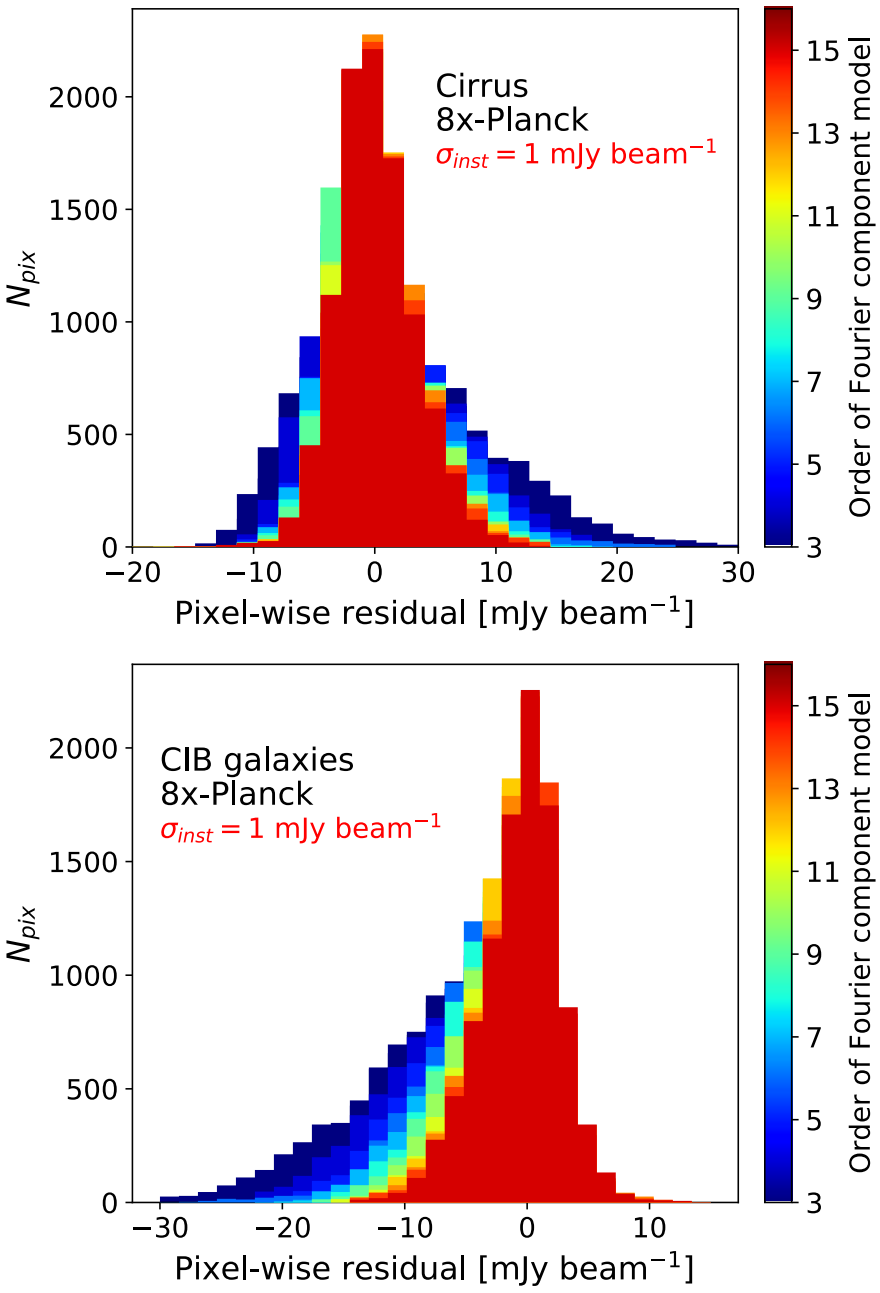

Figure 5.4: Histograms of pixel-wise residuals between cirrus (top), CIB galaxies (bottom) and the posterior mean of each respective model component.

Component separation

Figure 5.5 shows the input and recovered component maps for a mock CIB observation with cirrus at the 8x-*Planck* level, i.e., top 5% of 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 anti-correlated. The primary failure mode occurs when spurious sources compensate for spatially coincident residual diffuse signal, rather than from the diffuse signal concealing true point sources (this is supported by the results of §5.5). Likewise, we find that the CIB residual is weakly correlated with the cirrus spatial curvature (Pearson correlation coefficient $\rho = +0.2$) and the positions of spurious point sources. Regions of high curvature imply presence of higher angular frequency modes that may be difficult to capture with the truncated Fourier series, so this correlation is expected.

Component-wise power spectrum recovery

The `PCAT-DE` model separates signals effectively in both map space and Fourier space. Figure 5.6 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 4x- and 8x-*Planck* cases the recovered CIB has a slight positive bias, which can be attributed to leakage from the much brighter cirrus signals. While cirrus-dominated observations have more false detections and faint-end flux boosting on average (cf. §5.5), 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.

Computational requirements

Forward modeling approaches like probabilistic cataloging are computationally demanding but tractable for targeted science fields. Proposals that perturb the template and mean normalization components are the dominant computational expense for

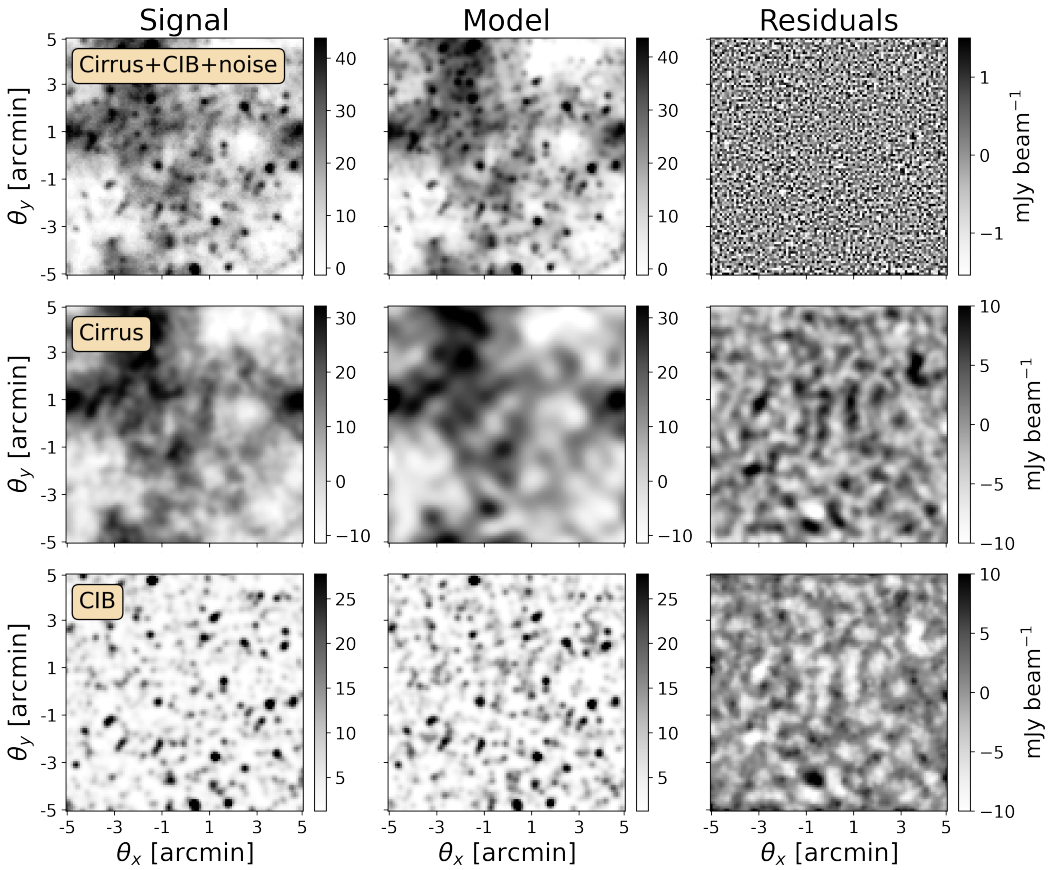

Figure 5.5: Component separation results for a mock CIB realization with injected cirrus dust (at the highest level, 8x-*Planck*) and SPIRE-like noise (σ_{inst} = 1 mJy beam^{-1}). From the left, columns show the data signal (left), the median PCAT model (middle) and corresponding residuals (right).

`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 [279]. 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 interest (ROI). In the absence of mean background and template-based proposals PCAT takes $\sim$ 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 $\sim$ 1 CPU hour without the MKL library. With mean background and template-based proposals included in the Metropolis-Hastings sampling, the

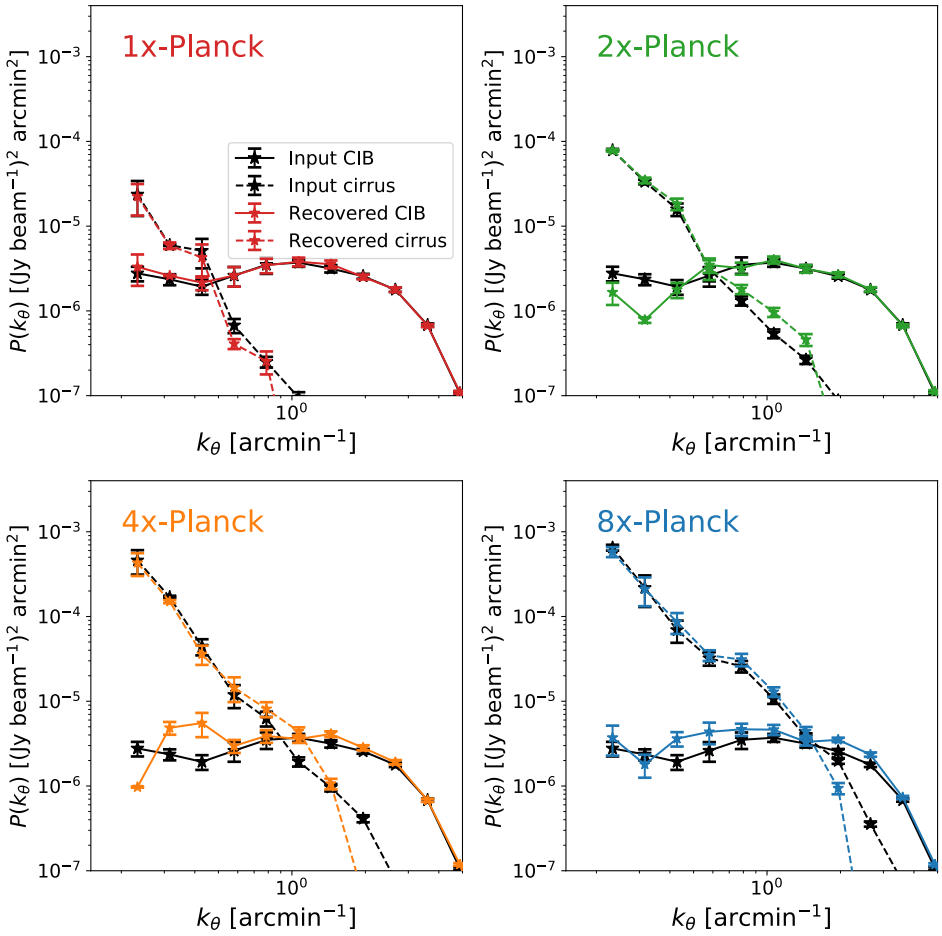

Figure 5.6: Comparison of input and recovered CIB (solid) and cirrus (dashed) power spectra for four cirrus realizations of increasing power. The recovered cirrus is represented by the Fourier component model image, while the CIB comes from the point source model. The per-pixel instrument noise is 10^{-3} Jy beam^{-1}, corresponding to a power spectrum amplitude of 10^{-8} Jy^2 beam^{-2} arcmin^2.

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, intermittently during sampling, making `PCAT-DE` similar to a collapsed Gibbs sampler [292]. The marginalization step integrates out the coefficients of the Fourier component model while fixing the remaining parameters at a given sample and 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.

5.5 Point source detection and population inference in the presence of diffuse emission

By incorporating a Fourier component model into source detection and deblending, probabilistic cataloging 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.

Probabilistic cataloging requires precise control over systematic effects in 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), the posterior on the number of components can diverge [293]. Within probabilistic cataloging 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 $S_{\rm min}$. 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 Fig. 5.7, 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 $S_{\rm min}$ (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 $S_{\rm min}$. 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 cross-matching procedure [see 251, for an outlined procedure], with posteriors obtained from the collection of samples near each source. One can see visually in Fig. 5.7 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 which deviate from idealized Gaussian uncertainties[4].

There is an intrinsic labeling degeneracy in probabilistic cataloging 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 the closest model source to each true source within 6″ (one-third of 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 *smin* after this cross-matching procedure is classified as spurious. While a more stringent cross-matching procedure might include a match on flux density or log-fluxes, we are primarily interested in trends of these statistics with varying Fourier order. The level of blending for SPIRE sources further complicates interpretation of more detailed cross-matches. (see §5.5 for more details).

Figure 5.8 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 1x-*Planck* low- (high-) noise configurations to 16 (35) mJy for 8x-*Planck*. Spurious sources are included when the improvement in the log-likelihood from modeling 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.

[4]The departure from idealized uncertainties (assuming well isolated point source, perfect background subtraction) is quantified with the *degradation factor* [see Appendix C of 251]. That work demonstrated that the degradation factors for positions and fluxes are highly correlated.

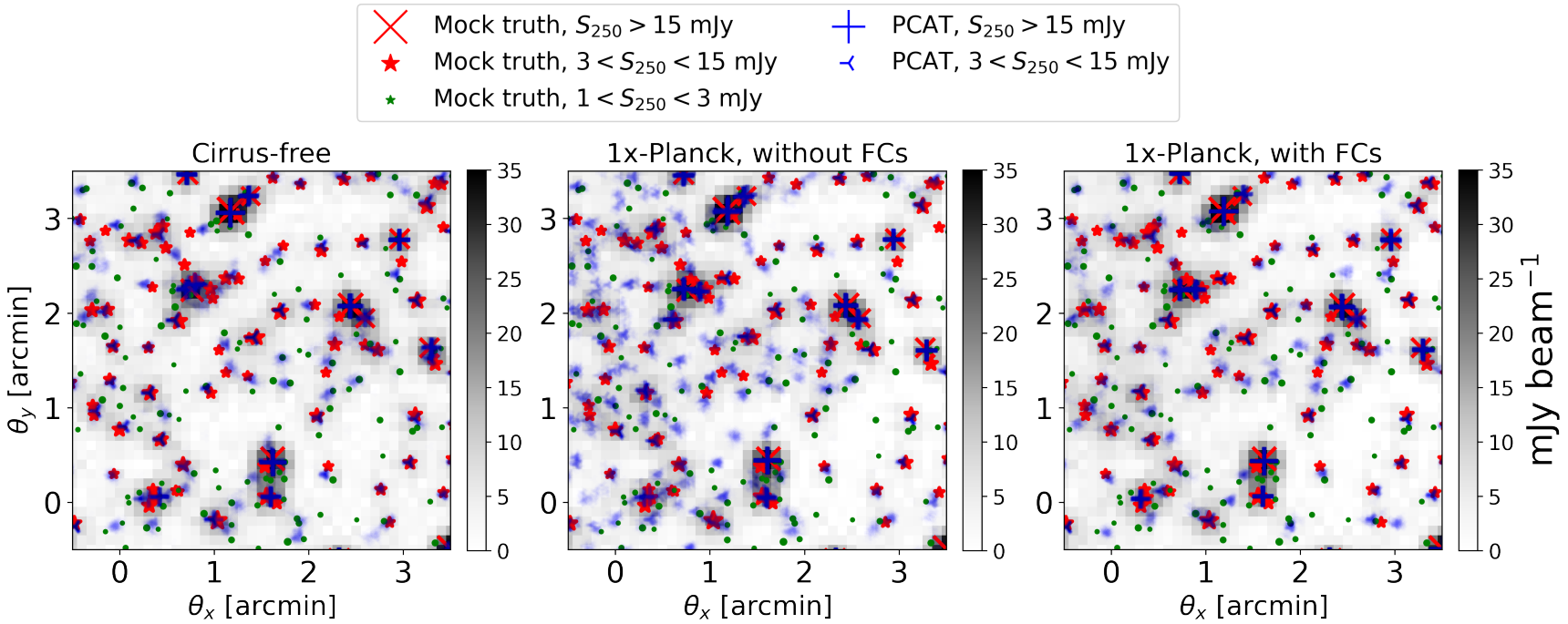

Figure 5.7: Comparison of recovered catalog ensembles for a fixed 250 μm SPIRE CIB observation in the cirrus-free and 1x-*Planck* cases, with and without the use of Fourier components. In each panel, red symbols indicate the true positions of the input catalog down to $S_{250} = 3$ mJy, while the blue shows 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 1x-*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).

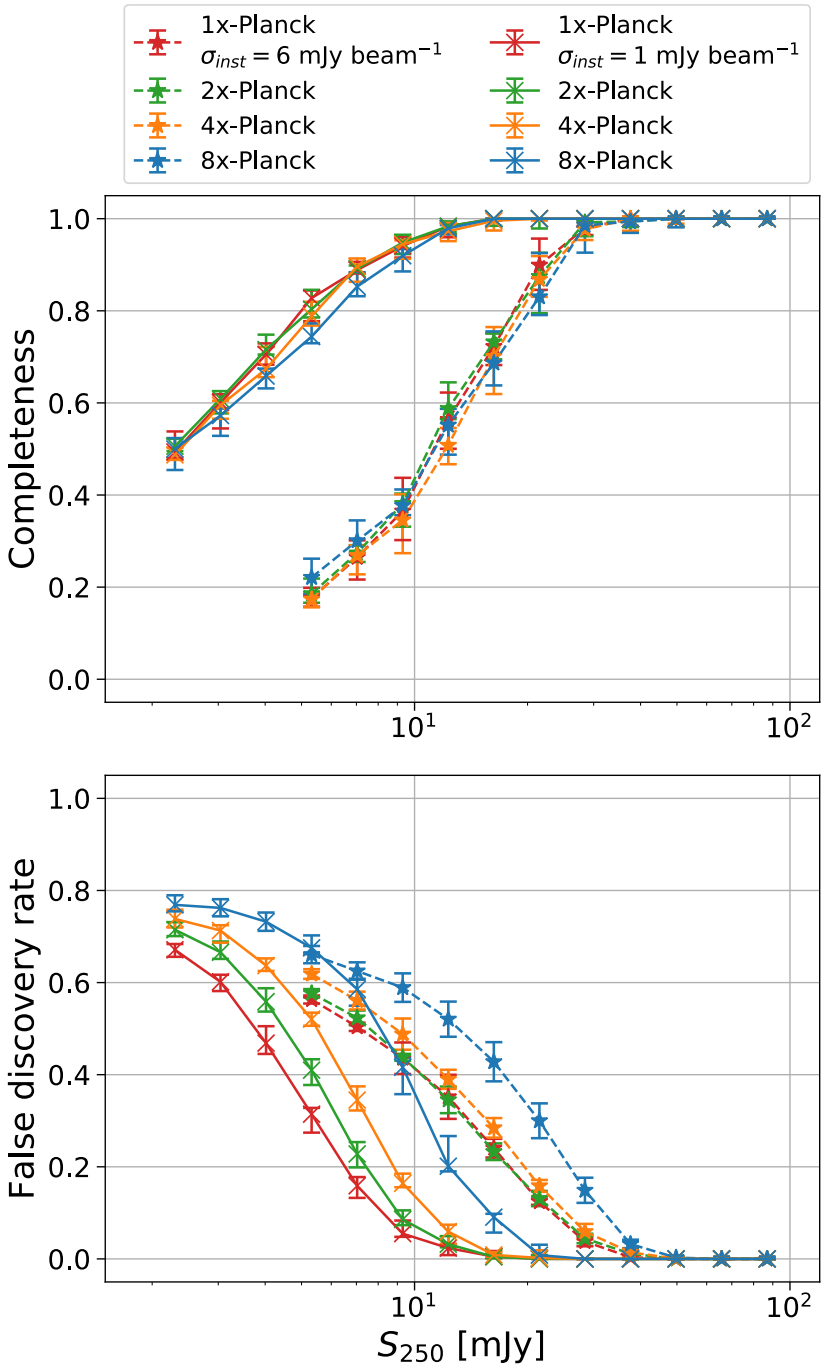

Figure 5.8: Completeness (top) and false discovery rate (bottom) of PCAT-detected sources for mock CIB data with $\sigma_{inst} = 1$ mJy beam^{-1} (solid) and $\sigma_{inst} = 6$ mJy beam^{-1} (dashed). Each line shows the mean and scatter from twenty mock CIB realizations.

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 luminosity function source an asymmetric scatter in the observed counts. In the context of sub-mm analyses, flux boosting can also be sourced by faint blended neighbors. Blending effects in single-dish sub-mm observations are often so severe that number counts are estimated from the one-point distribution of the maps (P(D) analysis, see [294, 295]) 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 to the full transdimensional inference performed with probabilistic cataloging. For a given blend of two (or more) sources, probabilistic cataloging 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, however 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 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 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 (5.7) 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:

$$p(N = 2) \approx \frac{1}{1 + e^{-\Delta \log \mathcal{P}}}. \tag{5.8}$$

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 $p(\mathrm{N} = 2)$. 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 probabilistic cataloging. 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 5.9 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 luminosity function of sub-millimeter galaxies and the angular resolution of SPIRE, this behavior is explainable by source blending. For a given blend of two (or more) sources, PCAT estimates the relative Bayesian evidence between models with different $\mathrm{N_{src}}$. If the likelihood does not

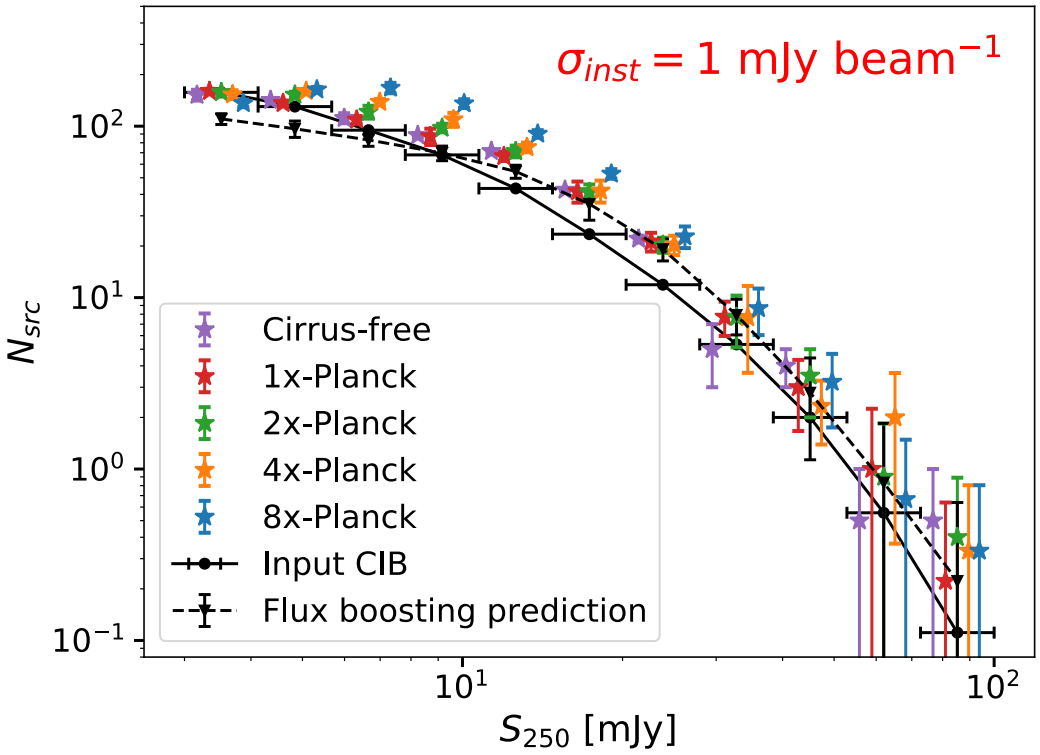

Figure 5.9: Input and recovered flux distributions for twenty mock CIB realizations with increasing cirrus contamination (~ 0.3 deg^2 in total area). The recovered flux distributions are shifted slightly along the x-axis to highlight the dependence on diffuse contamination for each bin. Error bars indicate 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.

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 de-blend 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 [296, 297]. The number counts predictions including 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, Fig. 5.9 shows that on the bright end ($S_{250} > 20$ mJy) the recovered flux distributions are insensitive to all levels of injected cirrus. For $S_{250} \leq 20$ 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 implementation, these results suggest it should be possible to empirically de-boost the observed number counts as a function of foreground contamination.

Sensitivity to N_{FC}

Figure 5.10 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 §4.1. For the 1x-*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 8x-*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.

5.6 Measuring the extended Sunyaev-Zel'dovich 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 (tSZ) effect and its relativistic corrections (rSZ) [298, 299, 300, 301] and cluster peculiar velocities through the kinematic SZ effect [kSZ effect, e.g., 302, 303, 304]. 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 [305].

The SZ effect has a spectral energy distribution (SED) that rises with wavelength across the SPIRE bandpasses. However at 500 μm, that 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

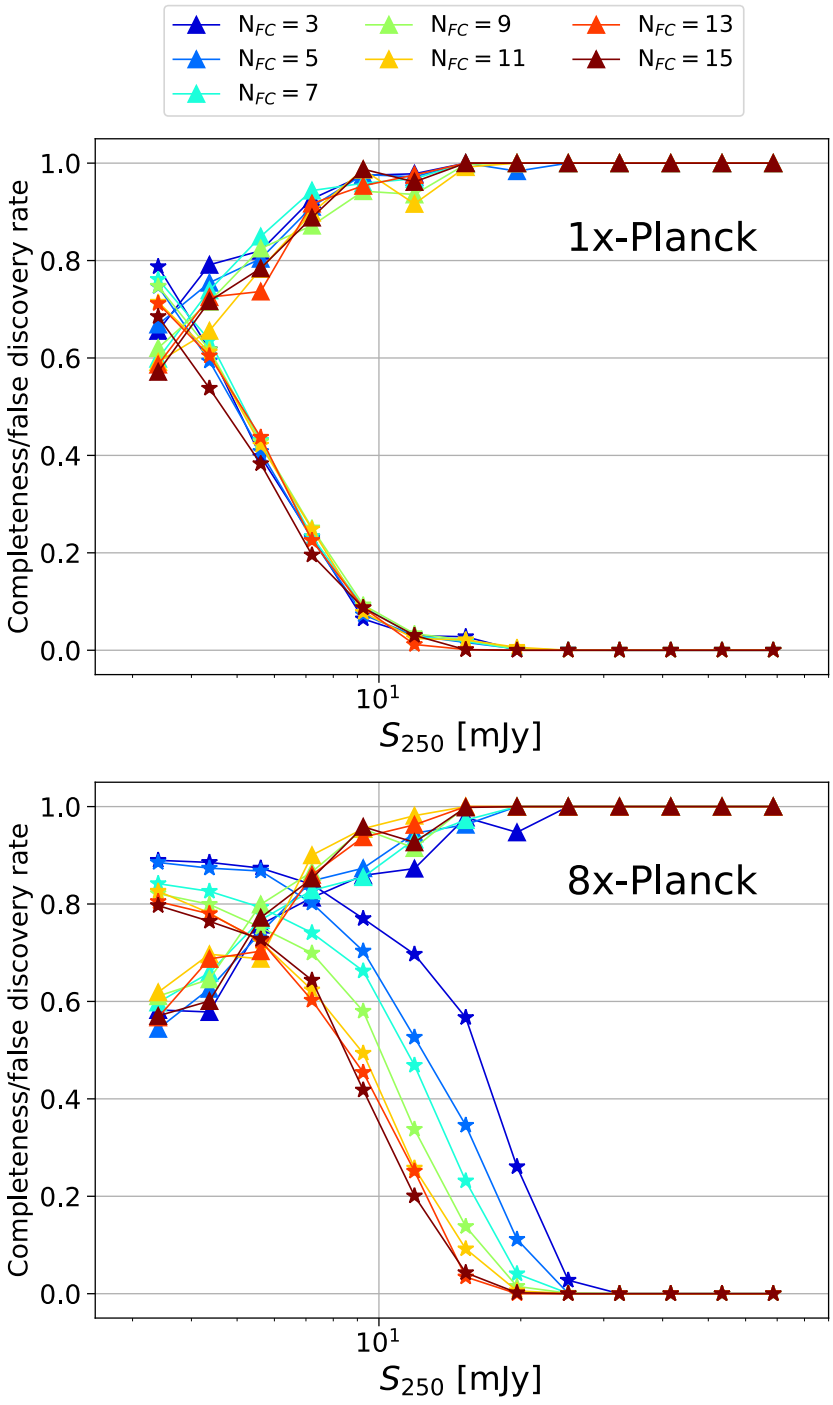

Figure 5.10: Completeness (triangles) and false discovery rates (stars) as a function of flux density. These are evaluated for a range of Fourier component models indicated by the different colors. The top panel shows the 1x-*Planck* results while the bottom shows the same for the 8x-*Planck* case.

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 [278].

Identifying and separating these components can be challenging even when instrument noise is low, especially for single-dish sub-mm measurements for which individual point sources are difficult to separate from truly diffuse emission. While ancillary catalogs provide information about the potential positions of sub-mm detected sources (e.g., deep optical or mid-infrared catalogs), extrapolations of source SEDs over a large wavelength range are required to predict sub-millimeter flux densities. The number of external counterparts per SPIRE beam can be as high as thirty per SPIRE beam [306], meaning some reduction of the external catalog is necessary if the SPIRE data are to be used to constrain the sub-mm 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 probabilistic cataloging, samples are drawn from the marginalized posterior on surface brightness template amplitudes,

$$P(\vec{A}_{SZ}|D) \propto \int P(C)P(\vec{A}_{SZ}|D,C)dC, \tag{5.9}$$

where C denotes the full catalog space.

Tests on mock galaxy clusters

`PCAT-DE` is tested in this section on mock data based on a set of clusters previously observed by *Chandra* and *Herschel* observatories. *Herschel*-SPIRE observed 56 galaxy clusters as part of the HerMES and HLS programs, with map depths of $\simeq$ 1–2 mJy beam^{-1} noise RMS [307, 308], which is subdominant to the SPIRE confusion noise which is $\sim$6 mJy beam^{-1} at 250 μm [283]. Three clusters from this sample are chosen with properties listed in Table 5.1. For each cluster we compute an effective angular FWHM, $\theta_{\rm FWHM}$, 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 [282]. 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 $\sim$2 mJy beam^{-1}. The SZ

signal component is modeled with a set of fixed templates $\mathrm{I}_b^{\mathrm{SZ}}$ with amplitudes $\mathrm{A}_b^{\mathrm{SZ}}$, convolved with the beam:

$$\lambda_{ij}^{b,SZ} = \mathcal{P}^b \circledast \left[\mathrm{A}_b^{\mathrm{SZ}} \mathrm{I}_b^{\mathrm{SZ}}(x_i, y_j)\right]. \tag{5.10}$$

The same templates are then included in the forward model, i.e., $\lambda_{ij}^{b,SZ}$ is added to λ_{ij}^{b} in Eq. 6.1. The SZ effect signal is negligible at 250 μm, so we only fit SZ template amplitudes for SPIRE's 350 μm and 500 μm bands, denoted $\mathrm{A}_{\mathrm{PMW}}^{\mathrm{SZ}}$ and $\mathrm{A}_{\mathrm{PLW}}^{\mathrm{SZ}}$, respectively. The morphology of the SZ profile is assumed to follow an elliptical generalized Navarro-Frenk-White (gNFW) profile [309]. This model is fit to ancillary Bolocam 140 GHz data, after which the best-fit profile is extrapolated and re-gridded to match SPIRE observations, following the method from [304]. 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 reality a small fraction of clusters are observed through lines of sight with significant cirrus contamination exists. Fortunately, cirrus is well constrained by the high resolution 250μm 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 [310] 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 [304] and [282], however 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.

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 [311, 312, 313, 314, 315, 316], including one that uses `PCAT-DE` [282]. To derive constraints on cluster properties like the temperature of gas comprising the intracluster medium (ICM) or cluster

peculiar velocity, multi-wavelength data from several instruments (e.g., *Bolocam*, *Planck*, *Chandra*, Hubble Space Telescope (*HST*)) are commonly employed, however in this work we focus on surface brightness measurements from SPIRE data alone.

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 (PSRF, also known as $\hat{R}$), from several Markov chains run on the same data. Twenty 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 $\hat{\mathrm{A}}^{\mathrm{SZ}}_{\mathrm{PMW}}$ and $\hat{\mathrm{A}}^{\mathrm{SZ}}_{\mathrm{PLW}}$, respectively, suggesting the chains are well mixed.

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 Fig. 5.11, 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 350μm), or is heavily confused as seen at 500μm, `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.

Sensitivity to cluster properties

For each of the three clusters listed in Table 5.1, the same set of twenty CIB + instrument noise realizations are used to generate mock cluster observations, after which samples from each set of twenty chains run on the data are aggregated and plotted in Figure 5.12. Computing this full distribution allows us to quantify systematic uncertainty associated with the CIB, identify any consistent biases and assess the sensitivity of our results to 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 maximum *a posteriori* (MAP) values and 68 per cent

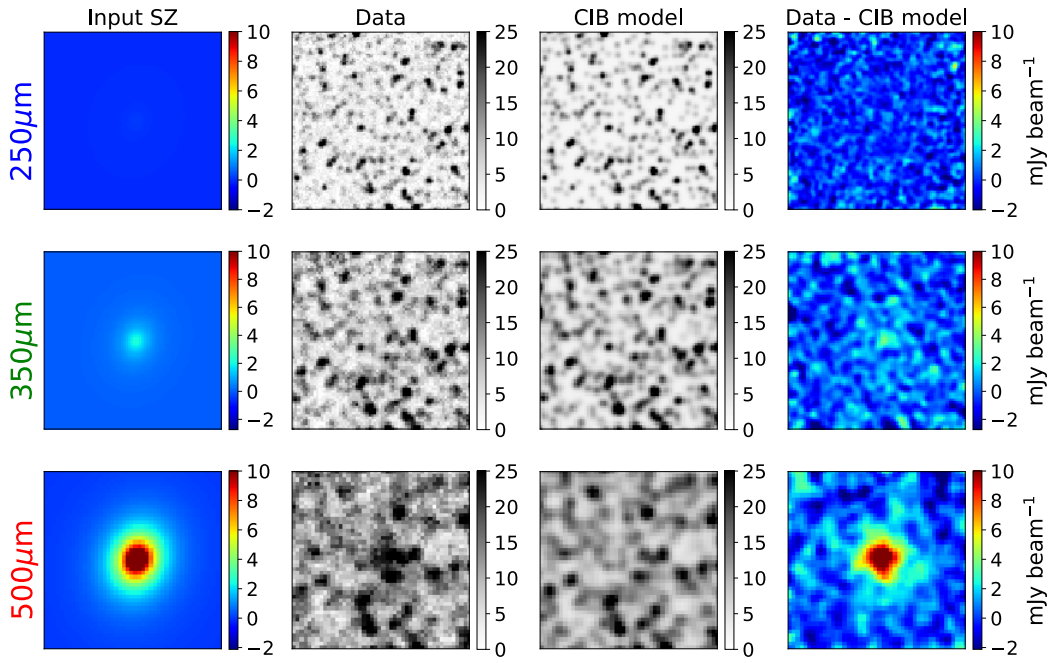

Figure 5.11: $10' \times 10'$ cirrus-free mock SPIRE observations toward galaxy cluster RX J1347.5-1145. The columns show (going from left to right) the input SZ effect signal, observed maps, median CIB model, and difference between observed maps and CIB model. By fitting both SZ and CIB simultaneously we can account for the presence of sub-mm point sources without overfitting the underlying SZ signal. The maps in the rightmost column are shown for visualization purposes only.

credible intervals for A^{SZ}_{PMW} and A^{SZ}_{PLW} are compared with input surface brightnesses in Table 5.1. The mean bias is $\lesssim -0.3\sigma$ for both A^{SZ}_{PMW} and A^{SZ}_{PLW}, 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, however a more thorough investigation of 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., smallest SZ distortion) and smallest angular extent, has larger uncertainties by ~40 and 20 per cent for PMW and PLW, respectively.

5.7 Summary and outlook

In this work we have considered a variety of inference tasks that rely on effective separation of point-like and diffuse signals. This is done by extending the forward modeling framework of probabilistic cataloging (PCAT) with a flexible, template-

Cluster name	Redshift	T_{PW} [keV]	θ_{FWHM} [arcmin]	Input A^{SZ}_{PMW} [MJy sr^{-1}]	$\hat{A}^{SZ}_{PMW}$ (MAP, 68% C.I.)		Input A^{SZ}_{PLW}	$\hat{A}^{SZ}_{PLW}$	
Abell 1835	0.25	8.9	3.6	0.012	$0.007^{+0.016}_{-0.014}$	(0.5σ)	0.068	$0.063^{+0.014}_{-0.026}$	(2.4σ)
RX J1347.5-1145	0.45	17.3	2.5	0.033	$0.028^{+0.016}_{-0.013}$	(2.2σ)	0.130	$0.125^{+0.014}_{-0.020}$	(6.3σ)
MACS J0025	0.59	8.3	2.1	0.010	$0.004^{+0.033}_{-0.018}$	(0.2σ)	0.055	$0.050^{+0.031}_{-0.018}$	(2.2σ)

Table 5.1: Input and recovered SZ effect surface brightnesses for the three test clusters Abell 1835, RXJ 1347 and MACS J0025. 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.

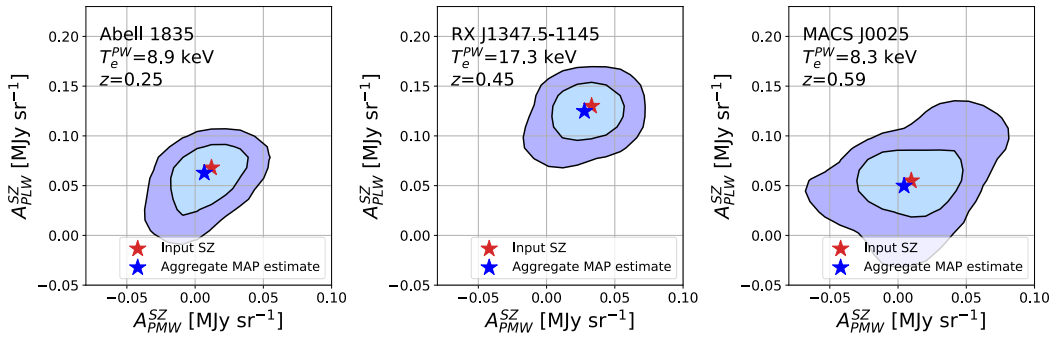

Figure 5.12: Recovered posteriors on A_b^{SZ} surface brightness parameters for clusters of varying redshift/angular extent. The cluster realizations corresponding to these results are cirrus-free, and input SZ surface values for each cluster (marked by red stars) are determined from a combined analysis of *Chandra* and Bolocam data. 1σ and 2σ density contours are plotted for each cluster.

based model for diffuse signals, which results in efficient separation of CIB emission from diffuse Galactic cirrus. Our point source completeness limits (§5.5) 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 robust recovery of number counts across a range of cirrus foreground levels. For our low noise (σ_{inst} = 1 mJy beam^{-1}) case we obtain consistent flux density number counts across all levels of injected cirrus for $S_{250} > 25$ mJy, and the counts in cases up to 4x-*Planck* are consistent for $S_{250} > 12$ 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 galaxy cluster RX J1347.5-1145, to which the SZ spectrum is sensitive through relativistic effects [282]. That work found a temperature of $\langle T_{sz}\rangle_{2500} = 22.4^{+10.6}_{-12.0}$ keV, consistent with the X-ray measured $\langle T_{x,pw}\rangle_{2500}$ = 17.3 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 §6, however more detailed

signal parameterizations may be used when appropriate.

Surveys with strict requirements on photometric accuracy may place stringent cuts on sources based on 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 the tools from this work can both expand the sample size of "usable" sources in astronomical catalogs when 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 Fig. 5.3). More flexible generative models may be able to capture nonlinear structures such as filaments, as have been demonstrated on CMB data to model foregrounds [317, 318]. 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 a 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 [319], on a given inference task.

`PCAT-DE` is publicly available on Github (`github.com/RichardFeder/pcat-de`), with corresponding documentation[5] and examples demonstrating applications from this work.

.1 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 $\boldsymbol{\mathcal{F}\beta}$ is a column vector with the

[5] `pcat-de.readthedocs.io`

flux in each pixel from each Fourier component. Let $\boldsymbol{\Sigma}$ be a diagonal matrix where Σ_{ii} is the variance in pixel i and K the corresponding data vector. Then the $\boldsymbol{\beta}$ that minimizes the chi-squared statistic $\chi^2(\boldsymbol{\beta})$ is given by the Moore-Penrose inverse:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\mathcal{F}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{F}})^{-1} \boldsymbol{\mathcal{F}}^T \boldsymbol{\Sigma}^{-1} K. \tag{11}$$

While the linear inversion from equation (11) 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:

$$\hat{\boldsymbol{\beta}}_{ridge} = (\boldsymbol{\mathcal{F}}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mathcal{F}} + \boldsymbol{\sigma} \boldsymbol{I})^{-1} \boldsymbol{\mathcal{F}}^T \boldsymbol{\Sigma}^{-1} K. \tag{12}$$

The vector $\boldsymbol{\sigma}$ acts as a Gaussian prior on the fluctuation amplitude of each component. In `PCAT-DE`, $\boldsymbol{\sigma}$ 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 (12) 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 15^{th} order Fourier component model, computing $\hat{\boldsymbol{\beta}}$ takes 8 ms per evaluation (with pre-computed 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.

.2 Parsimony prior in the weakly non-asymptotic limit

In probabilistic cataloging, 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 1/2 per additional degree of freedom. However, this asymptotic result may be ill-suited in the limit of severely confused observations.

We consider the F-statistic [320], 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:

$$F = \frac{\chi_1^2 - \chi_2^2}{p_\delta} \frac{N_{pix} - p_2}{\chi_2^2}. \tag{13}$$

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

$$\Delta\chi^2 \sim \frac{\chi_2^2 p_\delta}{N_{pix} - p_2} F(p_\delta, N_{pix} - p_2). \tag{14}$$

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

$$\langle \Delta \log \mathcal{L} \rangle = \frac{1}{2} \frac{\chi_2^2 p_\delta}{N_{pix} - p_2 - 2}. \tag{15}$$

In Figure .13 we plot Eq. 15 for both single- and three-band cases assuming good fits to the data, (i.e., $\chi_2^2 \approx N_{pix}$). 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

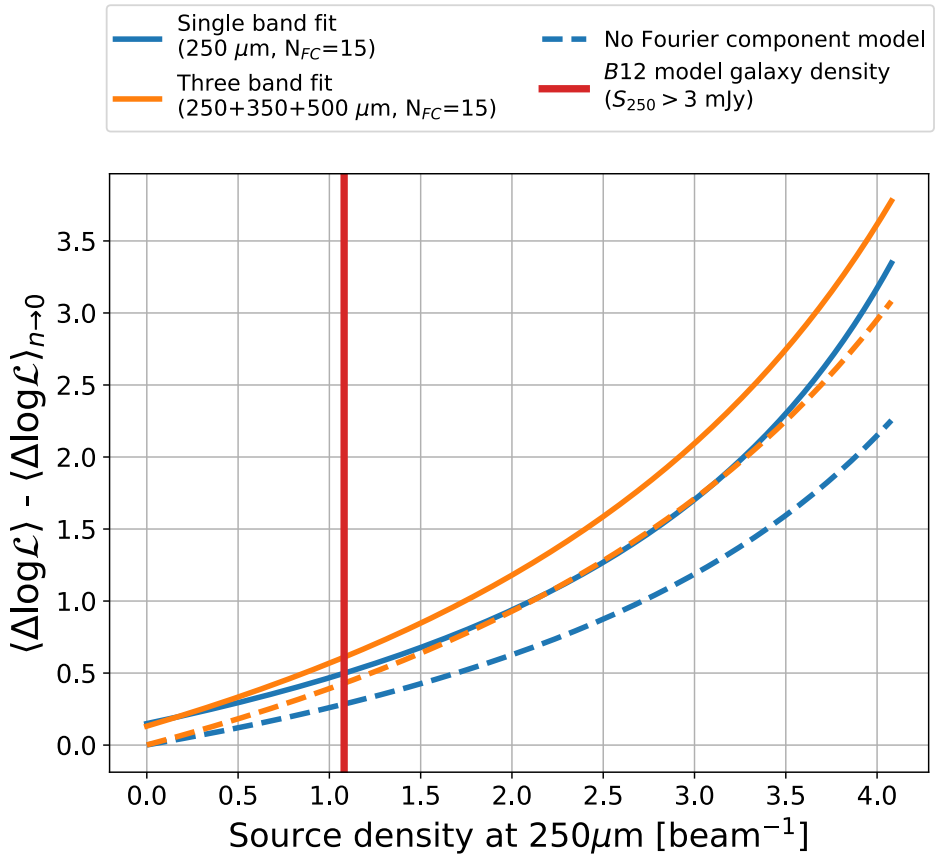

Figure .13: Expected improvement in log-likelihood from adding a point source to the model relative to sparse limit, $\langle \Delta \log \mathcal{L} \rangle_{n \to 0}$, as a function of source density. Three parameters are added to the model per source in the single band case, while each additional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

trace the black-body 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 [286] model (~ 1 beam^{-1}) to predict the expected delta log-likelihood relative to the sparse limit, which informs our choice of scaling parameter ρ defined in (5.7). Additional parameters for the Fourier component model modify the parsimony prior as well, as can be seen comparing the solid and dashed curves in Fig. .13.

.3 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 [321], 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:

$$\sigma_b^2(\theta_{ML}) = \frac{-\partial_f^2 \ln \mathcal{L}(\theta_{ML})}{\partial_b^2 \ln \mathcal{L}(\theta_{ML}) \partial_f^2 \ln \mathcal{L}(\theta_{ML}) - (\partial_b \partial_f \ln \mathcal{L}(\theta_{ML}))^2} = \frac{\sigma^2}{A - A_{psf}} \tag{16}$$

where $A = nm$ and $A_{psf} = \left(\sum_i^{N_{pix}} p_i(x,y)^2\right)^{-1}$ 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 $\ln \mathcal{L}^* = \ln \mathcal{L}(\theta_{ML})$. Our Fisher matrix can then be written as the following:

$$\mathcal{F}_{(\{f_i\}_{i=1}^n, b)}(x,y) = \begin{array}{c} \\ f_1 \\ f_2 \\ \vdots \\ f_n \\ b \end{array} \begin{array}{c} \begin{array}{cccccc} f_1 & f_2 & \dots & f_n & b \end{array} \\ \begin{bmatrix} \partial_{f_1}^2 \ln \mathcal{L}^* & & & & \partial_{f_1}\partial_b \ln \mathcal{L}^* \\ & \partial_{f_2}^2 \ln \mathcal{L}^* & & & \partial_{f_2}\partial_b \ln \mathcal{L}^* \\ & & \ddots & & \vdots \\ & & & \partial_{f_n}^2 \ln \mathcal{L}^* & \partial_{f_n}\partial_b \ln \mathcal{L}^* \\ \partial_b\partial_{f_1} \ln \mathcal{L}^* & \partial_b\partial_{f_2} \ln \mathcal{L}^* & \dots & \partial_b\partial_{f_n} \ln \mathcal{L}^* & \partial_b^2 \ln \mathcal{L}^* \end{bmatrix} \end{array}. \tag{17}$$

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

$$C_{(\{f_i\}_{i=1}^n, b)}(\theta_{ML}) = [\mathcal{F}_{(\{f_i\}_{i=1}^n, b)}(\theta_{ML})]^{-1} = \begin{bmatrix} D^{-1} & \\ & 0 \end{bmatrix} + \rho u u^T \tag{18}$$

where

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

Computing the covariance matrix from equation (18), and evaluating the diagonal background element, one obtains

$$\sigma_b^2(\theta_{ML}) = \frac{\sigma^2}{A - nA_{PSF}}. \tag{19}$$

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 uncertainties on a generic sky model is presented in Appendix C of [321].

References

[37] P. M. et al. In: *arXiv:1708.04058v1* (2017).

[243] C. L. H. et al. In: *MNRAS* 487.3 (Aug. 2019), pp. 3971–3989. DOI: 10.1093/mnras/stz1462. arXiv: 1902.01440 [astro-ph.GA].

[244] J. H. J. d. B. et al. In: *A&A* 576, A74 (Apr. 2015), A74. DOI: 10.1051/0004-6361/201424018. arXiv: 1502.00454 [astro-ph.IM].

[245] R. B. B. In: *Data Analysis in Cosmology*. Ed. by V. J. Martínez et al. Vol. 665. LNP, 2009, pp. 207–235. DOI: 10.1007/978-3-540-44767-2_7.

[246] M. Z. et al. In: *Science* 346.6210 (Nov. 2014), pp. 732–735. DOI: 10.1126/science.1258168. arXiv: 1411.1411 [astro-ph.CO].

[247] A. T. et al. In: *MNRAS* 416.4 (Oct. 2011), pp. 2932–2943. DOI: 10.1111/j.1365-2966.2011.19244.x. arXiv: 1106.0698 [astro-ph.IM].

[248] M. M. et al. In: *MNRAS* 422.2 (May 2012), pp. 1674–1689. DOI: 10.1111/j.1365-2966.2012.20742.x.

[249] M. M. et al. In: *Astronomy and Computing* 9 (Mar. 2015), pp. 10–19. DOI: 10.1016/j.ascom.2014.10.003.

[250] C. Z. et al. In: *MNRAS* 451.4 (Aug. 2015), pp. 4445–4459. DOI: 10.1093/mnras/stv1237.

[251] S. P. et al. In: *AJ, 154, 4* (2017).

[252] E. O. O. and B. Z. In: *AJ* 155.4, 169 (Apr. 2018), p. 169. DOI: 10.3847/1538-3881/aab265. arXiv: 1709.01524 [astro-ph.IM].

[253] A. S. G. R. et al. In: *MNRAS* 476.3 (May 2018), pp. 3137–3159. DOI: 10.1093/mnras/sty440. arXiv: 1802.00937 [astro-ph.IM].

[254] V. L., F. d. G., and M. B. In: *Galaxies* 8.1 (Dec. 2019), p. 3. DOI: 10.3390/galaxies8010003. arXiv: 1910.03631 [astro-ph.IM].

[255] G. H. C. et al. In: *arXiv e-prints* (Apr. 2021). arXiv: 2104.04529.

[256] R. L., J. D. M., and J. R. In: *arXiv e-prints*, arXiv:2102.02409 (Feb. 2021), arXiv:2102.02409. arXiv: 2102.02409 [astro-ph.IM].

[257] J. D. et al. In: *MNRAS* 511.3 (Apr. 2022), pp. 3377–3388. DOI: 10.1093/mnras/stac282.

[258] D. L. and D. W. H. In: *arXiv e-prints*, arXiv:2012.15836 (Dec. 2020), arXiv:2012.15836. arXiv: 2012.15836 [astro-ph.IM].

[259] A. M. et al. In: *A&A* 542, A81 (June 2012), A81. DOI: 10.1051/0004-6361/201218797. arXiv: 1204.4508 [astro-ph.IM].

[260] S. M. et al. In: *A&A* 530, A133 (June 2011), A133. DOI: 10.1051/0004-6361/201014752. arXiv: 1011.3946 [astro-ph.GA].

[261] A. D. et al. In: *A&A* 576, A7 (Apr. 2015), A7. DOI: 10.1051/0004-6361/201424602. arXiv: 1412.5387 [astro-ph.IM].

[262] A. E. et al. In: *A&A* 649, A38 (May 2021), A38. DOI: 10.1051/0004-6361/202038419. arXiv: 2101.03835 [astro-ph.GA].

[263] A. K. S. and D. P. F. In: *arXiv e-prints*, arXiv:2201.07246 (Jan. 2022), arXiv:2201.07246. arXiv: 2201.07246 [astro-ph.IM].

[264] A. K. S. et al. In: *arXiv e-prints* (June 2022). arXiv: 2206.11909.

[265] N. J. B. and R. R. In: *arXiv e-prints* (July 2011). arXiv: 1107.0740 [quant-ph].

[266] B., B. J., F. M., D., H., D. W. In: *AJ, 146, 7* (2013).

[267] B. J. B. and C. P. D. In: *MNRAS* 448.4 (Apr. 2015), pp. 3206–3214. DOI: 10.1093/mnras/stv199. arXiv: 1501.06952 [astro-ph.IM].

[268] T. D. et al. In: *ApJ* 854.2, 141 (Feb. 2018), p. 141. DOI: 10.3847/1538-4357/aaaa1e. arXiv: 1706.06111 [astro-ph.CO].

[269] D. E. J., V. L. K., and D. A. v. D. In: *ApJ* 808.2, 137 (Aug. 2015), p. 137. DOI: 10.1088/0004-637X/808/2/137. arXiv: 1411.7447 [astro-ph.IM].

[270] G., P. In: *Biometrika, 82, 4, 711-32* (1995).

[271] R. F. Version 0.0.2. June 2023. DOI: 10.5281/zenodo.8067131. URL: https://doi.org/10.5281/zenodo.8067131.

[272] M. J. G. et al. In: *A&A* 518, L3 (July 2010), p. L3. DOI: 10.1051/0004-6361/201014519. arXiv: 1005.5123 [astro-ph.IM].

[273] K. B. et al. In: *arXiv e-prints* (Aug. 2022). arXiv: 2208.02781.

[274] E. S. and M. W. In: *JCAP* 2021.5, 067 (May 2021), p. 067. DOI: 10.1088/1475-7516/2021/05/067. arXiv: 2103.01971 [astro-ph.CO].

[275] Y. T. C. et al. In: *ApJ* 877.2, 86 (June 2019), p. 86. DOI: 10.3847/1538-4357/ab1b2b. arXiv: 1809.06384 [astro-ph.CO].

[276] P. M. et al. In: *Nature Reviews Physics* 3.10 (Aug. 2021), pp. 712–718. DOI: 10.1038/s42254-021-00353-y.

[277] P. S. In: *The Astronomical Society of the Pacific* (1987).

[278] R. M. F. et al. In: *AJ* 159.4, 163 (Apr. 2020), p. 163. DOI: 10.3847/1538-3881/ab74cf. arXiv: 1907.04929 [astro-ph.IM].

[279] T. D. et al. In: *ApJ, 839, 4* (2017).

[280] E. B. and S. A. In: *A&AS* 117 (June 1996), pp. 393–404. DOI: 10.1051/aas:1996164.

[281] E. D. et al. In: *Astronomical Data Analysis Software and Systems IX*. Ed. by Nadine Manset, Christian Veillet, and Dennis Crabtree. Vol. 216. Astronomical Society of the Pacific Conference Series. Jan. 2000, p. 623. arXiv: astro-ph/9911354 [astro-ph].

[282] V. B. et al. In: *arXiv e-prints* (Oct. 2021). arXiv: 2110.13932.

[283] H. T. N. et al. In: *Astronomy and Astrophysics* 518 (July 2010), p. L5. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201014680. URL: http://dx.doi.org/10.1051/0004-6361/201014680.

[284] J. J. C. In: *ApJ* 188 (Mar. 1974), pp. 279–286. DOI: 10.1086/152714.

[285] C. M. C., D. N., and A. C. In: *PhysRep* 541.2 (Aug. 2014), pp. 45–161. DOI: 10.1016/j.physrep.2014.02.009. arXiv: 1402.1456 [astro-ph.CO].

[286] M. B. et al. In: *ApJL* 757.2, L23 (Oct. 2012), p. L23. DOI: 10.1088/2041-8205/757/2/L23. arXiv: 1208.6512 [astro-ph.CO].

[287] F. -X. D., F. B., and J. L. P. In: *A&A* 237 (Oct. 1990), p. 215.

[288] Planck Collaboration et al. In: *A&A* 571, A11 (Nov. 2014), A11. DOI: 10.1051/0004-6361/201323195. arXiv: 1312.1300 [astro-ph.GA].

[289] A. B. et al. In: *MNRAS* 412.2 (Apr. 2011), pp. 1151–1161. DOI: 10.1111/j.1365-2966.2010.17971.x. arXiv: 1011.0725 [astro-ph.CO].

[290] M. P. V. et al. In: *ApJ* 772.1, 77 (July 2013), p. 77. DOI: 10.1088/0004-637X/772/1/77. arXiv: 1208.5049 [astro-ph.CO].

[291] S. S. W. In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360. URL: https://doi.org/10.1214/aoms/1177732360.

[292] J. L. In: *Journal of The American Statistical Association - J AMER STATIST ASSN* 89 (Sept. 1994), pp. 958–966. DOI: 10.1080/01621459.1994.10476829.

[293] D. C., T. C., and T. B. In: *arXiv e-prints*, arXiv:2007.04470 (July 2020), arXiv:2007.04470. arXiv: 2007.04470 [math.ST].

[294] J. G. et al. In: *MNRAS* 409.1 (Nov. 2010), pp. 109–121. DOI: 10.1111/j.1365-2966.2010.17781.x. arXiv: 1009.5675 [astro-ph.CO].

[295] T. V. et al. In: *MNRAS* 440.3 (May 2014), pp. 2791–2809. DOI: 10.1093/mnras/stu470. arXiv: 1311.7451 [astro-ph.CO].

[296] K. C. et al. In: *MNRAS* 357.3 (Mar. 2005), pp. 1022–1028. DOI: 10.1111/j.1365-2966.2005.08723.x. arXiv: astro-ph/0412267 [astro-ph].

[297] T. M. C. et al. In: *ApJ* 718.1 (July 2010), pp. 513–521. DOI: 10.1088/0004-637X/718/1/513. arXiv: 0912.2341 [astro-ph.CO].

[298] R. A. S. and Ya. B. Z. In: *Comments on Astrophysics and Space Physics* 4 (Nov. 1972), p. 173.

[299] E. L. W. In: *ApJ* 232 (Sept. 1979), pp. 348–351. DOI: 10.1086/157294.

[300] N. I., Y. K., and S. N. In: *ApJ* 502.1 (July 1998), pp. 7–15. DOI: 10.1086/305876. arXiv: astro-ph/9712289 [astro-ph].

[301] J. C. et al. In: *Monthly Notices of the Royal Astronomical Society* 426.1 (Oct. 2012), pp. 510–530. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2012.21741.x. eprint: https://academic.oup.com/mnras/article-pdf/426/1/510/3278353/426-1-510.pdf. URL: https://doi.org/10.1111/j.1365-2966.2012.21741.x.

[302] R. K. S. and A. D. In: *Monthly Notices of the Royal Astronomical Society* 322.4 (Apr. 2001), pp. 901–917. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2001.04202.x. eprint: https://academic.oup.com/mnras/article-pdf/322/4/901/3339045/322-4-901.pdf. URL: https://doi.org/10.1046/j.1365-8711.2001.04202.x.

[303] J. S. et al. In: *ApJ* 832.1, 26 (Nov. 2016), p. 26. DOI: 10.3847/0004-637X/832/1/26. arXiv: 1605.03541 [astro-ph.CO].

[304] J. S. et al. In: *ApJ* 880.1, 45 (July 2019), p. 45. DOI: 10.3847/1538-4357/ab29ef. arXiv: 1812.06926 [astro-ph.CO].

[305] J. E. et al. In: *MNRAS* 476.3 (May 2018), pp. 3360–3381. DOI: 10.1093/mnras/sty327. arXiv: 1709.01187 [astro-ph.CO].

[306] I. G. R. et al. In: *MNRAS* 409.1 (Nov. 2010), pp. 48–65. DOI: 10.1111/j.1365-2966.2010.17634.x. arXiv: 1009.1658 [astro-ph.CO].

[307] S. J. O. et al. In: *MNRAS* 424 (Aug. 2012), pp. 1614–1635. DOI: 10.1111/j.1365-2966.2012.20912.x. arXiv: 1203.2562 [astro-ph.CO].

[308] E. E. et al. In: *A&A* 518, L12 (July 2010), p. L12. DOI: 10.1051/0004-6361/201014696. arXiv: 1005.3820 [astro-ph.CO].

[309] N. W. E. and J. H. A. In: *PhRvD* 73.2, 023524 (Jan. 2006), p. 023524. DOI: 10.1103/PhysRevD.73.023524. arXiv: astro-ph/0511687 [astro-ph].

[310] M. Z. et al. In: *ApJL* 769.2, L31 (June 2013), p. L31. DOI: 10.1088/2041-8205/769/2/L31. arXiv: 1303.6725 [astro-ph.CO].

[311] E. P. et al. In: *ApJL* 519.2 (July 1999), pp. L115–L118. DOI: 10.1086/312117.

[312] E. K. et al. In: *PASJ* 53.1 (Feb. 2001), pp. 57–62. DOI: 10.1093/pasj/53.1.57. arXiv: astro-ph/0006293 [astro-ph].

[313] T. K. et al. In: *PASJ* 56 (Feb. 2004), pp. 17–28. DOI: 10.1093/pasj/56.1.17. arXiv: astro-ph/0311574 [astro-ph].

[314] M. Z. et al. In: *ApJ* 749.2, 114 (Apr. 2012), p. 114. DOI: 10.1088/0004-637X/749/2/114. arXiv: 1202.0029 [astro-ph.CO].

[315] J. S. et al. In: *ApJ* 820.2, 101 (Apr. 2016), p. 101. DOI: 10.3847/0004-637X/820/2/101. arXiv: 1509.02950 [astro-ph.CO].

[316] T. K. et al. In: *PASJ* 68.5, 88 (Oct. 2016), p. 88. DOI: 10.1093/pasj/psw082. arXiv: 1607.08833 [astro-ph.CO].

[317] B. T., L. K., and K. P. In: *MNRAS* 504.2 (June 2021), pp. 2603–2613. DOI: 10.1093/mnras/stab1011. arXiv: 2101.11181 [astro-ph.CO].

[318] B. R. S. B. et al. In: *arXiv e-prints* (Aug. 2022). arXiv: 2208.03538 [astro-ph.CO].

[319] S J S. et al. In: *Monthly Notices of the Royal Astronomical Society* 499.2 (Sept. 2020), pp. 1587–1606. ISSN: 0035-8711. DOI: 10.1093/mnras/staa2799. eprint: https://academic.oup.com/mnras/article-pdf/499/2/1587/33890975/staa2799.pdf. URL: https://doi.org/10.1093/mnras/staa2799.

[320] R. A. F. In: *Philosophical Transactions of the Royal Society of London Series A* 222 (Jan. 1922), pp. 309–368. DOI: 10.1098/rsta.1922.0009.

[321] S. K. N. P., J. S. S., and D. P. F. In: *AJ* 159.4, 165 (Apr. 2020), p. 165. DOI: 10.3847/1538-3881/ab76ba. arXiv: 1902.02374 [astro-ph.IM].

[366] Planck Collaboration et al. In: *A&A* 594, A22 (Sept. 2016), A22. DOI: 10.1051/0004-6361/201525826. arXiv: 1502.01596 [astro-ph.CO].

[380] T. D., S. K. N. P., and D. P. F. In: *ApJ* 839.1, 4 (Apr. 2017), p. 4. DOI: 10.3847/1538-4357/aa679e. arXiv: 1607.04637 [astro-ph.IM].

[384] Planck Collaboration et al. In: *A&A* 596, A109 (Dec. 2016), A109. DOI: 10.1051/0004-6361/201629022. arXiv: 1605.09387 [astro-ph.CO].

[385] Planck Collaboration et al. In: *A&A* 586 (2016), A132. DOI: 10.1051/0004-6361/201424945. URL: https://doi.org/10.1051/0004-6361/201424945.

[392] J. S. et al. In: *ApJ* 728.1, 39 (Feb. 2011), p. 39. DOI: 10.1088/0004-637X/728/1/39. arXiv: 1010.1798 [astro-ph.CO].

[402] Planck Collaboration et al. In: *A&A* 607, A122 (Nov. 2017), A122. DOI: 10.1051/0004-6361/201630311. arXiv: 1612.07151 [astro-ph.EP].

[403] M. B. In: *arXiv:1701.02434* (2014).

[404] G., A., R., D. In: *Statistical Science, 7, 4, 457-511* (1992).

Chapter 6

MEASUREMENT OF THE RELATIVISTIC SUNYAEV-ZEL'DOVICH CORRECTION IN RX J1347.5-1145

Abstract: We present a measurement of the relativistic corrections to the thermal Sunyaev-Zel'dovich (SZ) effect spectrum, the rSZ effect, toward the massive galaxy cluster RX J1347.5-1145 by combining sub-mm images from *Herschel*-SPIRE with mm-wave Bolocam maps. Our analysis simultaneously models the SZ effect signal, the population of cosmic infrared background (CIB) galaxies, and galactic cirrus dust emission in a manner that fully accounts for their spatial and frequency-dependent correlations. Gravitational lensing of background galaxies by RX J1347.5-1145 is included in our methodology based on a mass model derived from *HST* observations. Utilizing a set of realistic mock observations, we employ a forward modelling approach that accounts for the non-Gaussian covariances between observed astrophysical components to determine the posterior distribution of SZ effect brightness values consistent with the observed data. We determine a maximum *a posteriori* (MAP) value of the average Comptonization parameter of the intra-cluster medium (ICM) within R_{2500} to be $\langle y \rangle_{2500} = 1.56 \times 10^{-4}$, with corresponding 68 per cent credible interval $[1.42, 1.63] \times 10^{-4}$, and a MAP ICM electron temperature of $\langle T_{sz} \rangle_{2500}$ = 22.4 keV with 68 per cent credible interval spanning $[10.4, 33.0]$ keV. This is in good agreement with the pressure-weighted temperature obtained from *Chandra* X-ray observations, $\langle T_{x,pw} \rangle_{2500} = 17.4 \pm 2.3$ keV. We aim to apply this methodology to comparable existing data for a sample of 39 galaxy clusters, with an estimated uncertainty on the ensemble mean $\langle T_{sz} \rangle_{2500}$ at the $\simeq 1$ keV level, sufficiently precise to probe ICM physics and to inform X-ray temperature calibration.

6.1 Introduction

Galaxy clusters are the largest bound objects in the Universe, and gravity is the dominant process driving their evolution and setting their overall physical characteristics. For our analysis, these objects are largely self-similar in their properties [322]. However, deviations from self-similarity can occur from processes outside of simple matter aggregation, such as major mergers between similar mass halos, feedback from active galactic nuclei (AGN), and gas sloshing within the cluster potential [323, 324, 325]. In some cases, such as cosmological studies relying on

precise halo mass estimates, it is important to quantify and understand the effects of these more complicated processes [326]. Such deviations from self-similarity are particularly evident in the spatially resolved thermodynamics of the gaseous intra-cluster medium (ICM), which have typically been studied via the thermal Bremsstrahlung emission in X-rays [327]. However, these studies are generally more difficult and/or impractical at higher redshifts and larger cluster-centric radii, due to $(1 + z)^4$ cosmological dimming and the density squared dependence of the X-ray surface brightness [328, 329]. In addition, the X-ray instruments on board *Chandra* and XMM-*Newton* have limited sensitivity to ICM gas above $\sim$ 10 keV, and such temperatures are expected to be relatively common among the population of massive galaxy clusters, for example, in the shock-heated gas that results from major mergers [330].

A complementary method to study ICM thermodynamics is through the Sunyaev-Zeldovich (SZ) effect signal, which is due to inverse Compton scattering of Cosmic Microwave Background (CMB) photons with energetic electrons in the ICM [331]. The electron population has both random thermal motions and coherent velocities, the latter due to internal turbulence or the overall peculiar velocity of the cluster. These two properties of the ICM give rise to a thermal and a kinematic SZ effect. In addition, the ICM is mildly relativistic, with an average normalized temperature equal to approximately 1% of the normalized electron mass. This gives rise to generally mild temperature-dependent relativistic corrections to the thermal SZ effect spectrum known as the rSZ effect signal [332, 333, 334, 335]. The typical signatures of these signals are shown in Figure 6.1. These relativistic corrections decrease the magnitude of the SZ effect signal at $\lesssim$ 500 GHz, while boosting the signal at higher frequencies.

While X-ray measurements are a more mature probe, with current satellites providing spectroscopic imaging with orders of magnitude better statistics and angular resolution than typical SZ effect observations [336], the SZ effect can play a critical role in several important areas of study. First, because the SZ effect is a fractional distortion of the CMB, the surface brightness is independent of redshift, and so it can more uniformly probe objects across cosmic time [337]. In addition, since the SZ effect signal is due to photon-electron scattering rather than the ion-electron acceleration that sources the X-ray Bremmstrahlung, it is relatively brighter in the lower density cluster outskirt regions [338]. Finally, the rSZ effect signal is sensitive to arbitrarily large temperatures above the band limitations of facilities such as

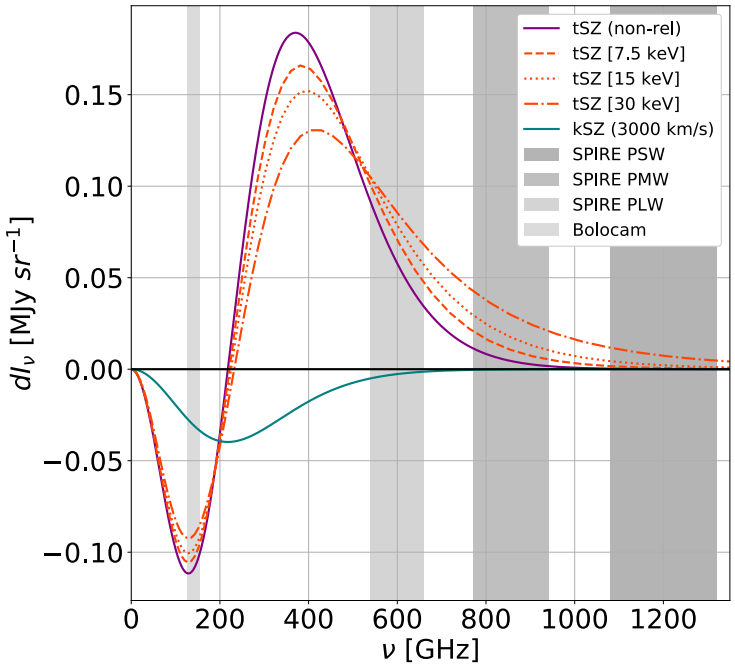

Figure 6.1: The thermal SZ effect spectrum with and without relativistic corrections, along with kSZ effect. The thermal SZ effect spectra were calculated assuming a total Comptonization parameter $y = 10^{-4}$ and four different temperature values. As temperature increases, the overall impact of the rSZ effect is to shift the thermal SZ effect spectrum to higher frequencies. The kSZ effect was determined using the same value of y combined with a peculiar velocity of +3000 km s^{-1} and an electron temperature of 17 keV.

Chandra and XMM-*Newton*.

While there are numerous examples of SZ effect studies pushing to new regimes at high redshift and/or large cluster-centric radii, relatively little effort has been directed toward measuring ICM temperatures with the rSZ effect. This is due to the combination of sensitivity needed to measure the rSZ effect signal and the myriad spatial and spectral data required to separate contamination from other astronomical sources. For example, the contrast between the rSZ effect signal and the canonical tSZ effect signal is maximized at higher frequencies in the sub-millimeter, where the total extragalactic signal is dominated by the Cosmic Infrared Background [CIB; 339].

The difficulty in measuring the rSZ effect is evidenced by the relatively modest constraints obtained to date. The first measurement to indicate a deviation from the non-relativistic tSZ effect spectrum (at $\simeq 2\sigma$) was reported by Z. et al. [340] based on SPIRE observations of the Bullet cluster [see also 341, 342]. Subsequent

observations using the Z-Spec spectrometer obtained a similar detection significance of the rSZ effect in RX J1347.5-1145 [343].

Recent attempts to measure relativistic SZ effect corrections have been based on stacking hundreds of galaxy clusters from the *Planck* all sky survey [344, 345], with these analyses measuring the rSZ effect signal with a statistical significance similar to what was achieved in the earlier single-object studies. Looking forward, planned and potential sub-mm/mm facilities, like LMT/TolTEC [346, 347], the upcoming Fred Young Sub-mm Telescope (FYST; formerly CCAT-prime) [348], the potential Atacama Large Aperture Sub-mm/mm Telescope (AtLAST; [349]) and the Chajnantor Sub/millimeter Survey Telescope (CSST; [350]) hold the promise of delivering the required sensitivity, angular resolution, and field of view to make high-significance rSZ effect measurements routine.

Here we report results from a measurement of the rSZ effect signal toward the galaxy cluster RX J1347.5-1145 using data from *Herschel*-SPIRE, *Bolocam Planck*, *Chandra* and the *Hubble Space Telescope* (*HST*). RX J1347.5-1145 is a famous system that has been the subject of numerous SZ effect studies over the past two decades [311, 312, 351, 313, 314, 315, 316]. It is one of the most massive galaxy clusters observed, with a well-defined cool core and a highly relaxed morphology over most of its projected surface [352, 353, 354]. However, there is clear evidence for a shock to the SE of the core, with a corresponding enhancement to the thermal SZ effect signal in that region [355, 356, 357, 358]. Detailed dynamical analyses of RX J1347.5-1145 indicate that this shock is the result of a 10-to-1 mass ratio merger occurring primarily in a direction orthogonal to the line of sight [359]. While this merger has produced sloshing features in the core gas, it has not significantly disrupted the cool core nor the overall relaxed morphology of the system. Furthermore, the SZ effect enhancement is relatively small in both amplitude and angular extent compared to the bulk signal and, as shown in S. et al. [315], the enhancement is not evident in the arcminute-resolution imaging we employ in this analysis.

This work has resulted in the development of several novel techniques to address the challenges of separating unwanted astrophysical contaminants from the rSZ effect signal. The reduction of the raw data products is described in Section 6.2, and our map fitting procedure is outlined in Section 6.3. A mock observation pipeline is required to accurately characterize our results, and the details of this pipeline are described in Section 6.4. Our analysis of the mocks to assess the uncertainties and biases in our measured SZ effect brightness values is reported in Section 6.5. We

present the procedure used to fit for the average Comptonization parameter and ICM temperature in Section 6.6. Finally, we discuss the implications of this work for current and future studies in Section 6.7. Throughout this work, we assume a flat ΛCDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.7$.

6.2 Data and reduction

SPIRE

SPIRE was an imaging photometer consisting of three focal plane arrays with approximate band centers at 600, 850, and 1200 GHz, which will be identified as PLW, PMW and PSW, respectively for the remainder of this manuscript G. et al. [360]. The point spread functions (PSFs) are well approximated as Gaussians, with full-width at half maxima (FWHMs) of 36″, 25″, and 18″. RX J1347.5-1145 was observed to a depth corresponding to instrument noise levels of $\lesssim 2$ mJy beam^{-1} for all three bands as part of the *Herschel* Multi-Tiered Extragalactic Survey [361, 362].

The SPIRE science archive data used in this analysis were processed through the SMAP/SPIRE-HerMES Iterative Map Maker (SHIM) following the description in L. et al. [363]. SHIM is favored over the *Herschel*-provided tools because it is optimized to separate large-scale correlated noise from the signal, making it better suited for the study of extended emission including the SZ effect. The transfer functions for diffuse astronomical signals were estimated using the methods described in V. et al. [290]. The typical difference between the input and output maps due to high-pass filtering is $\lesssim 1$ per cent. This is sub-dominant to the ~ 5 per cent absolute calibration uncertainty for SPIRE (see Section 6.5), so we assume the map-space signal transfer function is unity (other than the overall mean signal level, which is not preserved by SHIM).

Bolocam

We use Bolocam photometric imaging data collected at an SZ-emission weighted band center of 139 GHz with an approximately Gaussian PSF with a FWHM of 59.2″. The square images are 14′ in size, have an overall calibration uncertainty of 1.7 per cent, and astrometry accurate to 5″. To remove atmospheric fluctuations from the data, a template subtraction and high-pass filter are applied to the data timestreams. This results in an effective angular high-pass filtering of the SZ effect signal, which has been accurately characterized. To obtain an estimate of the SZ effect surface brightness, we follow the general method described by S. et al. [304,

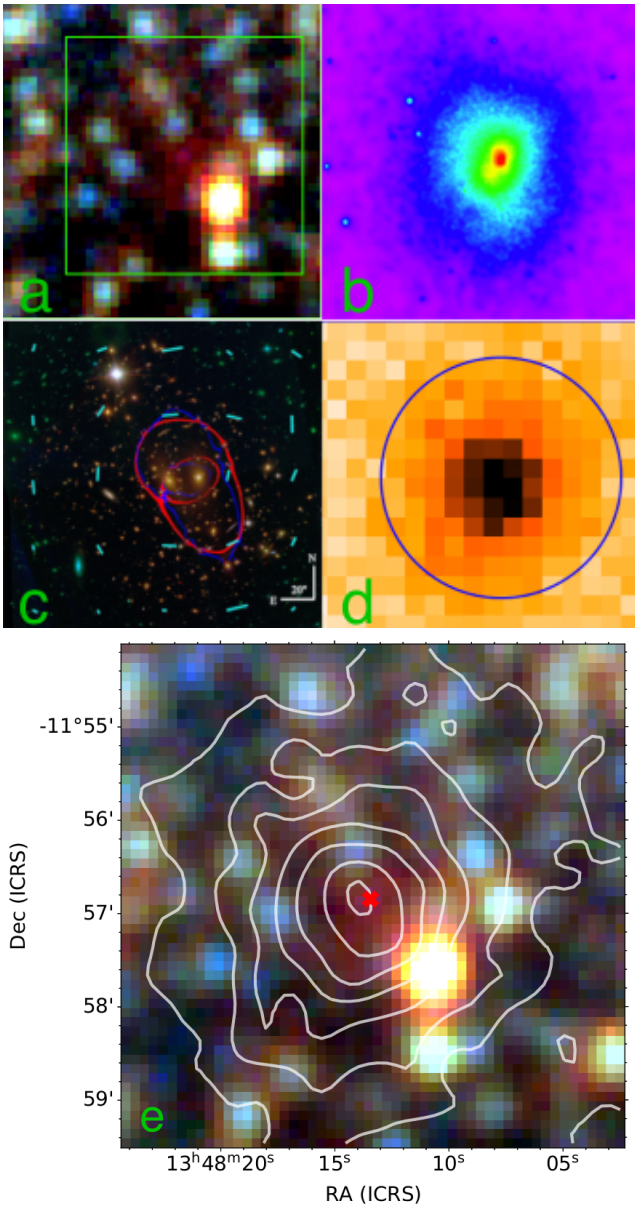

Figure 6.2: All images show the same $6 \times 6\ ^{\prime 2}$ field of view. a) SPIRE three-band false color image. The green box indicates where we fit for the SZ effect amplitude, is $4^{\prime} \times 4^{\prime}$ in size, and is centered on the X-ray center. b) *Chandra* X-ray surface brightness image with logarithmic scaling and smoothing applied. c) *HST* CLASH optical image with lensing model in red/blue contours. d) Bolocam SZ effect image. The circle indicates R_{2500}, within which we compute the average SZ effect signal brightness. e) SPIRE three-band false color image showing a contour overlay of the *Bolocam* image. The red cross denotes the X-ray center. Given the astrometric uncertainties, the peak of the SZ effect in the *Bolocam* image is coincident with the X-ray center (Figure Credit: Victoria Butler).

and references therein], which we briefly summarize here.

First, an elliptical generalized NFW (gNFW) profile [364], using the power-law exponents from A. et al. [365], is fit to the combination of the Bolocam data and the *Planck* MILCA y-map [303, 366]. As noted in S. et al. [303], the addition of the *Planck* data significantly improves the quality of this fit by constraining the SZ effect signal on large angular scales. The PSF of each instrument, along with the Bolocam filtering, are fully accounted for in these fits. Next, the fitted elliptical gNFW model is used to generate a 2-dimensional angular template of the SZ effect signal. This template is then fit to the Bolocam data, with its normalization as the only free parameter, to obtain the average surface brightness of the SZ effect signal within R_{2500} at a frequency of 139 GHz. R_{2500} corresponds to the spherical radius enclosing an average density 2500 times the critical density of the universe, and in RX J1347.5-1145 is measured to be 0.71 $\sim$ Mpc based on the analysis of C. et al. [367]. We note that the technique used to obtain R_{2500} in C. et al. [367] was based on a generalized scaling relation intended to be applicable to large, heterogeneous galaxy cluster samples. For highly relaxed objects, like RX J1347.5-1145, more accurate methods are available [see, e.g., 368, who find a value of 0.80 Mpc]. However, given that R_{2500} is employed in this work solely as a convenient aperture size that is well-matched to the observational data, and that C. et al. [367] had previously computed aperture photometry values from Bolocam data within $R_{2500}\sim=$ 0.71 Mpc, we retain that value for this analysis.

In performing the above fits, we assume the map-space noise in the Bolocam data can be described using a diagonal covariance matrix (i.e., there is no correlated noise between map pixels). Since this assumption is imperfect, we characterize the uncertainty on the SZ effect surface brightness by performing analogous fits to a set of 100 noise realizations. These noise realizations are identical to those used by [304], with the addition of a kinematic SZ effect signal (see Section 6.5). We find that the distribution of SZ effect surface brightness values obtained from the fits to these noise realizations is approximately Gaussian, and so we assign an uncertainty to the measured SZ effect surface brightness in the observed data based on the standard deviation of this distribution.

Chandra

ICM temperatures measured using the rSZ effect are approximately pressure-weighted [e.g., 369], and so we compute an analogous temperature from the available *Chan-*

dra X-ray spectroscopic imaging [1]. The procedure for reducing and cleaning the data is described in M. et al. [368, 354], although we use more recent versions of the *Chandra* analysis software and calibration files for this analysis (`CIAO` version 4.9 and `CALDB` version 4.7.4). From these data, we obtain deprojected density and temperature profiles using the techniques detailed in M. et al. [368, 371]. Using these profiles, we then compute the pressure-weighted mean temperature of the ICM within a cylindrical volume defined by R_{2500} in the plane of the sky and the maximum radial extent probed by the X-ray data along the line of sight, corresponding to $\sim 1.8\ R_{2500}$. As part of this calculation, we also apply the empirically-derived temperature calibration bias from W. et al. [372], equal to 0.09 ± 0.13. With this calibration correction, we obtain a value of $\langle T_{x,pw} \rangle_{2500} = 17.4 \pm 2.3$ keV.

HST

The galaxy cluster RX J1347.5-1145 was extensively imaged with *Hubble* as part of the Cluster Lensing and Supernova *HST* (CLASH) survey [373, 374]. The *HST* images were used to construct mass models according to the `PIEMDeNFW` formalism [375]. This parametric model incorporates elliptical NFW dark matter profiles to model the cluster dark matter halo, and double pseudo isothermal elliptical mass distributions to model cluster member galaxies. In order to construct a model that can be used over the full extent of the SPIRE maps, the best-fit parameters from the `PIEMDeNFW` model were used to extrapolate onto a larger grid [for more details, see 304]. These grids have sizes of $16' \times 16'$ and a angular resolution of $0.25''$.

6.3 Multi-component map fitting

A significant challenge to measuring the SZ effect from observations in the sub-mm/mm is constraining a signal that is contaminated with emission from dusty star-forming galaxies (DSFGs), including background DSFGs that are lensed by the cluster, galactic cirrus dust emission and potentially other components such as diffuse dust emission from the cluster itself [376, 345, 377, 378]. These spatially coincident signals can be disentangled through prior information on the spatial morphology and underlying spectral energy distributions (SEDs) of the emission components.

Spectral information has previously been used through multiband matched filtering to minimize the contribution from unwanted signals in a stacking analysis of galaxy

[1] For an example of the biases that can result from comparing the standard emission-weighted X-ray spectroscopic temperature to rSZ effect temperatures, see L. et al. [370].

clusters in *Planck* maps [345]. However, in general *Planck* does not resolve the galaxy clusters, and so the possibility of using spatial information is not available. SPIRE data, however, permit detailed spatial-spectral modeling to separate the unresolved point-like DSFGs from the diffuse SZ effect and cirrus dust emission. The combination of angular resolution, instrument sensitivity and the DSFG luminosity function set intrinsic limits on the efficacy of this type of approach, but for certain observations the gains in robustness and constraining power can be significant.

The multi-component fitting method employed in this work builds on the hierarchical modeling framework of probabilistic cataloging, or PCAT [379, 380, 381, 278]. PCAT is designed to explore the space of catalog models consistent with an observed image by fitting a Poisson mixture model directly to the data, where each component represents a point-like source. As a Bayesian hierarchical model, probabilistic cataloging combines prior information about the source population of interest with the data likelihood to estimate a posterior distribution of point-like source models consistent with our prior expectations and the data. This is represented as an ensemble of catalogs that naturally encode the often complicated model uncertainties that can arise in confusion-limited observations (such as from SPIRE) and crowded fields in general.

The major utility of PCAT as it applies to this analysis is that a point source model can be estimated directly from the data in such a way that marginalization over the point source model enables better estimation of a correlated signal of interest, where in this case the correlated signal is the SZ effect. This includes marginalization over uncertainties associated with point source positions and flux densities, as well as uncertainty due to our ignorance of the true number of DSFGs (down to a given flux density) in the observed field. This is necessary in the absence of ancillary data from deeper observations with finer angular resolution. In this section, we describe the forward model used in probabilistic cataloging and the extension to observations of the SZ effect toward galaxy clusters.

Generative model

For an image with dimension (W, H) observed in band b, the surface brightness sampled by pixel (i, j) is written as a sum over point-like and diffuse components convolved with the PSF $\mathcal{P}$:

$$\lambda_{ij}^{b} = \mathcal{P}^{b} \circledast \left[\sum_{n=1}^{\mathrm{N_{src}}} S_{n}^{b} \delta(x_i - x_n^b, y_j - y_n^b) \; + \mathrm{A}_{b}^{\mathrm{SZ}} I_{b}^{\mathrm{SZ}}(x_i, y_j) \; + \; B_{ij}^{b} \right]. \tag{6.1}$$

The first sum term denotes the contribution from $\mathrm{N_{src}}$ point-like sources with flux densities $\{S_n\}_{n=1}^{\mathrm{N_{src}}}$ and positions $\{x_n\}_{n=1}^{\mathrm{N_{src}}}$, and assumes that the galaxies in our image are well represented as point sources, which is reasonable given the SPIRE PSF size and the galaxy redshift distribution. The value of $\mathrm{N_{src}}$ fluctuates due to the transdimensional modeling employed by PCAT but is defined as the number of sources above the 5 mJy detection limit. The second term is the spatially extended SZ effect, included in our model as an angular template I_{SZ}^b scaled by amplitude $\mathrm{A}_b^{\mathrm{SZ}}$, with the angular template constructed from the elliptical gNFW fit to Bolocam and *Planck* (see Section 6.2). The last term, B_{ij}^b, captures additional diffuse components in a non-parametric fashion through the addition of a two-dimensional truncated Fourier series:

$$B_{ij}^b = B_0^b + \sum_{n_x=1}^{N_m} \sum_{n_y=1}^{N_m} \boldsymbol{\beta}_{n_x n_y} \cdot \boldsymbol{\mathcal{F}}_{ij}^{n_x n_y}, \tag{6.2}$$

where $\boldsymbol{\mathcal{F}}_{ij}^{n_x n_y}$ is a vector of values from Fourier components with wavevector $(k_x, k_y) = (\mathrm{W}/n_x, \mathrm{H}/n_y)$ evaluated at pixel (i, j):

$$\boldsymbol{\mathcal{F}}_{ij}^{n_x n_y} = \begin{pmatrix} \sin\left(\frac{n_x \pi x_j}{\mathrm{W}}\right) \sin\left(\frac{n_y \pi y_j}{\mathrm{H}}\right) \\ \sin\left(\frac{n_x \pi x_j}{\mathrm{W}}\right) \cos\left(\frac{n_y \pi y_j}{\mathrm{H}}\right) \\ \cos\left(\frac{n_x \pi x_j}{\mathrm{W}}\right) \sin\left(\frac{n_y \pi y_j}{\mathrm{H}}\right) \\ \cos\left(\frac{n_x \pi x_j}{\mathrm{W}}\right) \cos\left(\frac{n_y \pi y_j}{\mathrm{H}}\right) \end{pmatrix} \tag{6.3}$$

and $\boldsymbol{\beta}_{n_x n_y}$ are the component amplitudes. Including all four components in $\boldsymbol{\mathcal{F}}^{n_x n_y}$ is necessary because we do not wish to impose boundary conditions on the diffuse model. We determine an appropriate truncation scale for the Fourier component model empirically by minimizing scatter on the inferred SZ effect template amplitudes recovered from mock observations (see Section 6.4). In general, the choice of truncation scale depends on both the power spectrum of the diffuse signal contamination and the scale of the PSF, which acts as a low-pass filter through its convolution with B_{ij}^b. The mean surface brightness of each image, captured by B_0^b, is not physically meaningful because the SPIRE maps are not absolutely calibrated, and so B_0^b is treated as a nuisance parameter in our model.

Data likelihood and priors

Our data likelihood is assumed to be Gaussian and is written in map space as a product over pixels:

$$\mathcal{L} = \prod_{b=1}^{\mathrm{B}} \prod_{i=1}^{\mathrm{W}} \prod_{j=1}^{\mathrm{H}} \frac{1}{\sqrt{2\pi\vec{\sigma}_b^2}} \exp\left(-\frac{(\vec{d}_b - \vec{\lambda}_b)^2}{2\vec{\sigma}_b^2}\right) \tag{6.4}$$

where $\vec{d}_b$ is the observed data vector and $\vec{\lambda}_b$ is the generated model. The noise model that sets the pixel-wise variance σ_b^2 in Eq. (6.4) is estimated from the SPIRE timestream data following the procedure described in V. et al. [290]. The above expression then reduces to

$$\log \mathcal{L} \approx \sum_{b=1}^{\mathrm{B}} \sum_{i=1}^{\mathrm{W}} \sum_{j=1}^{\mathrm{H}} -\frac{(\vec{d}_b - \vec{\lambda}_b)^2}{2\vec{\sigma}_b^2}. \tag{6.5}$$

We assume galaxies are randomly distributed on the sky, placing uniform priors on source positions, although we note this is an imperfect assumption mainly due to the spatial inhomogeneities resulting from gravitational lensing of CIB sources by the galaxy cluster. The multiband flux density prior π is factorized into a flux density prior for the shortest wavelength band at 250 μm (PSW) and color priors for the remaining two bands:

$$\pi(\vec{S}) = \pi(S_{\mathrm{PSW}})\pi\left(\frac{S_{\mathrm{PMW}}}{S_{\mathrm{PSW}}}\right)\pi\left(\frac{S_{\mathrm{PLW}}}{S_{\mathrm{PSW}}}\right). \tag{6.6}$$

We assume the source flux density distribution follows a power law:

$$\pi(S_{\mathrm{PSW}}) \propto \left(\frac{S_{\mathrm{PSW}}}{S_{\mathrm{PSW}}^{\mathrm{min}}}\right)^{-\alpha} \tag{6.7}$$

and $\alpha = 3.0$ [382]. The color priors are modeled as Gaussian with means consistent with the typical DSFG SED. The widths of the color priors are optimized along with other hyperparameters using the mock observations described in Section 6.4 to minimize scatter in the inferred SZ effect brightness. The minimum source flux density permitted by the model is determined in a similar fashion and is fixed to $S_{\mathrm{PSW}}^{min} = 5$ mJy. The diffuse cirrus model is represented in image space as a linear combination of templates (one template per Fourier component), and the coefficients of those templates are sampled with uniform priors. Lastly, while the SZ effect increment between PMW and PLW has a well-defined range of colors for plausible temperatures, we choose to place independent, uniform priors on the

SPIRE SZ effect template amplitudes to capture potential systematic effects that could bias the inferred surface brightness values.

By sampling the space of models consistent with observed data D using Eq. (6.4) and regularizing the set of solutions with suitable priors, we can compute the posterior distribution of astrophysical models, $P(\mathcal{M}|D)$, through Bayes' rule:

$$P(\mathcal{M}|D) = \frac{\mathcal{L}(D|\mathcal{M})\pi(\mathcal{M})}{P(D)} \propto \mathcal{L}(D|\mathcal{M})\pi(\mathcal{M}). \tag{6.8}$$

Here, $\mathcal{L}(D|\mathcal{M})$ is the likelihood of observing data D given astrophysical signal model $\mathcal{M}$. PCAT uses a Markov Chain Monte Carlo (MCMC) sampler which has been optimized to efficiently explore the posterior distribution of catalogs consistent with image data. Details on implementation and the sampling algorithm can be found in [381] and [278].

Fitting procedure

Our procedure to fit the model to multiband SPIRE image data happens in two steps. First, `PCAT-DE` is run on the PSW-only data to determine a spatial model for cirrus dust emission. A sixth-order Fourier component model is fit to the data ($\theta_{\rm min} \sim 1.6'$, $n_{param} = 144$) along with the point source model and mean surface brightness level. Second, `PCAT-DE` is run on all three bands simultaneously, with the shape of the Fourier component model fixed. The Fourier component model from PSW is then scaled to PMW and PLW, assuming a constant cirrus SED across the field of view and constant B_0^b. The assumed cirrus SED is taken from SPIRE observations of the H-ATLAS SDP field [289]. In this step the SZ templates for PMW and PLW are added to and fit jointly with the rest of the model.

This two-step procedure is performed in order to mitigate the effect of degeneracies between the cirrus model and the other signal components. The spectrum of cirrus dust is well constrained and is brightest at short wavelengths, so determining the spatial structure of cirrus using PSW data alone is sufficient. The single-band fit also avoids possible bias of the cirrus dust model by the SZ effect itself, which is also spatially extended. The SZ effect template amplitudes are floated for PMW and PLW, but since the the SZ effect brightness is close to zero at high frequencies we fix $\mathrm{A}_{\rm PSW}^{\rm SZ}$ to the value predicated from the combination of the measured Bolocam SZ effect brightness $\langle d\mathrm{I}_{\rm Bcam}\rangle_{2500}$ and *Chandra* $\langle \mathrm{T}_{\rm x,pw}\rangle_{2500}$, with $\langle d\mathrm{I}_{\rm PSW}\rangle_{2500}$ =0.005 MJy/sr. This eliminates the possibility of CIB and/or cirrus emission in the PSW map from being incorrectly modeled as SZ effect signal.

Probabilistic cataloging is computationally expensive compared to other point source detection/extraction algorithms. This is because both the number of parameters in PCAT-DE's forward model and the degrees of freedom in the image data are large. To ensure we recover a well-sampled posterior on $\{ A_b^{SZ} \}$, we restrict our multi-component map fitting procedure to the $4' \times 4'$ region centered on the SZ-defined cluster centroid. However, we compute the best fit mean level for each band using the larger $10' \times 10'$ maps and then fix those values in fits of the $4' \times 4'$ maps. We confirm through tests on mock data that this procedure does not bias our SZ estimates, but we do include an additional statistical uncertainty determined by how much the inferred SZ template amplitudes vary using a range of mean surface brightness levels consistent with our $10' \times 10'$ map fits (see §6.5). Crucially, fits on the larger maps are done using the full model from Eq. (6.1), which allows for unbiased recovery of the mean level in each band.

The outputs of this fitting process comprise: a catalog of three-band point source flux densities and associated errors, three-band Fourier component amplitudes of the cirrus emission, and SZ effect template amplitudes in PMW and PLW bands expressed as a fitted brightness amplitude A_b^{SZ}. These surface brightnesses can then be corrected for the relevant biases and combined with the measured Bolocam surface brightness to constrain the SZ effect parameters $\langle y \rangle_{2500}$ and $\langle T_{sz} \rangle_{2500}$, as detailed in Section 6.6.

6.4 Mock observation pipeline

In order to validate our analysis pipeline, assess biases on the inferred SZ effect signal, and quantify uncertainties associated with instrument noise and astrophysical contaminants, synthetic multiband SPIRE maps of RX J1347.5-1145 are generated. The components in these maps include instrument noise, the SZ effect, diffuse dust emission from galactic cirrus, and random realizations of the CIB that include the effects of gravitational lensing. In the following subsections we describe how these various components are generated. The individual and combined signal components for the three SPIRE bands are shown in Figure 6.3 for one mock observation.

Instrument Noise

We use the model described in §6.2 to generate random instrument noise realizations for our mock observations. Specifically, we assume the noise fluctuations to be Gaussian and uncorrelated between map pixels, with per pixel uncertainties determined by the detectors' time stream variance and integration time in each pixel.

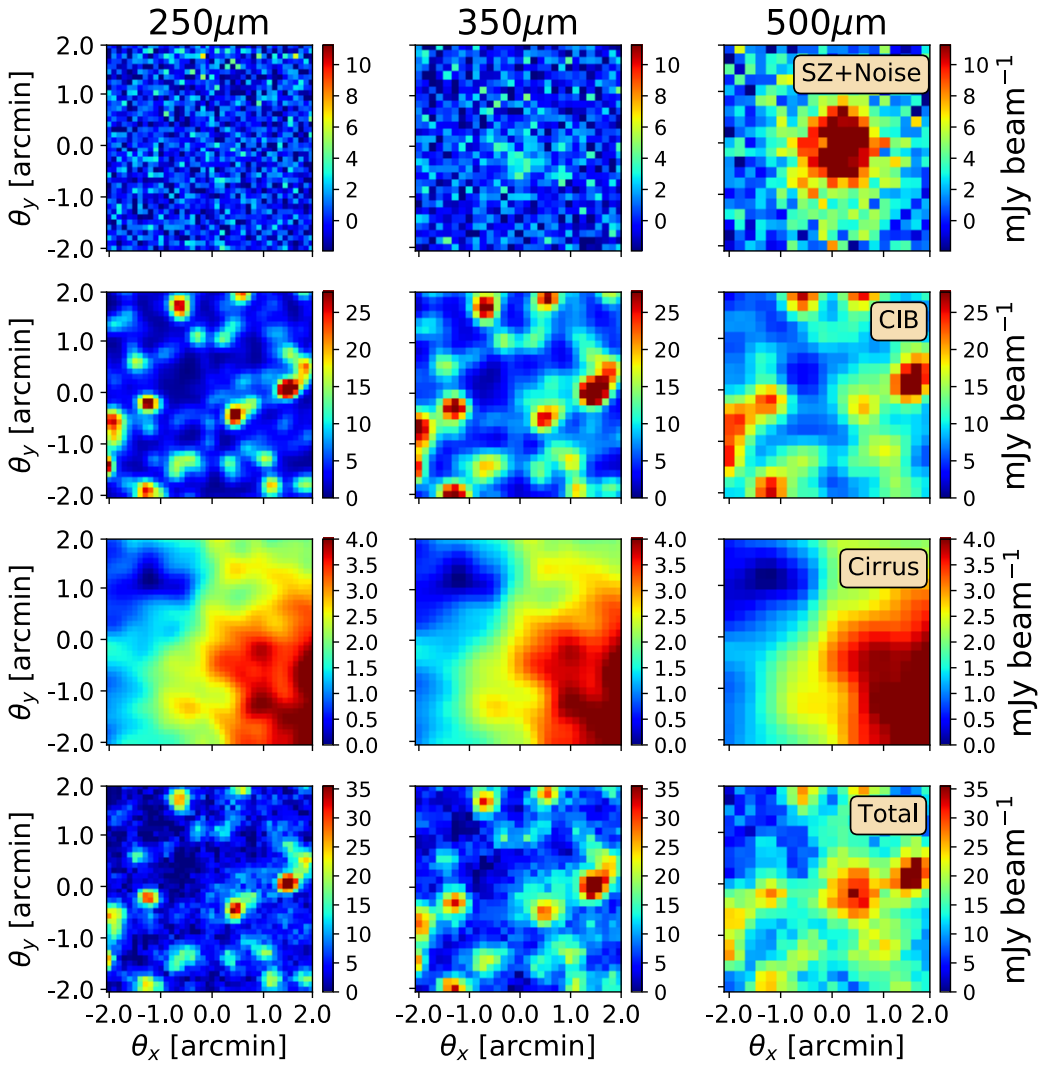

Figure 6.3: An example of one mock SPIRE observation of the galaxy cluster RX J1347.5-1145 in PSW (left), PMW (middle) and PLW (right).

The spatial scan pattern of SPIRE results in a non-uniform integration time across the field of view that is largest in the central region and decreases towards the edges [362].

Thermal SZ effect

To include the SZ effect signal in our mock images, we use the elliptical gNFW model described in §6.2 as a spatial template. The normalization of this template in the SPIRE bands is computed using `SZpack` based on the Bolocam measured surface brightness at 139 GHz and the *Chandra*-measured pressure-weighted temperature. The resulting images are then convolved with the appropriate PSF for each SPIRE

band.

Diffuse foregrounds

Initial fits to observed data in RX J1347.5-1145 with `PCAT-DE`, which assumed a mean signal level and not a more general diffuse component, produced image residuals with spatial structure on large angular scales. Data from IRAS [383] and *Planck* [384, 385] show emission with a consistent spatial and spectral signature, implying that the residual seen in our images is due primarily to thermal dust emission from galactic cirrus. However, the Planck beam size is very coarse compared to our map size, meaning only the largest modes are captured. Indeed, the *Planck*-interpolated maps were not sufficiently accurate as spatial templates to model the diffuse emission in the SPIRE maps. As `PCAT-DE` has a relatively low minimum source flux density threshold, residual diffuse emission can be misattributed to low-significance point sources. This was also observed using *Planck*-interpolated templates, motivating the Fourier component approach described in Section 6.3.

Random realizations of cirrus emission are created by drawing Gaussian random fields with angular power spectra consistent with the observed cirrus signal in RX J1347.5-1145. In particular, the power spectrum is assumed to scale as $P(k_\theta) \propto k_\theta^{-2.6}$ [289], with a normalization set by the measured amplitude of the large-scale power spectrum of the *Planck* dust template from [385], interpolated to the SPIRE footprint and extrapolated to SPIRE frequencies using the *Planck*-estimated parameters of a modified blackbody SED. As with the other astronomical signals included in our mock images, the resulting realizations are convolved with the PSF appropriate to each SPIRE band.

While the Planck maps show little spectral variation across the SPIRE field of view, there is a minor uncertainty introduced by fixing the spectrum in this manner. We estimated the impact of spectral variations by introducing a Gaussian-distributed random variation in the temperature and beta parameters and propagating those to changes in the surface brightness template. Since the templates are fixed to the values of τ in the Planck data, we do not vary those values here. Changing the parameters naturally changes the overall amplitude of the SED. We correct for this offset by calculating the amplitude change in PSW by taking the ratio of the fiducial brightness over the modified template. We then divide the modified template by this correction factor and subtract the result from the fiducial model in the other SPIRE bands. The maps are then mean subtracted, and the 68% confidence region

is calculated from the histogram of 100 map realizations. This region represents the uncertainty due to spectral variations. The results are aperture corrected to R2500 values to match the errors reported in Table 6.1, and are of order 10^{-4} MJy sr^{-1} for both changes in β and temperature. The squared sum of this contribution with the cirrus errors reported in Table 6.1 is negligible, and so we report only the larger of the two constituents.

CIB

The CIB, which is due primarily to DSFGs but also active galactic nuclei (AGN), is the brightest astrophysical source of emission at high galactic latitudes [386]. The depth of these observations and the angular resolution of SPIRE cause the image-space pixel fluctuations to be dominated by faint, undetected CIB sources rather than the instrument noise, commonly referred to as "confusion noise" [283]. Further, the vast majority of bright galaxies in the SPIRE images are not associated with the cluster, and are instead located behind it [387, 388]. Since RX J1347.5-1145 is an efficient gravitational lens [389, 375], most of the CIB sources in the SPIRE image have been deflected and magnified. As a result, it is necessary to consider not just the bright sources that can be detected individually, but also the undetected CIB sources, many of which have been lensed by the galaxy cluster.

We create mock observations of the CIB in the SPIRE bands using a two-step process wherein: bright sources individually detected by `PCAT-DE` are used to produce constrained CIB realizations and random realizations of the population of faint undetected sources are generated from an empirical model of the CIB, including the effects of gravitational lensing on sources behind RX J1347.5-1145. We describe these steps in detail below.

First, at flux densities above $2\sigma_{\rm conf}$, corresponding to 11.6, 12.6, and 13.6 mJy in PSW, PMW, and PLW [283], `PCAT-DE` detects sources with a completeness of $\gtrsim$ 90% and a false detection rate of $\lesssim$ 10% [100]. We thus expect sources above these thresholds in the `PCAT-DE` catalog to accurately describe the observed sky. Therefore, to create a single mock observation of the CIB, we extract the positions and flux densities of all catalog sources brighter than these thresholds in at least one SPIRE band. By populating the mocks with different realizations from `PCAT-DE`'s catalog ensemble, we ensure that the ensemble of mocks encodes the measured uncertainties from blind source extraction/photometry of the CIB.

Second, we generate random catalogs of the positions, SPIRE flux densities, and

redshifts of a set of CIB galaxies using the empirical "2SFM" model from B. et al. [286] and M. et al. [390] and S. et al. [391] therein. The flux densities are determined using an SED appropriate for the source type (starburst or main sequence galaxy), the redshift and the infrared luminosity. The 2SFM model produces SPIRE source distributions with the observed constraints.

All of the sources behind the galaxy cluster are gravitationally lensed according to the mass model derived in §6.2. To retain the correct CIB population statistics, we then remove all of the sources brighter than $2\sigma_{\rm conf}$ in at least one SPIRE band from the lensed source catalog. For a single mock, a full catalog of CIB sources is thus obtained from the combination of bright sources in the `PCAT-DE` catalog and fainter sources remaining in the B. et al. [286] catalog after this removal. The SPIRE PSF is then used to generate a mock observation from this catalog.

6.5 Sources of measurement errors

In this section, the components of the mock observations described in Section 6.4 are used to assess the impact of each type of signal (or noise) on the derived SZ results. We also isolate and study the effect of lensing on our inference by testing CIB realizations both with and without application of the lensing model.

Instrument noise

For Bolocam, we use the formalism described in [392] to produce random realizations that include both detector noise and fluctuations in the atmospheric emission. We note that the detector noise has an approximately flat spectrum, while the atmospheric fluctuations have a power spectrum that increases as a power law at low angular frequencies in the map.

For SPIRE, the contribution from instrument noise is isolated by generating random realizations from the SPIRE noise model, adding a fiducial SZ effect signal to each realization and fitting the signal template to the data within our analysis pipeline. This may be interpreted as an estimate of the raw sensitivity of the measurement, in the absence of other systematics. Quantitative estimates of the uncertainties on the $\langle d\mathrm{I}_b \rangle_{2500}$ due to instrument noise are given in Table 6.1.

Astrophysical contamination — Bolocam

The Bolocam images contain a small, but non-negligible, amount of signal from unwanted astrophysical contaminants. As originally described in [392], our noise model includes random realizations of the primary CMB anisotropies and CIB based

on their measured angular power spectra. In addition, the signal from the bright AGN in the BCG is modeled and removed according to the procedures detailed in [393]. For this analysis, we also add a contribution to the noise model due to the kinematic SZ effect signal resulting from the (unknown) bulk line of sight velocity of the galaxy cluster. Following the convention of [394], we assume a random velocity centered on zero with a standard deviation of 300 km sec^{-1} based on the simulations of [302]. In computing the signal from this velocity, we assume the ICM is isothermal with a temperature equal to the X-ray measured value of 17.4 keV. Quantitative estimates of the overall uncertainties on $\langle d\mathrm{I}_{\mathrm{Bcam}}\rangle_{2500}$ arising from these signals are given in Table 6.1.

Astrophysical contamination — SPIRE

The individual galaxies comprising the CIB contaminate our measurement of the SZ effect, particularly the spatially correlated emission arising due to gravitational lensing of the background population. Using the 100 mock CIB realizations described in Section 6.4, we estimate the associated uncertainty on the $\langle d\mathrm{I}_b\rangle_{2500}$ from the aggregate posterior for the values of $\mathrm{A}_b^{\mathrm{SZ}}$. Because the ensemble of CIB realizations is well approximated as a collection of independent, identically distributed draws from an underlying luminosity function, the aggregate posterior from these mocks should capture the effect of instrument noise, per-realization CIB model uncertainties, and any error due to intrinsic scatter from cosmic variance in the CIB. While the `B12` model does not include a clustering term, the fluctuations due to Poisson noise from the CIB dominate at the scales of interest to our map, and so is assumed to be negligible [290, see Figure 9].

To quantify the contribution of cirrus dust contamination to our error budget, we compute the scatter on the inferred $\langle d\mathrm{I}_b\rangle_{2500}$ from an ensemble of 100 lensed mocks without cirrus, and then compare against the same mock observations with cirrus included. This constrains the effect of diffuse cirrus emission to the upper limits given in Table 6.1.

Unlike in the Bolocam data, emission from the AGN in the BCG is relatively dim compared to the other signals in the SPIRE bands. Specifically, its brightness is approximately equal to the instrument noise per beam and almost two orders of magnitude dimmer than the integrated SZ effect signal within R_{2500} based on the SED fits from [395] and [393]. Therefore, we have not attempted to specifically model the AGN emission in the SPIRE data.

Gravitational lensing

Gravitational lensing has a significant impact on our analysis of the CIB and the resulting SZ effect constraints. Uncertainties on the $\langle dI_b \rangle_{2500}$ obtained from performing our analysis on lensed CIB mock realizations are presented in Table 6.1, and equal to 0.014 and 0.012 MJy sr^{-1} for PLW and PMW. In addition, we measure a significant bias in the value of the $\langle dI_b \rangle_{2500}$, equal to -0.019 and -0.021 MJy sr^{-1} in the two bands. For comparison, we also performed our analysis using unlensed CIB mock realizations, obtaining slightly higher $\langle dI_b \rangle_{2500}$ uncertainties of 0.016 and 0.014 MJy sr^{-1} and significantly smaller (and positive) biases equal to +0.007 MJy sr^{-1} in both bands.

These differences between the lensed and unlensed mocks are due primarily to the effect of "depletion" noted by [396], where a lack of CIB emission is observed within the strong lensing region near the center of the galaxy cluster [397, 304]. The effective subtraction of bright individual sources from the images using `PCAT-DE` further enhances this depletion. Thus, the level of CIB emission coincident with the SZ effect signal is slightly lower than what is observed in unlensed regions of the sky. In addition, the inferred $\langle dI_b \rangle_{2500}$ are biased low due to the on-average deficit of CIB emission coincident with the SZ effect signal. For our final analysis, we have corrected the values of $\langle dI_b \rangle_{2500}$ for the measured bias, which, as noted above, is due primarily to lensing.

If the effects of gravitational lensing are not included in the mocks, then the uncertainty on $\langle dI_b \rangle_{2500}$ attributed to variations in the CIB is slightly larger (0.016 MJy sr^{-1} for PLW and 0.014 MJy sr^{-1} for PMW). In addition, the magnitude of the mean bias on $\langle dI_b \rangle_{2500}$ is significantly reduced and its value is slightly positive (+0.007 MJy sr^{-1} for both PLW and PMW). Thus, the on-average decrease in CIB brightness near the galaxy cluster center due to gravitational lensing strongly biases the inferred $\langle dI_b \rangle_{2500}$ to lower values while also minimally decreasing the associated uncertainty.

Instrumental flux calibration

The SPIRE data are first calibrated according to the procedure described by [398]. This calibration is then adjusted using the empirical cross-calibration factors between *Planck*-HFI and SPIRE determined by B. et al. [399]. As detailed by B. et al. [399], the *Planck*-HFI calibration has a statistical uncertainty of 1.1 and 1.4 per cent at 545 and 857 GHz, along with an absolute uncertainty of $\simeq$ 2 per cent and $\simeq$ 5 per cent,

where the former is obtained from a measurement of the first two acoustic peaks in the primary CMB anisotropy power spectrum and the latter is obtained from the ESA2 planetary model of Uranus and the ESA3 planetary model of Neptune. In translating the *Planck*-HFI calibration to SPIRE, there is an additional $\simeq 4$ per cent uncertainty due to the SPIRE PSF calibration, along with a sub-per cent statistical uncertainty in the cross calibration. Adding these terms in quadrature, we estimate the absolute SPIRE calibration to be accurate to 4.6 per cent and 6.6 per cent for PLW and PMW. Furthermore, based on this calibration scheme, we expect a negligible correlation in the calibration uncertainty between PLW and PMW.

The Bolocam data are calibrated according to the procedure described in S. et al. [304], which is accurate to 1.7 per cent. In brief, the empirical model derived in [400] is used to correct for variations in atomospheric transmission based on observing conditions. The planetary model of G. and O. [401], corrected based on the *Planck*-HFI planetary brightness measurements of Planck Collaboration et al. [402], is then employed to determine the absolute calibration.

Additional instrumental calibrations

While there are additional sources of potential systematic errors related to instrumental calibration (e.g., the measured spectral bandpasses, PSF shape measurements, astrometric corrections), all of these are likely to be sub-dominant to the flux calibration accuracy. In addition, many of these potential sources of systematic error have already been subsumed into the flux calibration model, and are thus largely accounted for. We therefore do not explicitly include them in our overall error budget.

6.6 Results

`PCAT-DE` SPIRE results

The quantities $\langle d\mathrm{I}_{\mathrm{PMW}}\rangle_{2500}$ and $\langle d\mathrm{I}_{\mathrm{PLW}}\rangle_{2500}$ are inferred using a combination of the observed SPIRE data and the set of constrained mock observations. To ensure full coverage over the SZ posterior, we run 100 randomly initialized Markov chains on the observed data in parallel. Each chain is run for 5×10^6 samples, which are then thinned by a factor of 1000. The `PCAT-DE` model provides a reasonably good but imperfect description of the data, with typical reduced-χ^2 values, computed using the pixel-wise log-likelihood in Eq. (6.5) and the number of parameters at each step of the chain, in the range of 1.2–1.3.

Running several independent chains allows us to assess the level of convergence

Source	Uncertainty on $\langle dI_b \rangle_{2500}$ (MJy sr^{-1})	Notes
Bolocam		
Instrument+Atmosphere	0.008	See [392]
CMB+CIB	0.003	See S. et al. [392]
Kinematic SZE	0.007	Assuming ±300 km sec^{-1} S. and D. [302]
Absolute Calibration	0.003	±1.7 per cent, see S. et al. [304]
Total	0.012	Measured $\langle dI_{\rm Bcam} \rangle_{2500} = -0.153$ MJy sr^{-1}
SPIRE PLW		
Instrument	0.003	Mock SZ + Instrument noise run through pipeline
CIB	0.014	Mean bias of −0.019 MJy sr^{-1} (primarily due to lensing)
Cirrus	0.003	Mocks with and without cirrus have consistent posteriors
Absolute Calibration	0.003	±4.6 per cent based on B. [403]
Model fitting	0.002	Fixed mean signal level
Total	0.014	Measured $\langle dI_{\rm PLW} \rangle_{2500} = 0.104$ MJy sr^{-1}
SPIRE PMW		
Instrument	0.003	Mock SZ + Instrument noise run through pipeline
CIB	0.014	Mean bias of −0.019 MJy sr^{-1} (primarily due to lensing)
Cirrus	0.003	Mocks with and without cirrus have consistent posteriors
Absolute Calibration	0.003	±4.6 per cent based on B. [403]
Model fitting	0.002	Fixed mean signal level
Total	0.014	Measured $\langle dI_{\rm PMW} \rangle_{2500} = 0.037$ MJy sr^{-1}

Table 6.1: Sources of Measurement Error. While the CIB-related uncertainty on $\langle dI_b \rangle_{2500}$ is dominated by cosmic variance, a small amount is also related to fluctuations in the value of $N_{\rm src}$ in `PCAT-DE` (corresponding to 0.004 MJy sr^{-1} for PLW and 0.003 MJy sr^{-1} for PMW).

in our model. We check that our chains have completed the "burn-in" phase by visually inspecting the ensemble of trace plots and confirming that they are well mixed. We discard the first 50% of samples and combine the remaining ones from all chains to produce the aggregate posterior distribution of A_b^{SZ} values. The mean auto-correlation lengths from the collection of observed chains are $\langle\tau\rangle_{PMW} = 6$ (thinned) samples and $\langle\tau\rangle_{PLW} = 22$, which results in effective sample sizes large enough that Monte Carlo error on the results are $< 1\%$. We further validate that our chains are well mixed using the Gelman-Rubin statistic, also known as the potential scale reduction factor (PSRF) [404]. For the 100 randomly initialized chains run on the observed data, $\hat{R} = 1.14$ for PMW and $\hat{R} = 1.2$ for PLW, which suggests sufficient convergence to use the aggregated samples from all observed chains to compute an unbiased estimate of the posterior variance. By running several chains on individual mock observations, we measure $\hat{R} < 1.1$ and so we do not include any additional correction to the 100 chains run on the mocks, where only one chain is assigned to each mock realization.

Figure 6.4 shows the marginalized posterior on A_{PMW}^{SZ} and A_{PLW}^{SZ}, along with the inferred number of point-like sources. Both the mock and observed posteriors on A_b^{SZ} in Fig. 6.4 are uncorrected for the bias due primarily to lensing depletion (see Section 6.5). In both the observed data and mock realizations, A_{PMW}^{SZ} and A_{PLW}^{SZ} are positively correlated, with Pearson correlation coefficients of +0.18 and +0.52, respectively. Since the SZ effect is independently observed across different SPIRE bands, the correlation between the two is due to the presence of correlations in the unresolved source population that are degenerate with the SZ effect signal. In the observed data, A_{PMW}^{SZ} and N_{src} appear uncorrelated. However, A_{PLW}^{SZ} is anti-correlated with the number of detected CIB sources with a correlation coefficient of -0.25. This comports with the coarser PLW angular resolution, which results in the CIB being more spatially degenerate with extended emission from the SZ effect. The anti-correlation between A_{PLW}^{SZ} and N_{src} is washed out in the aggregate mock posterior, where intrinsic scatter from cosmic variance in the CIB is dominant.

Probing the cross-model covariance between the SZ effect and the union of point source models with varying N_{src} is straightforward in the framework of probabilistic cataloging, since for each SZ posterior sample there is an associated catalog of CIB sources. To assess the impact of our imperfect knowledge of N_{src}, we take the quadrature difference of the fully marginalized uncertainties on A_b^{SZ} with those from the conditional uncertainty assuming the median inferred N_{src}, i.e.,

$\sqrt{\sigma^2(\mathrm{A}_b^{SZ}) - \sigma^2(\mathrm{A}_b^{SZ}|\mathrm{med}(\mathrm{N_{src}}))}$. By this metric, our results suggest that imperfect knowledge of $\mathrm{N_{src}}$ results in an uncertainty on $\langle d\mathrm{I}_b\rangle_{2500}$ that is ~ 4 times smaller than the uncertainty due to cosmic variance in the CIB.

Additional uncertainties due to the use of a fixed mean level estimated from the larger $10' \times 10'$ maps, inflating the constraints by and the convergence of the Markov chains are added as a random Gaussian component to the set of corrected samples, inflating the constraints on $\langle d\mathrm{I_{PMW}}\rangle_{2500}$ and $\langle d\mathrm{I_{PLW}}\rangle_{2500}$ by 17% and 20%, respectively. After correcting for the lensing depletion bias, we measure the SZ effect brightness to be $\langle d\mathrm{I_{PLW}}\rangle_{2500} = 0.104 \pm 0.014$ MJy/sr and $\langle d\mathrm{I_{PMW}}\rangle_{2500} = 0.037 \pm 0.013$ MJy/sr, corresponding to 7.4σ and 2.9σ detections of the SZ effect, respectively.

Finally, the difference between the posterior SZ effect samples and the *Chandra*+Bolocam SZ effect amplitude injected into the mocks defines a bias distribution, which is used to propagate estimates from the observed (biased) value of $\mathrm{A}_b^{\mathrm{SZ}}$ to the underlying $\langle d\mathrm{I}_b\rangle_{2500}$ (see Figure 6.5 and also the discussion in Section 6.5). To test for model dependence in the bias, mock realizations with different injected SZ effect amplitudes were analyzed. For amplitudes spanning the range $[0, 2]$ times the nominal *Chandra*+Bolocam value, the mean bias is found to be constant within our measurement precision. Thus, it is valid to apply the linear bias correction to any measured values of $\mathrm{A}_b^{\mathrm{SZ}}$ within this range.

SZ spectral likelihood analysis

To obtain constraints on the average Comptonization parameter $\langle y\rangle_{2500}$ and ICM temperature $\langle \mathrm{T_{sz}}\rangle_{2500}$, we perform a likelihood analysis of our observed $\langle d\mathrm{I}_b\rangle_{2500}$ over a range of $\langle y\rangle_{2500}$ and $\langle \mathrm{T_{sz}}\rangle_{2500}$ values, with $\langle \mathrm{T_{sz}}\rangle_{2500}$ varying between $[0, 60]$ keV. The model $\langle d\mathrm{I}_b\rangle_{2500}$ values for each $\langle y\rangle_{2500}/\langle \mathrm{T_{sz}}\rangle_{2500}$ grid pair are computed using `SZpack` [334, 335]. In all cases, the effective band center of each $\langle d\mathrm{I}_b\rangle_{2500}$ used in this calculation was obtained by weighting the measured spectral bandpass of the instrument by the temperature-dependent shape of the SZ effect spectrum at the given value of $\langle \mathrm{T_{sz}}\rangle_{2500}$. The likelihood assumes independence between the Bolocam and SPIRE measurements of $\langle d\mathrm{I}_b\rangle_{2500}$, which is a good approximation given that the CIB and cirrus signals are negligible in the Bolocam image and the primary CMB anisotropies and kinematic SZ effect signal are negligible in the SPIRE images.

Our results using `PCAT-DE` demonstrate that the inferred distribution of $\langle d\mathrm{I_{PLW}}\rangle_{2500}$ and $\langle d\mathrm{I_{PMW}}\rangle_{2500}$ values is non-Gaussian to such a degree that approximating the PDF with a Gaussian covariance matrix significantly distorts the final constraints.

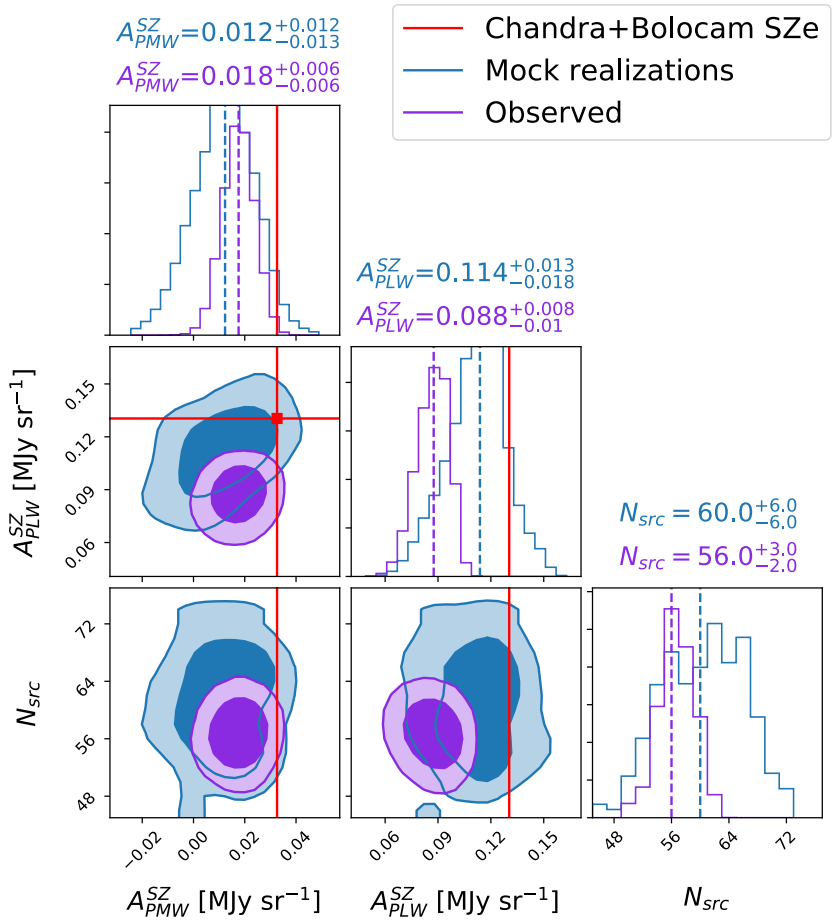

Figure 6.4: 68 and 95 per cent credible regions for A^{SZ}_{PMW} and A^{SZ}_{PLW}. The posteriors are composed from the distribution of template amplitudes sampled by `PCAT-DE`. Credible regions are shown for the observed RX J1347.5-1145 SPIRE data (purple) and mock SPIRE maps with instrument noise only (black), instrument noise + unlensed CIB (green) and instrument noise + lensed CIB (blue). The red star indicates the injected SZ effect surface brightness values in the mock observations.

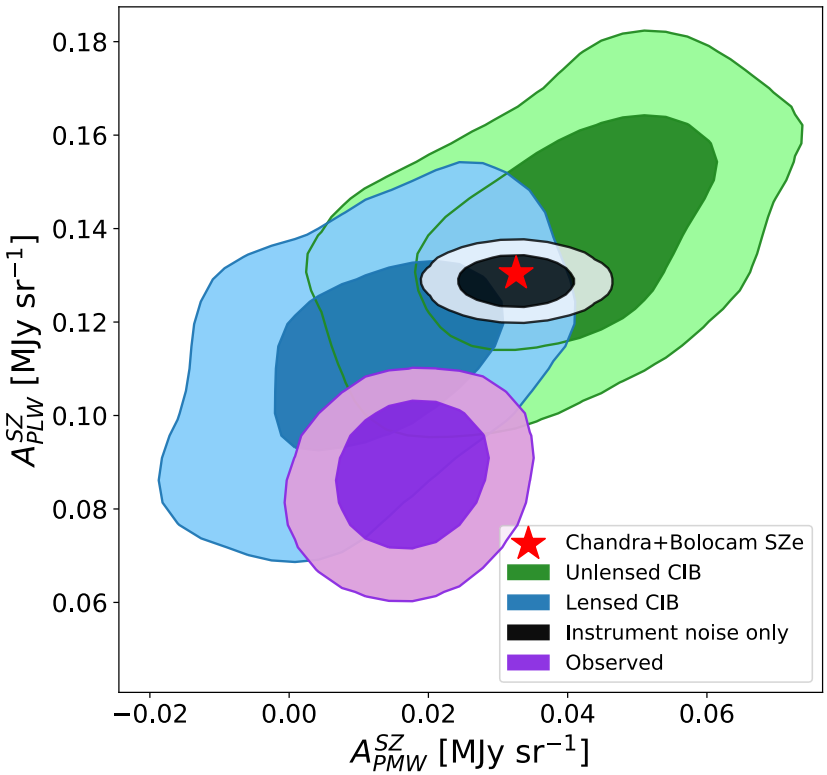

Figure 6.5: 68 and 95 per cent credible regions for A^{SZ}_{PMW} and A^{SZ}_{PLW}. The posteriors are composed from the distribution of template amplitudes sampled by `PCAT-DE`. Credible regions are shown for the observed RX J1347.5-1145 SPIRE data (purple) and mock SPIRE maps with instrument noise only (black), instrument noise + unlensed CIB (green) and instrument noise + lensed CIB (blue). The red star indicates the injected SZ effect surface brightness values in the mock observations.

Instead, when computing the posterior $p(y, T|d\mathrm{I})$ with Bayes' rule:

$$p(y, T|d\mathrm{I}) \propto p(d\mathrm{I}|y, T)p(y, T), \tag{6.9}$$

the likelihood $p(d\mathrm{I}|y, T)$ is evaluated directly from the gridded, bias-corrected samples of $\langle d\mathrm{I}_{\mathrm{PLW}}\rangle_{2500}$ and $\langle d\mathrm{I}_{\mathrm{PMW}}\rangle_{2500}$ (also referred to as the "empirical PDF"). The prior $p(y, T)$ places constraints on the range of spectra that are consistent with our underlying model. In our current analysis, no prior is placed on the amplitudes of the SZ templates, nor on their color, at the map-fitting level. While this helps in identifying and correcting for systematic effects, it does mean a subset of samples from the posterior fall outside the range of $\langle d\mathrm{I}_b\rangle_{2500}$ computed by `SZPack` for $0 \leq \langle \mathrm{T_{sz}}\rangle_{2500} \leq 60$ keV. These samples, which reside on the tails of the PDF, comprise only $\sim$ 7% of the full sample and are given zero weight in the final constraints.

Parameter	MAP value	68% credible interval	95% credible interval
$\langle y \rangle_{2500}$ [$\times 10^{-4}$]	1.56	$1.42 < \langle y \rangle_{2500} < 1.63$	$1.29 < \langle y \rangle_{2500} < 1.71$
$\langle \mathrm{T_{sz}} \rangle_{2500}$ [keV]	22.4	$10.4 < \langle \mathrm{T_{sz}} \rangle_{2500} < 33.0$	$0.0 < \langle \mathrm{T_{sz}} \rangle_{2500} < 39.5$

Table 6.2: Fitted values of $\langle y \rangle_{2500}$ and $\langle \mathrm{T_{sz}} \rangle_{2500}$ for RX J1347.5-1145.

The measured $\langle d\mathrm{I}_b \rangle_{2500}$ and constrained set of SZ effect spectra are shown in Figure 6.6, and the posterior on $\langle y \rangle_{2500}$ and $\langle \mathrm{T_{sz}} \rangle_{2500}$ is shown in Figure 6.7. Looking separately at the contributions from Bolocam and SPIRE, one can see that the individual measurements suggest different constraints on $\langle y \rangle_{2500}$ and $\langle \mathrm{T_{sz}} \rangle_{2500}$. The Bolocam data constrain the parameters along a positively correlated axis in the $\langle y \rangle_{2500}$/$\langle \mathrm{T_{sz}} \rangle_{2500}$ plane. The $\langle y \rangle_{2500}$/$\langle \mathrm{T_{sz}} \rangle_{2500}$ constraints using the empirical SZ posterior from SPIRE data have a more complicated structure. The preference of a smaller value of $\langle y \rangle_{2500}$ compared to Bolocam is largely driven by the value of $\langle d\mathrm{I_{PLW}} \rangle_{2500}$, which is approximately $\sim 2\sigma$ lower than the value expected from the combination of Bolocam and *Chandra*. While the Bolocam data are in good agreement with the maximum *a posteriori* model estimate, the delta log-likelihood from the SPIRE points is $\Delta \ln \mathcal{L} = -2.4$, indicating tension with the spectral model.

The joint posterior contains the most probability mass near the tails of the SPIRE-only posterior, and when combined with Bolocam data leads to bimodal constraints. This is due to the positive covariance between $\langle d\mathrm{I_{PLW}} \rangle_{2500}$ and $\langle d\mathrm{I_{PMW}} \rangle_{2500}$ combined with the measured values falling on opposite sides of the SZ effect spectrum for that range of $\langle y \rangle_{2500}$/$\langle \mathrm{T_{sz}} \rangle_{2500}$. That is, the measured values scatter in an anti-correlated manner relative to the model while the covariance matrix implies a positive correlation, which leads to two local peaks in $p(d\mathrm{I}|y, T)$. There is mild preference for the mode with larger values of the pair.

Our maximum *a posteriori* (MAP) estimate for the ICM temperature is $\langle \mathrm{T_{sz}} \rangle_{2500}$= 22.4 keV, with a 68% highest posterior density credible interval (HDPI) of $10.4 < \langle \mathrm{T_{sz}} \rangle_{2500} < 33.0$ keV (see Figure 6.8). This estimate is consistent with the X-ray predicted, pressure-weighted temperature of $\langle \mathrm{T_{x,pw}} \rangle_{2500} = 17.4 \pm 2.3$ keV, indicating good agreement between measurements.

To understand the sensitivity of our data in the absence of the bimodality noted above, we also repeat the same analysis using the $\langle d\mathrm{I_{PLW}} \rangle_{2500}$ and $\langle d\mathrm{I_{PMW}} \rangle_{2500}$ values determined from the combination of Bolocam and *Chandra* (i.e., $\langle d\mathrm{I_{PLW}} \rangle_{2500}$ = 0.134 MJy sr^{-1} and $\langle d\mathrm{I_{PMW}} \rangle_{2500}$ = 0.033 MJy sr^{-1}). This eliminates the issue of positively-covariant values falling on opposite sides of the SZ effect spectrum,

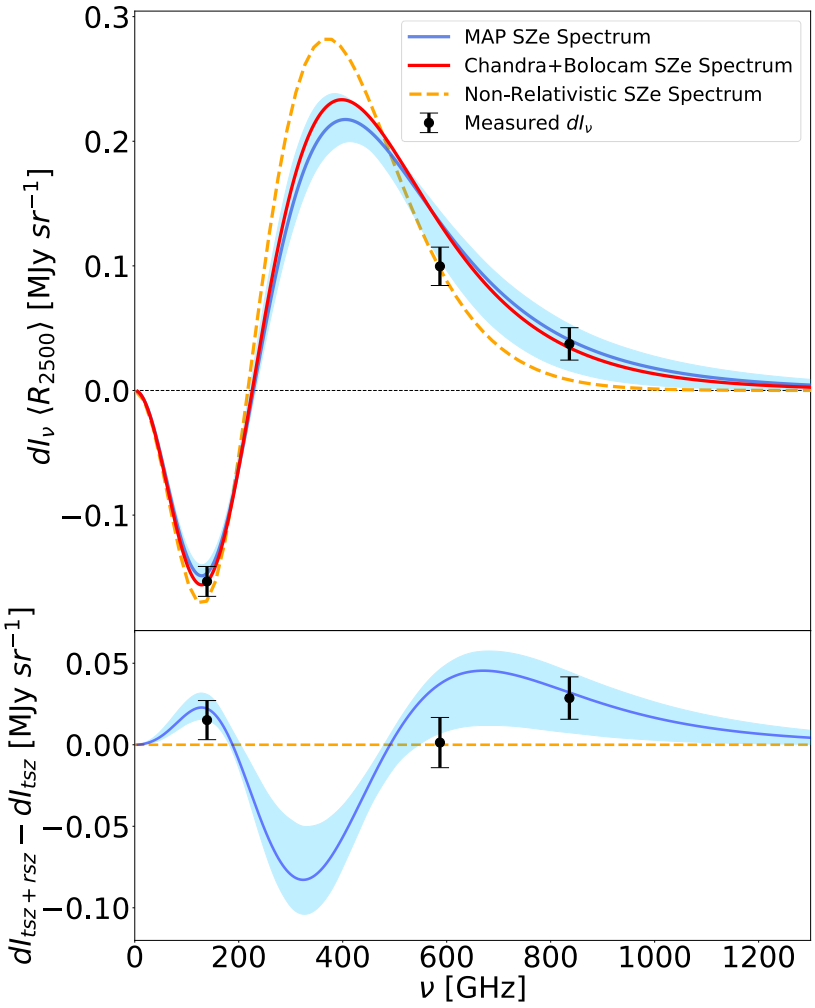

Figure 6.6: Top: SZ effect spectrum for the Bolocam+SPIRE MAP $\langle y\rangle_{2500}$ and $\langle \mathrm{T_{sz}}\rangle_{2500}$ values (thin blue), along with the spectrum obtained from the Bolocam $\langle d\mathrm{I_{Bcam}}\rangle_{2500}$ and *Chandra* $\langle \mathrm{T_{x,pw}}\rangle_{2500}$ (red). Also shown is the best-fit non-relativistic SZ effect spectrum (dashed orange). The blue shaded region bounds the range of SZ effect spectra within the 68 per cent credible region of $\langle y\rangle_{2500}$ and $\langle \mathrm{T_{sz}}\rangle_{2500}$. The $\langle d\mathrm{I}_b\rangle_{2500}$ values are shown as black points, with reported errors. Bottom: Residual spectra after subtraction of the non-relativistic SZ effect spectrum. Note that, due to the negative sign of the SZ effect in the Bolocam band, a measured $\langle d\mathrm{I_{Bcam}}\rangle_{2500}$ below the MAP spectrum indicates a larger rather than a smaller SZ effect amplitude. The PSW band center is at approximately 1200 GHz, where there is very little SZ effect signal. (Figure Credit: Victoria Butler)

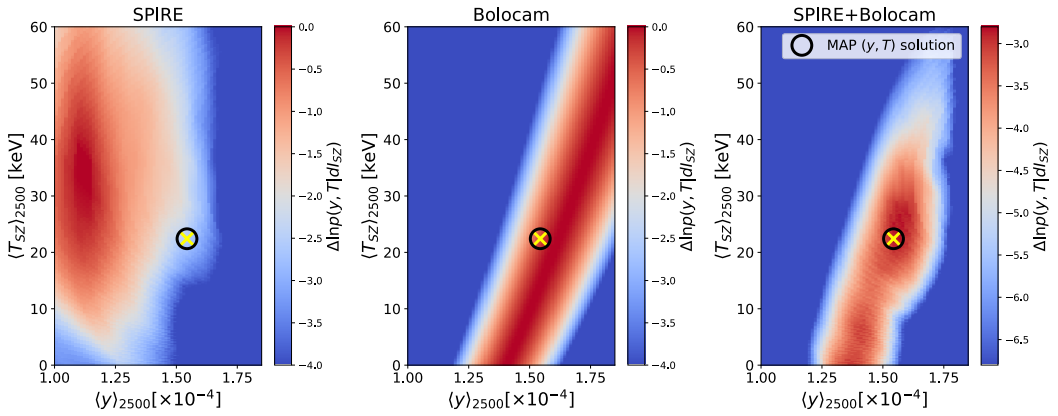

Figure 6.7: The joint posterior distribution (right) is computed as the product of the SPIRE posterior (left), which is sampled directly with `PCAT-DE`, and the Bolocam likelihood (middle) which is assumed to be Gausisan-distributed with respect to both parameters. The black circle + yellow cross indicates the position of the maximum *a posteriori* (MAP) estimate of $\langle y \rangle_{2500}$ and $\langle \mathrm{T_{sz}} \rangle_{2500}$. Log-probabilities are plotted for visualization purposes.

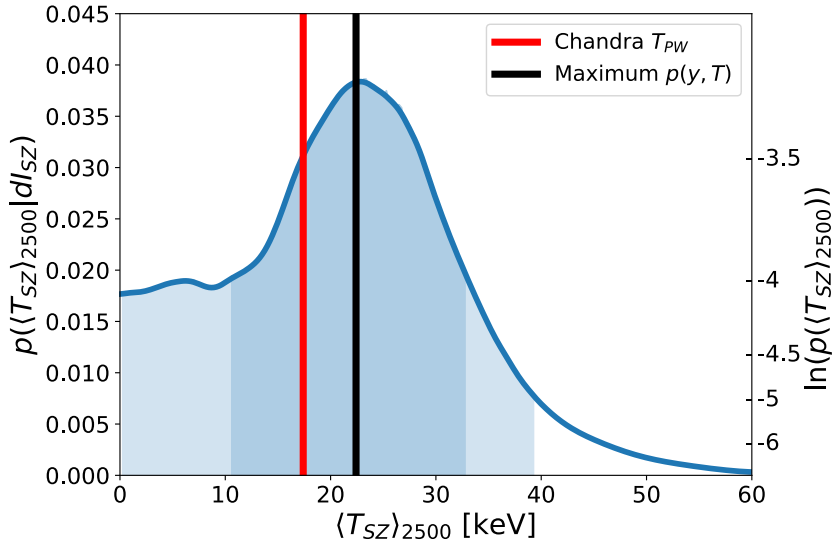

Figure 6.8: 1D marginalized probability distribution for cluster temperature $\langle \mathrm{T_{sz}} \rangle_{2500}$, with joint MAP temperature (black) and credible intervals shaded in blue.

and removes the bimodality, while fully preserving the noise as characterized by the empirical PDF. Under these conditions, the posterior is single-peaked and the constraints marginalized over $\langle y \rangle_{2500}$ yield a MAP $\langle \mathrm{T_{sz}} \rangle_{2500} = 12.3$ keV with a 68 per cent credible interval of $5.8 < \langle \mathrm{T_{sz}} \rangle_{2500} < 20.5$ keV. This represents a reduction in the range of the credible interval by 35 per cent compared to the observed data.

If we instead repeat the analysis using the observed values of $\langle d\mathrm{I}_b \rangle_{2500}$, but with randomized samples to remove the correlation between $\langle d\mathrm{I_{PLW}} \rangle_{2500}$ and $\langle d\mathrm{I_{PMW}} \rangle_{2500}$ in the empirical PDF, then we find a 68 per cent credible interval of $5.1 < \langle \mathrm{T_{sz}} \rangle_{2500} < 20.2$ keV. This is nearly identical to the interval obtained using *Chandra*+Bolocam predicted $\langle d\mathrm{I}_b \rangle_{2500}$ values to eliminate the bimodality, and suggests that the $\langle d\mathrm{I_{PLW}} \rangle_{2500}$/$\langle d\mathrm{I_{PMW}} \rangle_{2500}$ correlation similarly degrades our constraints on $\langle \mathrm{T_{sz}} \rangle_{2500}$.

As illustrated in Table 6.1, CIB fluctuations dominate the uncertainties on $\langle d\mathrm{I_{PLW}} \rangle_{2500}$ and $\langle d\mathrm{I_{PMW}} \rangle_{2500}$. To quantitatively assess the impact of the CIB on our derived $\langle \mathrm{T_{sz}} \rangle_{2500}$ constraints, we also perform our analysis on the SZ + instrument noise only mock posterior (see Figure 6.5), finding a MAP estimate of $\langle \mathrm{T_{sz}} \rangle_{2500}$=17.1 keV with a 68 per cent credible interval of $12.8 < \langle \mathrm{T_{sz}} \rangle_{2500} < 19.0$ keV.

6.7 Discussion

We have combined observations from SPIRE, Bolocam, and *Planck* to constrain the thermodynamic properties of the ICM in the galaxy cluster RX J1347.5-1145 through a measurement of the rSZ effect. In order to probe the desired SZ effect signal, we accounted for significant contamination from unwanted astrophysical components, in particular the CIB. Not only is the CIB brighter than the SZ effect signal in the SPIRE bands, but the individual sources that comprise it have a range of SEDs. In addition, most CIB sources reside behind RX J1347.5-1145 and have thus been deflected and magnified due to gravitational lensing. To properly characterize the effects of the CIB and other relevant signals such as galactic cirrus emission, we developed a mock observation pipeline to fully assess the uncertainties and biases in measuring the SZ effect brightness $\langle d\mathrm{I}_b \rangle_{2500}$. Our mock pipeline includes a mass model derived from *HST*, allowing us to accurately capture the effects of gravitational lensing, which we found to be the largest source of bias on the derived values of $\langle d\mathrm{I}_b \rangle_{2500}$.

Rather than treating each astrophysical source in isolation, we employ an extension of the forward modeling framework, `PCAT-DE`, to simultaneously constrain the signal

components. This enables us to make more efficient use of the available information in the observed data while robustly capturing correlated, non-Gaussian uncertainties on $\langle dI_b \rangle_{2500}$. We propagate samples from the PCAT-DE posterior to estimates of $\langle dI_b \rangle_{2500}$ and use the resulting empirical PDF to constrain the Comptonization parameters. This was important in our analysis — we discovered that approximating the SPIRE posterior with a Gaussian covariance matrix significantly biased the final constraints.

As noted in Section 6.6, our constraints on $\langle T_{sz} \rangle_{2500}$ are strongly influenced by the relatively low measured value of $\langle dI_{PLW} \rangle_{2500}$. A large positive correlation between the measured $\langle dI_{PMW} \rangle_{2500}$ and $\langle dI_{PLW} \rangle_{2500}$ values, coupled with them falling on opposite sides of the SZ effect spectrum for intermediate temperatures, results in a bimodal posterior for $\langle T_{sz} \rangle_{2500}$ and a significant expansion of the 68 per cent credible interval range (22.6 keV compared to 14.7 keV when a more likely value of $\langle dI_{PLW} \rangle_{2500}$ is assumed). We have not been able to identify any significant biases in the measured $\langle dI_{PLW} \rangle_{2500}$, but it is possible that one or more do exist. For instance, the particular spatial distribution of astrophysical signals toward RX J1347.5-1145 may result in PCAT-DE incorrectly associating some of the PLW SZ effect signal with CIB sources. It is also possible that some unmodeled astrophysical components are significant enough to produce a bias, such as the excess SZ effect signal associated with an ICM shock SE of the cluster center [312, 355, 316, 358]. RX J1347.5-1145 also has a pronounced cool core, giving rise to large temperature variations as a function of radius within the ICM. This invalidates the assumption of isothermality along the line of sight, and causes a slight change in the SZ spatial structure between the Bolocam and SPIRE observing bands. Since we are utilizing Bolocam for the SZ effect spatial template, this could change the location and significance of the actual SZ effect "peak". Another possibility is that the unexplained cluster-coincident emission with a dust-like SED noted by [345] has not been fully described by the combination of the PCAT-DE source catalog and diffuse Fourier component model.

Applying our analysis method to a larger sample of galaxy clusters will allow us to better assess if the somewhat low measured value of $\langle dI_{PLW} \rangle_{2500}$ is due to a statistical fluctuation, something particular to RX J1347.5-1145, or a more general bias in our methodology. Multi-probe SPIRE, Bolocam, *HST*, *Chandra*, and *Planck* observations exist for an additional 20 galaxy clusters, and another 18 objects can be studied using data from the ACT survey in place of Bolocam [405]. In addition to enabling checks for subtle, emergent systematics, this sample

of 39 galaxy clusters will allow us to place tight constraints on the ensemble-average temperature measured using the rSZ effect. Taking the raw sensitivity from this study in the absence of a bimodal $\langle T_{sz} \rangle_{2500}$ posterior, we estimate that the sample mean temperature will be measured to a precision of approximately 1 keV. By comparing with pressure-weighted X-ray temperatures, an empirical temperature calibration of *Chandra*and/or XMM-*Newton* should thus be possible. This level of sensitivity would also allow us to probe the expected differences between rSZ effect and X-ray emission-weighted temperatures due to spatial and thermal inhomogeneities in the ICM [370].

Looking forward, potential future wide-field and high angular resolution facilities such as AtLAST and CSST will provide better decoupling of the SZ effect from unwanted astrophysical signals, resulting in higher signal-to-noise ratio rSZ effect measurements of the ICM temperature while probing smaller physical scales. These data would complement future X-ray facilities and open exciting new areas of study, such as routine temperature measurements in the highest redshift galaxy clusters and resolved imaging of the hottest shock-heated gas produced in major merger events. In this work, we have demonstrated an analysis framework that can help maximize the science reach of these facilities.

References

[322] N. K. In: *MNRAS* 222 (Sept. 1986), pp. 323–345. DOI: 10.1093/mnras/222.2.323.

[323] H. Z. In: *arXiv e-prints*, astro-ph/0312165 (Dec. 2003), astro–ph/0312165. arXiv: astro-ph/0312165 [astro-ph].

[324] B. R. M. and P. E. J. N. In: *ARA&A* 45.1 (Sept. 2007), pp. 117–175. DOI: 10.1146/annurev.astro.45.051806.110625. arXiv: 0709.2152 [astro-ph].

[325] M. M. and A. V. In: *PhysRep* 443.1 (May 2007), pp. 1–53. DOI: 10.1016/j.physrep.2007.01.001. arXiv: astro-ph/0701821 [astro-ph].

[326] G. W. P. et al. In: *Space Sci. Rev* 215.2, 25 (Feb. 2019), p. 25. DOI: 10.1007/s11214-019-0591-0. arXiv: 1902.10837 [astro-ph.CO].

[327] C. L. S. In: *Reviews of Modern Physics* 58.1 (Jan. 1986), pp. 1–115. DOI: 10.1103/RevModPhys.58.1.

[328] M. M. et al. In: *ApJ* 843.1, 28 (July 2017), p. 28. DOI: 10.3847/1538-4357/aa7740. arXiv: 1702.05094 [astro-ph.CO].

[329] T. H. R. et al. In: *Space Sci. Rev* 177.1-4 (Aug. 2013), pp. 195–245. DOI: 10.1007/s11214-013-9983-8. arXiv: 1303.3286 [astro-ph.CO].

[330] M. M. et al. In: *ApJL* 567.1 (Mar. 2002), pp. L27–L31. DOI: 10.1086/339619. arXiv: astro-ph/0110468 [astro-ph].

[331] R. A. S. and Ya. B. Z. In: *Comments on Astrophysics and Space Physics* 4 (Nov. 1972), p. 173.

[332] E. L. W. In: *ApJ* 232 (Sept. 1979), pp. 348–351. DOI: 10.1086/157294.

[333] N. I., Y. K., and S. N. In: *ApJ* 502.1 (July 1998), pp. 7–15. DOI: 10.1086/305876. arXiv: astro-ph/9712289 [astro-ph].

[334] J. C. et al. In: *MNRAS* 426.1 (Oct. 2012), pp. 510–530. DOI: 10.1111/j.1365-2966.2012.21741.x. arXiv: 1205.5778 [astro-ph.CO].

[335] J. C. et al. In: *MNRAS* 430.4 (Apr. 2013), pp. 3054–3069. DOI: 10.1093/mnras/stt110. arXiv: 1211.3206 [astro-ph.CO].

[336] A. C. F. et al. In: *MNRAS* 418.4 (Dec. 2011), pp. 2154–2164. DOI: 10.1111/j.1365-2966.2011.19402.x. arXiv: 1105.5025 [astro-ph.CO].

[337] L. E. B. et al. In: *ApJS* 216.2, 27 (Feb. 2015), p. 27. DOI: 10.1088/0067-0049/216/2/27. arXiv: 1409.0850 [astro-ph.CO].

[338] Planck Collaboration et al. In: *A&A* 550, A131 (Feb. 2013), A131. DOI: 10.1051/0004-6361/201220040. arXiv: 1207.4061 [astro-ph.CO].

[339] M. G. H. et al. In: *ApJ* 508.1 (Nov. 1998), pp. 25–43. DOI: 10.1086/306379. arXiv: astro-ph/9806167 [astro-ph].

[340] M. Z. et al. In: *A&A* 518, L16 (July 2010), p. L16. DOI: 10.1051/0004-6361/201014685. arXiv: 1005.3824 [astro-ph.CO].

[341] D. A. P. et al. In: *MNRAS* 416.1 (Sept. 2011), pp. 302–310. DOI: 10.1111/j.1365-2966.2011.19037.x. arXiv: 1105.4271 [astro-ph.CO].

[342] D. A. P. and S. C. In: *MNRAS* 424.1 (July 2012), pp. L49–L53. DOI: 10.1111/j.1745-3933.2012.01284.x. arXiv: 1207.3729 [astro-ph.CO].

[343] M. Z. et al. In: *ApJ* 749.2, 114 (Apr. 2012), p. 114. DOI: 10.1088/0004-637X/749/2/114. arXiv: 1202.0029 [astro-ph.CO].

[344] G. H. In: *A&A* 596, A61 (Dec. 2016), A61. DOI: 10.1051/0004-6361/201629726. arXiv: 1701.09020 [astro-ph.CO].

[345] J. E. et al. In: *MNRAS* 476.3 (May 2018), pp. 3360–3381. DOI: 10.1093/mnras/sty327. arXiv: 1709.01187 [astro-ph.CO].

[346] D. H. H. et al. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 11445. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Dec. 2020, 1144522, p. 1144522. DOI: 10.1117/12.2561893.

[347] G. W. W. et al. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 11453. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Dec. 2020, 1145302, p. 1145302. DOI: 10.1117/12.2562331.

[348] CCAT-Prime collaboration et al. In: *arXiv e-prints*, arXiv:2107.10364 (July 2021), arXiv:2107.10364. arXiv: 2107.10364 [astro-ph.CO].

[349] P. D. K. et al. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 11445. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Dec. 2020, 114452F, 114452F. DOI: 10.1117/12.2561315. arXiv: 2011.07974 [astro-ph.IM].

[350] S. P. In: 53.20 (July 2014), pp. 4431–4439. DOI: 10.1364/AO.53.004431.

[351] E. P. et al. In: *ApJ* 552.1 (May 2001), pp. 42–48. DOI: 10.1086/320447. arXiv: astro-ph/0012309 [astro-ph].

[352] S. W. A., R. W. S., and A. C. F. In: *MNRAS* 335.2 (Sept. 2002), pp. 256–266. DOI: 10.1046/j.1365-8711.2002.05554.x. arXiv: astro-ph/0111368 [astro-ph].

[353] M. G. and S. S. In: *A&A* 427 (Nov. 2004), pp. L9–L12. DOI: 10.1051/0004-6361:200400086. arXiv: astro-ph/0409627 [astro-ph].

[354] A. B. M. et al. In: *MNRAS* 449.1 (May 2015), pp. 199–219. DOI: 10.1093/mnras/stv219. arXiv: 1502.06020 [astro-ph.CO].

[355] B. S. M. et al. In: *ApJ* 716.1 (June 2010), pp. 739–745. DOI: 10.1088/0004-637X/716/1/739.

[356] T. J. P. et al. In: *ApJ* 770.2, 112 (June 2013), p. 112. DOI: 10.1088/0004-637X/770/2/112. arXiv: 1203.2175 [astro-ph.CO].

[357] S. U. et al. In: *ApJ* 866.1, 48 (Oct. 2018), p. 48. DOI: 10.3847/1538-4357/aadd9d. arXiv: 1808.09232 [astro-ph.HE].

[358] L. D. M., E. C., and T. M. In: *MNRAS* 487.3 (Aug. 2019), pp. 4037–4056. DOI: 10.1093/mnras/stz1550. arXiv: 1812.01034 [astro-ph.CO].

[359] R. E. J. et al. In: *ApJ* 751.2, 95 (June 2012), p. 95. DOI: 10.1088/0004-637X/751/2/95. arXiv: 1106.3489 [astro-ph.CO].

[360] M. J. G. et al. In: *A&A* 518, L3 (July 2010), p. L3. DOI: 10.1051/0004-6361/201014519. arXiv: 1005.5123 [astro-ph.IM].

[361] S. J. O. et al. In: *A&A* 518, L21 (July 2010), p. L21. DOI: 10.1051/0004-6361/201014697. arXiv: 1005.2184 [astro-ph.CO].

[362] S. J. O. et al. In: *MNRAS* 424.3 (Aug. 2012), pp. 1614–1635. DOI: 10.1111/j.1365-2966.2012.20912.x. arXiv: 1203.2562 [astro-ph.CO].

[363] L. L. et al. In: *MNRAS* 409.1 (Nov. 2010), pp. 83–91. DOI: 10.1111/j.1365-2966.2010.17771.x. arXiv: 1010.0020 [astro-ph.CO].

[364] D. N., A. V. K., and A. V. In: *ApJ* 668.1 (Oct. 2007), pp. 1–14. DOI: 10.1086/521328. arXiv: astro-ph/0703661 [astro-ph].

[365] M. A. et al. In: *A&A* 517, A92 (July 2010), A92. DOI: 10.1051/0004-6361/200913416. arXiv: 0910.1234 [astro-ph.CO].

[367] N. G. C. et al. In: *ApJ* 806.1, 18 (June 2015), p. 18. DOI: 10.1088/0004-637X/806/1/18. arXiv: 1406.2800 [astro-ph.CO].

[368] A. B. M. et al. In: *MNRAS* 440.3 (May 2014), pp. 2077–2098. DOI: 10.1093/mnras/stu368. arXiv: 1402.6212 [astro-ph.CO].

[369] S. T. K. et al. In: *MNRAS* 386.4 (June 2008), pp. 2110–2114. DOI: 10.1111/j.1365-2966.2008.13183.x. arXiv: 0706.3668 [astro-ph].

[370] E. L. et al. In: *MNRAS* 493.3 (Apr. 2020), pp. 3274–3292. DOI: 10.1093/mnras/staa450. arXiv: 1912.07924 [astro-ph.CO].

[371] A. B. M. et al. In: *MNRAS* 456.4 (Mar. 2016), pp. 4020–4039. DOI: 10.1093/mnras/stv2899. arXiv: 1509.01322 [astro-ph.CO].

[372] J. T. W. et al. In: *arXiv e-prints* (Jan. 2021). arXiv: 2101.09389 [astro-ph.CO].

[373] M. P. et al. In: *ApJS* 199.2, 25 (Apr. 2012), p. 25. DOI: 10.1088/0067-0049/199/2/25. arXiv: 1106.3328 [astro-ph.CO].

[374] Anton M. Koekemoer et al. In: *The Astrophysical Journal Supplement Series* 197.2 (Dec. 2011), p. 36. ISSN: 1538-4365. DOI: 10.1088/0067-0049/197/2/36. URL: http://dx.doi.org/10.1088/0067-0049/197/2/36.

[375] A. Z. et al. In: *The Astrophysical Journal* 801.1 (Mar. 2015), p. 44. ISSN: 1538-4357. DOI: 10.1088/0004-637x/801/1/44. URL: http://dx.doi.org/10.1088/0004-637X/801/1/44.

[376] M. V. et al. In: *MNRAS* 487.4 (Aug. 2019), pp. 4870–4883. DOI: 10.1093/mnras/stz1644. arXiv: 1811.05477 [astro-ph.GA].

[377] Planck Collaboration et al. In: *A&A* 596, A104 (Dec. 2016), A104. DOI: 10.1051/0004-6361/201628522. arXiv: 1603.04919 [astro-ph.CO].

[378] Planck Collaboration et al. In: *A&A* 594, A23 (Sept. 2016), A23. DOI: 10.1051/0004-6361/201527418. arXiv: 1509.06555 [astro-ph.CO].

[379] B. J. B., D. F. M., and D. W. H. In: *AJ* 146.1, 7 (July 2013), p. 7. DOI: 10.1088/0004-6256/146/1/7. arXiv: 1211.5805 [astro-ph.IM].

[381] S. K. N. P. et al. In: *AJ* 154.4, 132 (Oct. 2017), p. 132. DOI: 10.3847/1538-3881/aa8565. arXiv: 1703.01303 [astro-ph.IM].

[382] C. M. C., D. N., and A. C. In: *PhysRep* 541.2 (Aug. 2014), pp. 45–161. DOI: 10.1016/j.physrep.2014.02.009. arXiv: 1402.1456 [astro-ph.CO].

[383] M. A. M. and G. L. In: *ApJS* 157.2 (Apr. 2005), pp. 302–323. DOI: 10.1086/427938. arXiv: astro-ph/0412216 [astro-ph].

[386] M. G. H. and E. D. In: *ARA&A* 39 (Jan. 2001), pp. 249–307. DOI: 10.1146/annurev.astro.39.1.249. arXiv: astro-ph/0105539 [astro-ph].

[387] T. D. R. et al. In: *MNRAS* 442.1 (July 2014), pp. 196–206. DOI: 10.1093/mnras/stu868. arXiv: 1405.1046 [astro-ph.GA].

[388] T. D. R. et al. In: *MNRAS* 459.2 (June 2016), pp. 1626–1645. DOI: 10.1093/mnras/stw712. arXiv: 1508.00586 [astro-ph.GA].

[389] M. B et al. In: *A&A* 437.1 (July 2005), pp. 49–60. DOI: 10.1051/0004-6361:20042234.

[390] G. E. M. et al. In: *The Astrophysical Journal* 760.1 (Oct. 2012), p. 6. ISSN: 1538-4357. DOI: 10.1088/0004-637x/760/1/6. URL: http://dx.doi.org/10.1088/0004-637X/760/1/6.

[391] M. T. S. et al. In: *ApJL* 747.2, L31 (Mar. 2012), p. L31. DOI: 10.1088/2041-8205/747/2/L31. arXiv: 1202.0290 [astro-ph.CO].

[393] J. S. et al. In: *ApJ* 764.2, 152 (Feb. 2013), p. 152. DOI: 10.1088/0004-637X/764/2/152. arXiv: 1209.5129 [astro-ph.CO].

[394] E. M. M. et al. In: *ApJ* 808.1, 47 (July 2015), p. 47. DOI: 10.1088/0004-637X/808/1/47. arXiv: 1408.6248 [astro-ph.CO].

[395] E. K. et al. In: *ApJL* 516.1 (May 1999), pp. L1–L4. DOI: 10.1086/311983. arXiv: astro-ph/9902351 [astro-ph].

[396] A. W. B. In: *MNRAS* 330.1 (Feb. 2002), pp. 219–224. DOI: 10.1046/j.1365-8711.2002.05058.x. arXiv: astro-ph/0110403 [astro-ph].

[397] M. Z. et al. In: *ApJL* 769.2, L31 (June 2013), p. L31. DOI: 10.1088/2041-8205/769/2/L31. arXiv: 1303.6725 [astro-ph.CO].

[398] G. J. B. et al. In: *MNRAS* 433.4 (Aug. 2013), pp. 3062–3078. DOI: 10.1093/mnras/stt948. arXiv: 1306.1217 [astro-ph.IM].

[399] B. B. et al. In: *A&A* 588, A107 (Apr. 2016), A107. DOI: 10.1051/0004-6361/201527313. arXiv: 1509.01784 [astro-ph.IM].

[400] J. S., N. G. C., and S. R. G. In: *ApJ* 744.2, 169 (Jan. 2012), p. 169. DOI: 10.1088/0004-637X/744/2/169. arXiv: 1110.3473 [astro-ph.IM].

[401] M. J. G. and G. S. O. In: 105.2 (Oct. 1993), pp. 537–547. DOI: 10.1006/icar.1993.1147.

[405] S. A. et al. In: *JCAP* 2020.12, 047 (Dec. 2020), p. 047. DOI: 10.1088/1475-7516/2020/12/047. arXiv: 2007.07288 [astro-ph.CO].

Appendix A

CONFIGURATION SPACE ESTIMATORS FOR INTENSITY MAPPING

While the pseudo-C_ℓ approach to fluctuation-based measurements has been successfully used in a variety of contexts, there are challenges related to its application. One example is masking; in the limit where a large fraction of pixels are masked, the standard mode de-convolution through mixing matrices can become more unstable. One needs to account for additional mode coupling from any filtering (as done in Chapter 2) as well as any systematics that couple through the mask. A benefit of pseudo-C_ℓ methods is that they are relatively fast to compute and can accommodate a variety of effects through the MASTER formalism.

Configuration space estimators, such as the two-point correlation function, bypass the standard challenges of masking by down-weighting pairs (or higher order multiplets) which include masked pixels, avoiding biases and mode coupling induced by the mask. One trade-off is that configuration space estimators can be more computationally demanding, however progress has been made to accelerate the computation, which naively scales as N_{pix}^n for a n-point correlation function.

The angular correlation function $w(\theta)$ is defined as the Fourier transform of the angular power spectrum C_ℓ:

$$w(\theta) = \int_0^\infty \frac{d\ell}{2\pi} \ell C_\ell J_0(\ell\theta). \tag{A.1}$$

Conversely, the angular power spectrum can be expressed as the Hankel transform of the two point correlation function:

$$C_\ell = 2\pi \int_0^\infty \theta d\theta w(\theta) J_0(\ell\theta). \tag{A.2}$$

Through these transformations it is possible to derive the angular power spectrum while bypassing the Discrete Fourier Transform (DFT).

We investigate a configuration space estimator based off the code `TreeCorr` [406]. Initial tests on 1024×1024 pixel maps with nineteen log-spaced bins suggest `TreeCorr` is fast, taking $O(10)$ seconds. In Figure A.1 we show the recovered angular two point correlation function for a single mock CIBER point source map,

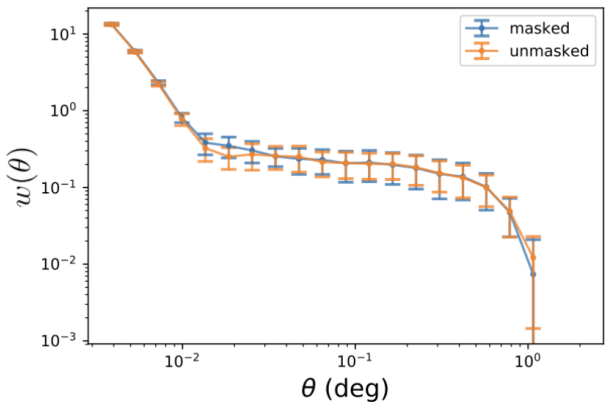

Figure A.1: Masked (blue) and unmasked (orange) angular two-point correlation function estimates $w(\theta)$ for a set of fifty IGL realizations using `TreeCorr`. Errorbars indicate the dispersion across realizations.

with and without masking. The close agreement between cases (with the exception of one angular bin at $\theta = 2 \times 10^{-2}$ deg) highlights the benefit of this approach. The slightly larger uncertainties in the masked case capture the loss of information due to the mask.

However, it may be desirable to down-weight noisy pairs in bins of (θ_x, θ_y) before computing 1D correlation functions, as done within the CIBER pipeline in Fourier space (see Chapters 2 and 3). This becomes more memory- and time-intensive. The finest 2D resolution available with `TreeCorr` is 256×256 bins, which places limitations on the effectiveness of pair weighting. In Figure A.2 we show an example of the estimated $w(\theta_x, \theta_y)$ from an ensemble of 256×256 pixel read noise realizations. Ultimately the computational performance of such an estimator should be compared against what is required to estimate mode mixing matrices, which dominates the CIBER computational budget.

Window function

In the case where the angular correlation function is known at all scales θ, C_ℓ and $w(\theta)$ can be computed faithfully. However, in practice we use a finite range of scales bounded by the pixel resolution and field of view. To handle this, we write the Hankel transform with filter function $T(\theta)$:

$$\hat{C}_\ell = \int_0^\infty d\theta J_0(\ell\theta) w(\theta) T(\theta) \tag{A.3}$$

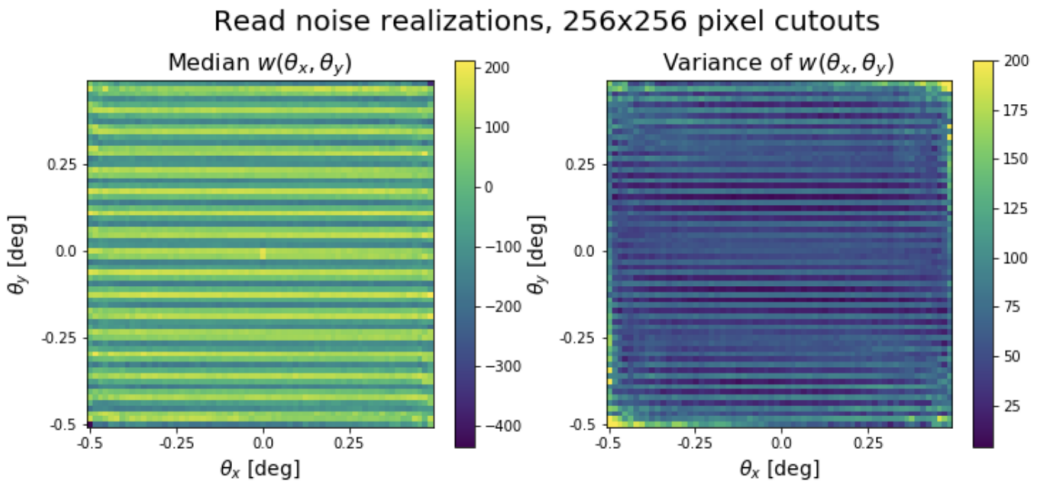

Figure A.2: Median $w(\theta_x, \theta_y)$ estimates for fifty read noise realizations (left), along with the associated per-bin variance (right).

$T(\theta)$ is equal to 1 within θ_{min} and θ_{max} and vanishes otherwise. With infinite coverage, $T(\theta)$ is simply unity for all θ. Because of finite angular covarage, the computed bandpowers will have a different response than if the full integration were performed. With knowledge of the bandpowers and the integration range of the correlation function, we can compute a window matrix that can be used to relate theory curves to measured bandpowers. Specifically, let $W_{b\ell}$ be a $N_{bandpower} \times \ell_{max}$ matrix that projects a high dimensional theory curve to a smaller number of bandpowers. This is written as

$$\hat{C}_b = W_{b\ell} C_\ell^{th} \tag{A.4}$$

where $\hat{C}_b$ is the average of power spectral modes within bandpower b and C_ℓ^{th} is the theory spectrum. If we had full sampling of the two point correlation function at all lags, then the filter function $T(\theta) = 1$ for all θ.

Bandpass apodization

The sharp integration cutoff of the bandpasses may produce ringing after the Hankel transform to C_ℓ is calculated. Integrating against the zeroth order Bessel function, which exhibits oscillatory decay, can produce ringing artifacts. [407] suggests the use of a Hann function to apodize the correlation function such that out of band sensitivity is reduced. We use an upper Hann window to apodize on large scales, which can be defined as a piecewise function of $x = \log(\theta)$:

Apodizing on large scales reduces ringing and reduces sensitivity of the window functions to scales larger than the FOV. This looks like

$$T_{upper}(x) = \begin{cases} 0; & x < x_{lo} \\ 1; & x_{lo} \leq x < x_{up} - \frac{\Delta_x}{2} \\ \cos^2\left[\frac{\pi}{2}\frac{x-(x_{up}-\Delta_x/2)}{\Delta_x}\right]; & x_{up} - \frac{\Delta_x}{2} \leq x < x_{up} + \frac{\Delta_x}{2} \\ 0; & x \geq x_{up} + \frac{\Delta_x}{2} \end{cases} . \tag{A.5}$$

Δ_x is a tunable parameter for the log of the taper width. The transformation from configuration to Fourier space changes slightly in the case where we want to compute bandpowers instead of the angular power spectrum at a single multipole ℓ. By computing a given bandpower, we are integrating the underlying power spectrum against the band power response function. A simple bandpower response function is a top hat function bounded by the bin multipoles:

$$S_l(\ell) = \begin{cases} 1; & \ell_{lo,l} \leq \ell \leq \ell_{hi,l} \\ 0; & \text{otherwise} \end{cases} . \tag{A.6}$$

To find the corresponding filters in configuration space, we convolve $S_l(\ell)$ with the zeroth order Bessel function, which has the following closed form:

$$g_b(\theta) = \int_0^\infty d\ell\ell S_l(\ell) J_0(\ell\theta) \tag{A.7}$$

$$= \frac{1}{\theta^2}\left[\theta\ell_{up,l}J_1(\theta\ell_{up,l}) - \theta\ell_{lo,l}J_1(\theta\ell_{lo,l})\right]. \tag{A.8}$$

We then express the bandpower $\hat{C}_b$ as

$$\hat{C}_b = \frac{\pi}{\mathcal{N}_l}\int_0^\infty d\theta\theta T(\theta) w(\theta) g_b(\theta), \tag{A.9}$$

where the normalization can be written as

$$\mathcal{N}_l = \int_0^\infty d\ell\ell^{-1} S_l(\ell) = \ln(\ell_{up,l}/\ell_{lo,l}), \tag{A.10}$$

such that the bandpowers trace $\ell^2 C_\ell$ at the center of log-spaced bin l. We can express the band power angular power spectrum as

$$C_b = \frac{1}{\mathcal{N}_b}\int_0^\infty d\ell\ell W_b(\ell) C_\ell. \tag{A.11}$$

In practice, we will represent $W_b(\ell)$ as a matrix with some multipole resolution, such that the integral above is really just a fancy way of writing (A.4). In any case, we can compute the window function for a given band power as

$$W_b(\ell) = \int_0^\infty d\theta\theta T(\theta) J_0(\ell\theta) g_b(\theta). \tag{A.12}$$

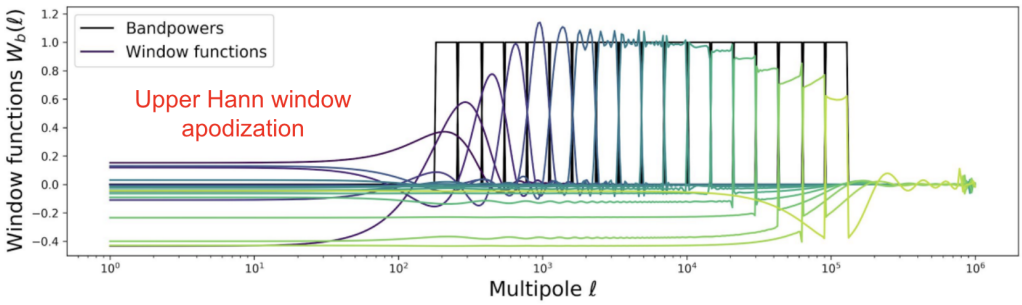

Figure A.3: Window functions for eighteen log-spaced bandpowers using upper Hann apodization.

In Figure A.3 we show the derived $\{W_b(\ell)\}$ for a set of bandpowers using apodization. There are many features to note. While the window functions are roughly centered at the centers of each bandpower, we see considerable out of band response which is more pronounced in the three lowest bandpowers. These window functions do not fold in image filtering, which typically has the effect of suppressing response from scales larger than the FOV. The higher frequency oscillatory behavior arises due to the set of Bessel functions used to compute $W_b(\ell)$.

While configuration space estimators have benefits, namely robustness to masking, there is more work to be done to realize their potential in analyzing existing and near-future datasets. The computational requirements for these estimators at full CIBER resolution are considerable using existing tools such as `TreeCorr`, though this may change with time. It should be possible to incorporate the effects used in pseudo-C_ℓ analyses, e.g., beam corrections, mode weighting, corrections for filtering. Certain effects, such as mask-signal correlations, should be kept in mind despite the lack of an explicit mode coupling correction in this setup. Numerical instabilities from translating correlation functions to power spectra should also be considered carefully, and may motivate performing certain corrections or even a full analysis in configuration space.

BIBLIOGRAPHY

[406] M. J. Astrophysics Source Code Library, record ascl:1508.007. Aug. 2015.

[407] B. J., A. N. T., and A. K. In: *MNRAS* 418.1 (Nov. 2011), pp. 145–169. DOI: 10.1111/j.1365-2966.2011.19472.x. arXiv: 1104.1399 [astro-ph.CO].