Draft version May 24, 2016
Preprint typeset using LaTeX style AASTeX6 v. 1.0

SURVEYSIM: A MARKOV CHAIN MONTE CARLO CODE FOR CONSTRAINING THE EVOLUTION OF GALAXY LUMINOSITY FUNCTIONS

Noah Kurinsky[1,5], Anna Sajina[2], Matteo Bonato[2], Allison Kirkpatrick[3], Alexandra Pope[3], Jed McKinney[2], Andrea Silva[2,6], & Lin Yan[4]

[1]Department of Physics, Stanford University, Stanford, CA

[2]Department of Physics and Astronomy, Tufts University, Medford, MA

[3]Department of Astronomy, University of Massachusetts Amherst, Amherst, MA

[4]California Institute of Technology, Pasadena, CA

[5]Kavli Institute for Particle Astrophysics and Cosmology, SLAC National Accelerator Laboratory, Menlo Park, CA

and

[6]National Astronomical Observatory of Japan 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

(Received; Revised May 24, 2016; Accepted)

ABSTRACT

We present SurveySim: a new public Markov Chain Monte Carlo (MCMC) code which uses only color and flux information to constrain the evolution of galaxy luminosity functions. This code is designed to be easily adaptable, with luminosity function and SED model prescriptions fed into it externally. Here we adopt our new IR SED template library which includes SED models for star-forming galaxies, composites, and AGN, as well as their redshift and luminosity evolution. The code converges on a best model by simulating the density of galaxies in diagnostic color-magnitude plots. We demonstrate SurveySim's use by using the HerMES COSMOS survey data and two diagnostics: $\log(S_{350}/S_{250})$ vs. $\log(S_{250})$ and $\log(S_{250}/S_{24})$ vs. $\log(S_{24})$. We find that a luminosity function consisting of a double power-law with a relatively steep faint-end slope ($\beta = 0.6$) gives the best results. Consistent with previous studies, we find strong luminosity evolution ($\propto (1+z)^{3.8}$) up to $z \sim 2$ but at $z > 2$ both the luminosity and density evolution are negative. Our models are consistent with a range of observational data including luminosity functions, number counts, redshift distributions, CIB intensity and redshift breakdown. This agreement is achieved without incorporating any knowledge of the redshifts of our sources, or prior knowledge of how the IR luminosity function should evolve with redshift. Therefore, SurveySim has the potential to evolve into a useful tool to analyze data from current and upcoming large extragalactic surveys where only limited photometric coverage and even more limited redshift coverage exists.

Keywords: galaxies: evolution, luminosity function, statistics – infrared: galaxies – methods: statistical

1. INTRODUCTION

The recent generation of sensitive infrared space and ground-based survey facilities (e.g. *Spitzer*, *Herschel*, *WISE*[1], ALMA) has allowed us to move from a few tens of dusty high-redshift galaxies just a couple of decades ago, to tens of thousands today. This has truly opened up the study of dust obscured star-formation throughout the history of the

[1] For details on each see Werner et al. (2004); Pilbratt et al. (2010); Wright et al. (2010). Also see Casey et al. (2014) for a more extensive list.

Universe (see Casey et al. 2014, for a review). However, we are not yet taking full advantage of this wealth of data, since the vast majority of these sources likely never will have individual spectroscopic redshift measurements. Moreover, we do not yet have a clear understanding of the full selection biases associated with the different IR-bright populations studied in the literature - a limitation largely due to our lack of consensus of the range of SEDs of dusty, high-z galaxies. The latter includes an incomplete accounting of the role of AGN in the IR emission of galaxies (Kirkpatrick et al. 2015, Roebuck et. al. 2016, subm.). These factors lead to significant uncertainties in the IR luminosity function estimates even locally (see Figure 1).

This paper presents a new MCMC-based code, SurveySim, which is designed to specifically address the above limitations. It is inspired by earlier efforts to use MCMC to model the BLAST number counts (Marsden et al. 2011) which recover the luminosity function evolution parameters as well as SED template color evolution. SurveySim is also inspired by Milky Way structure studies which do not model individual stars, but rather stellar populations using Hess diagrams, where optimizing between the model and observed Hess diagram recovers the 3D stellar distribution as well as the star-formation history (e.g. de Jong et al. 2010). A Hess diagram is the binned density of stars in a color-magnitude diagram - ideal when modelling large populations, such as Galactic stars detected by the SDSS. Similarly, SurveySim models the density of galaxies in color-magnitude or color-color plots, recovering the evolution of the luminosity function. Useful by-products are the number counts, redshift distribution, and contribution to the Cosmic Infrared Background (CIB; for a review see Hauser & Dwek 2001). We stress that this is a statistical approach, not designed to recover parameters for individual galaxies, only for whole populations. With sufficient statistics, it can be complementary to the traditional, observationally-expensive spectroscopic follow-up of relatively small sub-samples (e.g. Casey et al. 2012). Indeed, Christlein et al. (2009) already demonstrate the ability of a maximum likelihood fit to a color-magnitude diagram to recover the evolution of the optical luminosity function without galaxy redshifts. SurveySim does this, currently, for the infrared luminosity function, but importantly is designed to be adaptable and scalable.

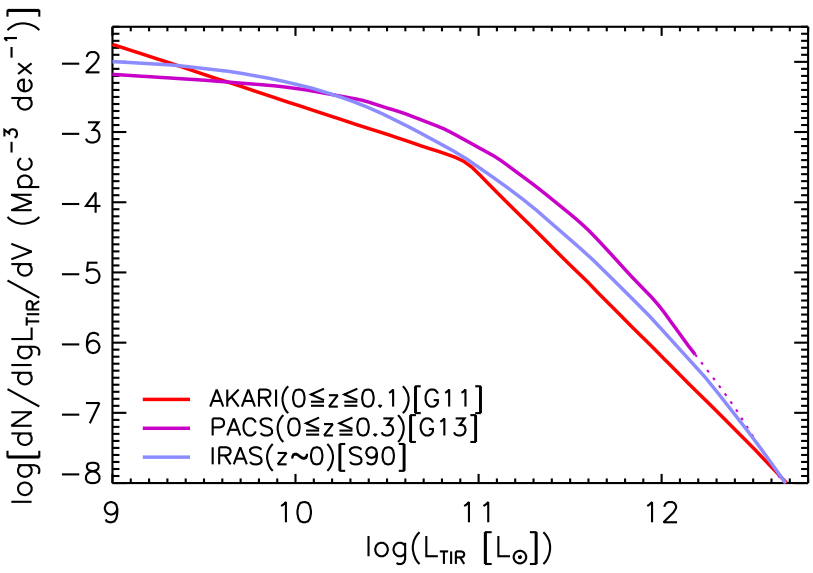

Figure 1. Local total IR luminosity function measurements restricted to those that include both star-forming galaxies and AGN (Saunders et al. 1990; Goto et al. 2011; Gruppioni et al. 2013, here S90, G11, G13). Solid lines correspond to L_{IR} ranges where the LF is directly constrained by data, dotted lines represent extrapolations. We have rescaled all to match our adopted cosmology and $L_{\rm TIR}$ definition. While there is broad agreement, the spread is a factor of $\sim$3 for most luminosities, due to selection biases.

As in all efforts to model galaxy evolution, our results are dependent on the choice of SED templates. To address that, our group has been engaged in an effort to define a new SED template library[2] which, crucially, does a careful

[2] This library is now publicly available through: http://www.astro.umass.edu/~pope/Kirkpatrick2015/

decomposition of the effects of AGN on the IR SED (Kirkpatrick et al. 2015). However, to ensure that any future improvements of our knowledge of the SEDs of dusty galaxies can be easily incorporated, the SED library is not hardwired into the code, but rather fed into it externally. A key and unique feature of SurveySim is its very general design, allowing for a wide range of different galaxy populations, luminosity function forms, SED templates, and photometric filters to be used. The current implementation is limited to the infrared (3–1000μm), although we are planning on extending this range (e.g. Silva et al. 2016 in prep.).

The MCMC approach allows us to examine exactly the constraining power of the particular galaxy population being modeled. It also allows for detailed examination of the degeneracies between the fitted parameters. In upcoming papers, we consider dusty high-z galaxies selected at different parts of the IR SED giving us insight into the relative selection biases (Bonato et al. 2016 in prep; Silva et al. 2016 in prep.).

The structure of this paper is as follows. In Section 2, we discuss the two *Herschel* SPIRE datasets being modeled. In Section 3, we present details of the galaxy evolution modeling, including luminosity function parametrization and SED template library. In Section 4 we describe our MCMC implementation. In Section 5, we present the results of simulating our samples. The best-fit luminosity functions, number counts, redshift distributions, SFR density and contribution to the CIB are compared with a range of published results. In Section 6, we discuss the broader implications of the current results as well as future improvements of SurveySim. Section 7 presents the summary and conclusions. We include two appendices for the more technical aspects of SurveySim. Throughout this paper we adopt a flat Universe with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$, and $H_0 = 73.8$km/s/Mpc (based on the WMAP results Komatsu et al. 2011). We chose the WMAP cosmological model here for ease of comparison with earlier studies.

2. DATA

We model two *Herschel* SPIRE samples, one 24 μm-selected and one 250μm-selected. Both are drawn from the *Herschel* Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2012) COSMOS field and the catalogs are obtained from the *Herschel* Database in Marseille[3]. The 24μm-selected sample is the published "*Spitzer*-prior" catalog which is based on the 2 deg^2 MIPS 24μm-coverage of the field. For the 24μm sample, the 250μm flux densities were estimated at the positions of the known 24μm sources, including deblending of the SPIRE signal as needed. The 250μm-selected sample is the "blank" catalog based on the total 4.78 sq.deg. SPIRE coverage of the field (which reaches a nominal 5 σ depth of 8 mJy). We use the band-merged point-source catalog from the second data release (Oliver et al. 2012) including photometry for the three SPIRE wavelengths. This catalog is constructed by convolving the 250, 350 and 500μm filters at locations corresponding to the single-band 250μm sources detected by HerMES, as described in Oliver et al. (2012). These two catalogs are affected by different potential biases; their joint fits should thus be more robust than a fit to either alone.

The confusion noise, which is significant for SPIRE (especially the blank-sky photometry), is described in Wang et al. (2014), along with the methods to account for and partially mitigate its effect in the published catalogs. This confusion leads to significantly higher 1 σ uncertainties on the blank-sky SPIRE photometry as opposed to the *Spitzer*-prior photometry. For the 24μm-selected sample, we adopt a 24μm flux density limit of 60 μJy and a 1 σ uncertainty of 0.16 μJy – this is coupled with a 250μm flux density limit of 8 mJy (to match the 250μm-selected survey) with a 1 σ uncertainty of 2 mJy. For the 250μm sample the 1 σ uncertainty is 6.95 and 6.63 mJy at 250μm, and 350μm respectively (the uncertainties are the median values in the respective catalogues). Note that obviously an 8 mJy source in the blank-sky 250μm-selected catalog does not constitute a formal detection. However, we also incorporate the survey incompleteness in our modelling, which translates to a negligible fraction of modelled sources with such flux densities being included when we simulate the 250μm-selected sample. Therefore a flux density limit of 8 mJy here is largely a formality.

3. SIMULATING A GALAXY SURVEY

We simulate a survey by stepping through the binned redshift-luminosity space, generating N galaxies according to the luminosity function, and assigning each galaxy a random luminosity and redshift within the $L - z$ bin. Here "L" refers to the total IR luminosity, L_{TIR}, which is the integrated 5-1000μm luminosity. The number of galaxies expected in a given bin is prescribed by the luminosity function, specifically:

$$N(L, z, \Omega) \approx \Phi(L, z) \frac{dV_c}{dz d\Omega}(z) \Omega \Delta z \Delta \log L \quad (1)$$

[3] http://hedam.lam.fr/HerMES/index/download

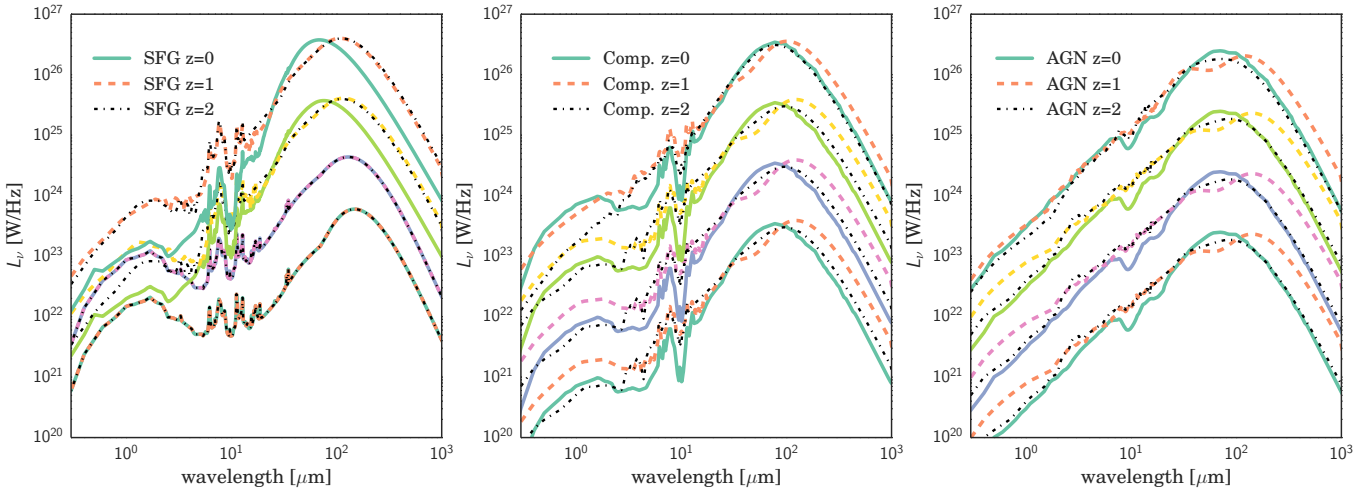

Figure 2. Our default library is binned by $L_{\rm TIR}$ in 17 bins (see section 3.3 for details). For clarity we show only 4 bins here: $\log(L_{\rm TIR}/L_\odot) = 9.75$, 10.75, 11.75 and 12.75. The templates in the higher-z bins are based on Kirkpatrick et al. (2015) as described in the text. *Left:* The SFG templates, where for the $z = 0$ and the higher-z $\log(L_{\rm TIR}) < 11.5$ ones we adopt the Rieke et al. (2009) templates. *Middle:* The Composite templates, where the $z = 0$ ones are all rescaled from the NGC6240 SWIRE library template. *Right:* The AGN templates, where the $z = 0$ templates are all rescaled from the Mrk231 SWIRE library template. The differences between the solid, dashed, and dot-dashed curves, especially among the SFGs highlight the SED evolution build into our SED library. However, note that for the $z > 0$ Composite and AGN templates there is limited data in the far-IR/sub-mm translating into substantial uncertainties in the templates in this regime.

For this calculation, we use the $\log(L)$ and z values corresponding to the center of the L-z bin, such that a given redshift bin i is defined as $z \in (z_i - \Delta z/2, z_i + \Delta z/2]$ and a given luminosity bin j is defined as $\log(L) \in (\log(L_j) - \Delta\log(L)/2, \log(L_j) + \Delta\log(L)/2]$. [4] For each redshift-luminosity bin, we determine N, which is inherently a non-integer value. This is rounded-off to an integer by using a uniform random variate between 0 and 1 to determine whether to round downward or upward.

In addition, each of the N sources is ascribed a luminosity and redshift according to a triangular distribution within the given luminosity redshift bin, which approximates the change in density as linear with L and z within each bin. This implies that the luminosity and redshift bins should not be too coarse or else the triangular approximation fails, nor too fine or else the probability of finding at least 1 galaxy per bin becomes too small.

The SED template library gives us the observed flux density (see e.g. Hogg et al. 2002) for a given luminosity, redshift, and spectral type (e.g. starburst or AGN). This is done via:

$$S_\nu = \frac{(1+z)}{4\pi d_L^2} \frac{\int_0^\infty L_{\nu(1+z)} T_\nu d\nu}{\int_0^\infty T_\nu d\nu} \tag{2}$$

where $\nu \equiv \nu_{obs}$, d_L is the luminosity distance[5], $L_{\nu(1+z)}$ is the specific luminosity for a given SED template in the emitted frame, T_ν is the filter transmission curve, and the $(1+z)$ factor arises from $d\nu_{em}/d\nu_{obs} = (1+z)$. We apply observational error to these flux densities in the form of a random Gaussian variate with a σ corresponding to the typical uncertainty in the survey being simulated. Lastly, we apply the same selection (e.g. limiting flux density in a particular band) to these simulated galaxies as the observed survey we are fitting.

3.1. *Completeness Corrections*

SurveySim includes completeness correction curves to account for the different source detection efficiency as a function of flux density. For example, for the *Herschel* SPIRE 250μm-selected sample, we adopt the completeness curves derived in Wang et al. (2014) where $S_{250} \sim 12$ mJy corresponds to 50% completeness and 100% completeness is reached by $\sim$40 mJy. We tested for incompleteness among the 24μm-selected sample by comparing our measured differential 24μm number counts with published values, and found this sample to be complete above the adopted 60 μJy flux density limit, therefore no completeness correction was applied to it. In SurveySim, we accept a simulated

[4] In this paper, we adopt a luminosity range of $logL_{\rm TIR} = 9.00$ - 13.00 with $\Delta logL_{\rm TIR} = 0.25$, corresponding to the sampling of our SED library. Our default redshift range is 0.1-5.0 with $\Delta z = 0.1$.

[5] We compute the luminosity distance and co-moving volume using numerical integration in redshift steps of $z = 0.001$.

source with a given flux density with probability given by the completeness function.

3.2. *The Luminosity Function*

SurveySim allows for three potential functional forms for the luminosity function: a Schechter function ("S"; Eq 3; e.g. Blanton et al. 2003), a modified Schechter function ("MS"; Eq 4; e.g. Gruppioni et al. 2013), and a double power-law ("PL"; Eq 5; e.g. Negrello et al. 2013).

$$\Phi_S = \Phi^* \Big(\frac{L}{L^*}\Big)^{\alpha} \exp\Big[-\frac{L}{L^*}\Big] \tag{3}$$

$$\Phi_{MS} = \Phi^* \Big(\frac{L}{L^*}\Big)^{1-\alpha} \exp\Big[-\frac{1}{2\sigma^2} log_{10}^2\Big(1+\frac{L}{L^*}\Big)\Big] \tag{4}$$

$$\Phi_{PL} = \Phi^* \left[\left(\frac{L}{L^*}\right)^{\alpha} + \left(\frac{L}{L^*}\right)^{\beta}\right]^{-1} \tag{5}$$

In all cases, Φ^* and L^* are functions of redshift and we adopt the parameterizations

$$\Phi^*(z) = \Phi_0^*(1+z)^p \tag{6}$$

$$L^*(z) = L_0^*(1+z)^q \tag{7}$$

where p denotes density evolution, and q denotes luminosity evolution. We allow for two break redshifts, where the power of the density (z_{bp}) and luminosity (z_{bq}) evolution changes[6]. In subsequent sections, p_1 and q_1 denote the evolution from these break redshifts until the present day, whereas p_2 and q_2 denote the evolution at redshifts higher than the respective break redshifts.

3.3. *SED Templates*

Since our understanding of the range of high-z galaxy SEDs is still evolving, it is important to note that our SED Template library is not hardwired into SurveySim. Instead, it is read into it as a stand-alone FITS file with different FITS extensions corresponding to different SED types, and/or redshift bins. We chose this structure in order to allow us to easily update the SED library without a need to modify the code itself. SurveySim adopts its wavelength and luminosity arrays as well as the SED type mix from this external FITS file, which will facilitate future extension into other wavelength regimes. Our default SED library includes 3 SED types: star-forming galaxies (SFG), Composites and AGN. It also incorporates 3 redshift bins: "$z \sim 0$" or $z = 0 - 0.499$, "$z \sim 1$" or $z = 0.5 - 1.2$, and "$z \sim 2$ or $z > 1.2$. These SED types and redshift bins are determined based on the redshift bins in the Kirkpatrick et al. 2015 comprehensive SED templates which form the core of our library. This library is based on an exceptional sample of 343 dusty galaxies with mid-IR IRS spectra as well as a wealth of ancillary data covering the near- to far-IR regime. The IRS spectra allow for accurate spectral classification, separating out star-forming galaxies (SFG), AGN and composites. This library however does not include $z < 0.5$ nor $L_{\rm IR} \lesssim 10^{11} L_{\odot}$ sources.

For the $z \sim 0$ SFG templates, we adopt the SED library of Rieke et al. (2009), who construct a set of templates from averaged fits of theoretically based models to local IRAS 60 μm-selected galaxies. Their templates are binned in terms of total-IR luminosity, $L_{\rm TIR}$, from $10^{9.75} - 10^{13} L_{\odot}$ in bins of 0.25 dex. We extend this down to $10^{9.00}$, by assuming constant SED shape below $10^{9.75}$. With this modification, we adopt this luminosity binning for our default SED library, re-scaling the Kirkpatrick et al. (2015) templates as needed. For the $z \sim 0$ Composite templates, we adopt the SWIRE library NGC6240 template (Polletta et al. 2007), re-scaled to each of our luminosity bins. Similarly, for the $z \sim 0$ AGN templates, we adopt the SWIRE library Mrk231 template, re-scaled to each of our luminosity bins. We analyzed the IRS spectra of NGC6240 and Mrk231 in the same manner as our supersample above and found them to be consistent with our definition of a composite and AGN sources respectively. Figure 2 illustrates our template library highlighting the apparent SED evolution within the SFG, Composite and AGN sub-classes. Note that for Composites and AGN at $z > 0$, there are typically few individual detections in the far-IR therefore effects in this regime such as seen among the $z \sim 1$ sources should be viewed with caution. The limitations of the current library are discussed in Section 6.1, along with planned future improvements.

[6] To allow for smooth transition, after the break, we use the parametrization $\Phi^*(z) = \Phi^*(z_{bp})(1 + z - z_{bp})^{p_2}$.

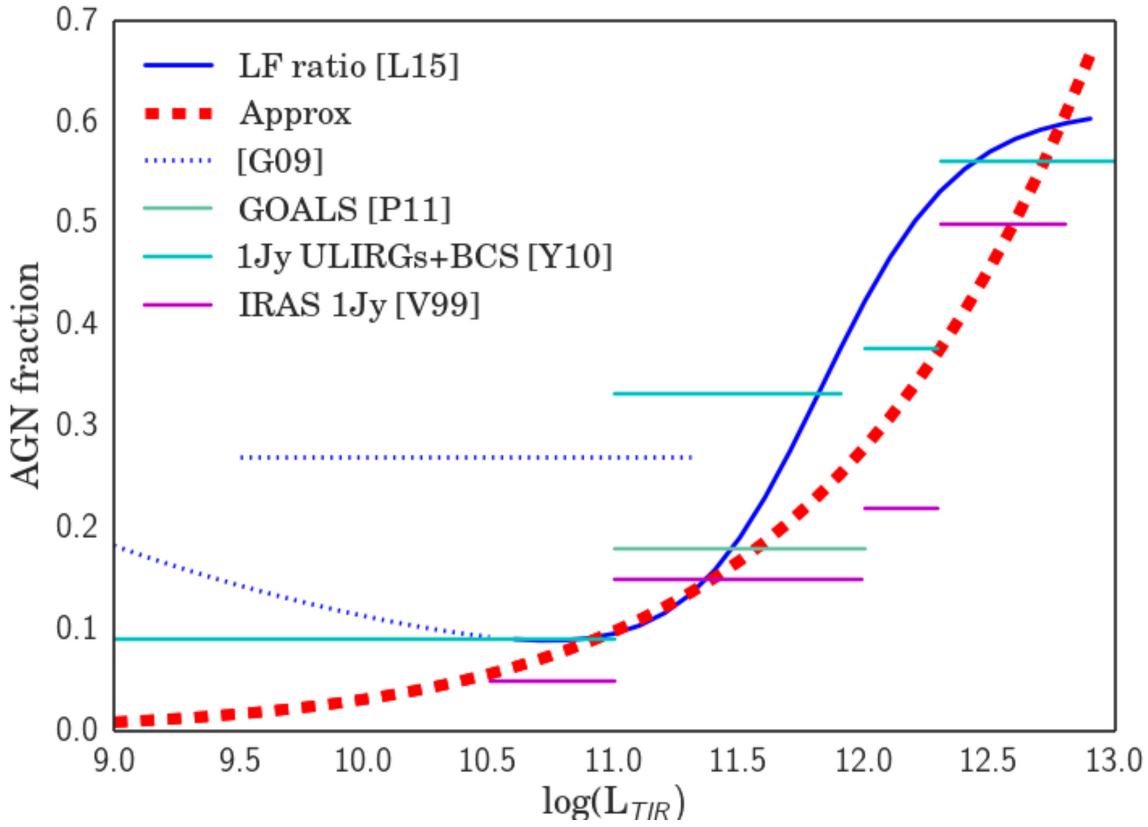

Figure 3. The ratio of the AGN (Lacy et al. 2015) to total LF (Goto et al. 2011). To correct for the difference in redshift bins between the two, we evolved the Φ_* and L_* values in the Goto et al. (2011) LF to the mean redshift of the bin in Lacy et al. (2015), using $q = 4.9$ and $p = -0.98$. We also show the local AGN fractions in different luminosity ranges estimated directly from the presence of AGN-excited optical and/or mid-IR lines (Veilleux et al. 1999; Goulding & Alexander 2009; Yuan et al. 2010; Petric et al. 2011, ; V99, G09, Y10, P11). The thick dashed red curve represents the power-law approximation we adopt in SurveySim (see Equation 8).

3.3.1. *Including different SED types*

To incorporate different galaxy populations, we choose to model the total LF and parametrize the fraction of it due to the different populations (as opposed to modelling several different LFs for example). However, in either approach, simplifying approximations need to be made both to avoid an excessive number of free parameters (and associated degeneracies) and to bolster our incomplete information.

We have no knowledge of the LF of AGN-SFG composites as opposed to AGN-dominated sources, so we will aim to parametrize the fraction of sources, as a function of $L_{\rm TIR}$ that contain AGN regardless of their energetic dominance. The most recent and complete estimate of the local AGN LF is that based on mid-IR selected, spectroscopically-confirmed AGN of Lacy et al. (2015)[7]. We confirmed that this selection largely includes both our AGN and AGN+SFG composites since in the Kirkpatrick et al. (2015) sample, upon which our templates are based, 95% of the AGN and 70% of the Composites fall within the Lacy wedge for IRAC color-color selection of AGN.

Figure 3 shows the ratio of the AGN luminosity function to the total LF, where we adopt the Goto et al. (2011) AKARI-based local LF. We chose this one as it has the best high-L coverage, thanks to the large volume sampled. We note however, that the Lacy et al. (2015) LF is computed over the range $z = 0.05$-0.4, whereas Goto et al. (2011) covers $z = 0$-0.1. To allow for direct comparison, we apply luminosity and density evolution to the Goto et al. (2011) consistent with our best-fit model presented in Section 5 to shift it to an effective redshift of 0.23. The ratio between the AGN and total LF is then compared with direct measurements of the AGN fraction based on the detectability of AGN-excited optical and/or mid-IR lines.

Reassuringly, these direct measurements are roughly consistent with the LF ratio. The only outlier is the Goulding

[7] Lacy et al. (2015) present both a total LF, and individual LFs based on AGN type all in terms of L_{5um}. We apply conversion factors to $L_{\rm TIR}$ specific to each type and then add up the resultant LFs to obtain the total AGN LF in terms $L_{\rm TIR}$. We also correct for differences in the adopted cosmology

& Alexander (2009) measurement who used [NeV] 14.3μm detectability to find a large AGN fraction for $L_{IR} < 10^{11} L_{\odot}$ sources. Their sample however is quite small (15 objects). In a sample more than an order of magnitude larger, Petric et al. (2011) find a smaller fraction using the same line as a diagnostic. The difference may also be due to differences in relative sensitivity, so the Goulding & Alexander (2009) result may suggest more low-level deeply obscured AGN activity (see also below). We note the upturn in the AGN fraction at low-L; however, the Lacy et al. (2015) data do not sample this range and the uncertainties in the faint end slope of both the Lacy et al. (2015) and Goto et al. (2011) LFs make this extrapolation very unreliable.

We caution that the optical line diagnostics in the above studies effectively separate out Seyfert nuclei. The recent study of Yuan et al. (2010) accounts also for optical starburst-AGN composites. Including these gives roughly twice the AGN fraction given above, while preserving the trend of increasing AGN fraction with IR luminosity. We did not adopt this for this paper, because it is not entirely clear yet how these optically-defined composites compare with mid-IR defined composites. In addition, the question of the AGN fraction is degenerate with the details of the SED library adopted (see Section 6 for further discussion). To test this issue, we ran the models presented here with a modification allowing for double the $z \sim 0$ AGN fraction than allowed for by default, but found the results to be worse. As stated above, this conclusion is degenerate with the SED library adopted.

With these caveats in mind, Figure 3 suggests that a reasonable approximation to these different estimates of the AGN fraction is a power-law of the form:

$$f_{AGN,z} = f_{AGN,0} * [log(L_{\rm TIR})/12]^{fp} * (1+z)^{t} \quad (8)$$

where $f_{AGN,0}$ represents the $z \sim 0$ AGN fraction at $log(L_{\rm TIR})$=12, fp is the fraction power index, and t parametrizes the evolution of the AGN fraction with redshift. In Figure 3, we adopt $f_{AGN,0}$=0.28, and fp=12 respectively. The functional form is assumed to be fixed, but the fraction of AGN is allowed to evolve as shown in Equation 8. As done previously, we take $t \equiv t_1$ until a break redshift, z_{bt} and $t \equiv t_2$ after the break redshift[8]. Lastly, we assume that the Composite sources represent a fixed fraction of all AGN. This fraction is given by a single parameter f_{Comp}.

It is clear that the constancy of the AGN fraction functional form (i.e. that the total and AGN functional forms do not evolve in their shape, only in their relative amplitudes) maybe incorrect. Lacy et al. (2015) study the evolution of the AGN LF and a model with fixed slopes and only the usual luminosity and density evolution, although not unique, is consistent with the data. An even bigger assumption, is that the composite fraction is constant with redshift, but this represents our lack of knowledge on this issue. The parametrization of the different SED types and the associated redshift evolution represent the largest uncertainties in the current implementation of SurveySim. As our knowledge of the evolution of the AGN LF and the role of composites therein improves in the future, this can be easily incorporated into SurveySim.

Lastly, we note that the above definition of f_{AGN} differs from what we use in Kirkpatrick et al. 2015 or Roebuck et al. (2016, subm). In these papers, we use f_{AGN} to mean the fraction of L_{TIR} for a given galaxy that is AGN-powered. In this paper, it is the fractino of galaxies in a given $L-z$ bin that are classified as either AGN or starburst-AGN composites. However, when we estimate the evolution of the global star-formation rate density, we make use of the median values for the fraction of L_{TIR} due to star-formation vs. AGN as a function of galaxy type derived in Roebuck et al. (2016, sum.).

3.4. *Fitting Metrics*

The fitting metrics allowed in SurveySim are either a color-color or color-magnitude plot. In this paper, we use the latter. Specifically, for both the observations and the simulated data, we produce a two dimensional histogram, with each axis being either a color (defined as $\log(f_x/f_y)$) or "magnitude" (defined as $\log(f_x)$). Examples for the two SPIRE samples are shown in Fig. 4. The best-fit solution is then found by minimizing a χ^2-like statistic between the two histograms. The details of implementing such 2D histogram fitting are non-trivial and are discussed in detail in Appendix A.

If one axis is a magnitude, the 2d histogram fitting is essentially a simultaneous fit to the number counts in that magnitude and the color distribution. We choose this approach as including colors in the fit allows us to partially alleviate the degeneracy between SED templates and the evolution of the LF. Without the right SEDs, the color distribution cannot be reproduced by merely adjusting the level of luminosity or density evolution. A second motivation is that rare populations with unusual SEDs can in principle be detected in the residuals of the observed and simulated

[8] Similar to the density and luminosity evolution, to have a smooth transition, after the break we use $f_{AGN,z} = f_{AGN,z_{bt}}(1+z-z_{bt})^{t_2}$.

2D histograms, which cannot be done fitting number counts alone (see Section 6.1).

4. MARKOV CHAIN MONTE CARLO FITTING

Markov Chain Monte Carlo (MCMC) consists of random deviations from an initial guess (drawn from proposal distributions), and an algorithm to decide whether to accept these deviations (the sampling algorithm). The accepted guesses form a Markov chain as each guess only depends on the previous one and not on the earlier history or initial state. The location of the initial guesses, the width of the proposal distributions, and the leniency of the sampling algorithm in accepting guesses (which is controlled by a temperature-like parameter) all determine the speed with which the code converges on the best solution. To ensure good coverage of the parameter space, and avoid getting stuck in a local minimum, multiple chains are often initiated (see e.g. Dunkley et al. 2005; Ross 2013, for more details). In the following sub-sections, we describe our particular implementation of MCMC.

4.1. *Sampling*

For our proposal distribution, we adopt a multivariate Gaussian distribution, which generates correlated random deviates. This function has the form:

$$P(\mathbf{X_R}; \mathbf{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{\Sigma}|}} \exp\left[\frac{-\mathbf{x^T \Sigma^{-1} x}}{2}\right] \tag{9}$$

where $\mathbf{x}$ is an input vector containing the n sampled parameters, $\mathbf{\Sigma}$ is the $n \times n$ parameter covariance matrix, and $\mathbf{X_R}$ is the randomly generated parameter vector. The variance of each parameter is initially chosen based on the user set limits at run-time, with parameters assumed independent ($\Sigma_{ij} = \sigma_i^2 \delta_{ij}$) but is adjusted during the course of simulation to take into account early priors and make sampling more efficient through regular re-computation of the covariance matrix between all parameters.[9]. The trade-off between small and large variance is completeness of sampling versus resolution of sampling for a given time period.

Next, we adopt the Metropolis-Hastings sampling algorithm (Metropolis et al. 1953; Hastings 1970), which consists of the following steps (see also Johnson et al. 2013):

1. Draw a random deviate from the proposal distribution thus displacing each parameter from its last accepted value, X_i, to a new one, X_{i+1}.

2. Calculate the acceptance likelihood, based on the difference in χ^2 achieved with the last accepted and proposed parameter sets and a simulation temperature which controls how likely are large deviations to be accepted ("hotter" runs mean the code can make bigger jumps, and vice versa):

$$P_{acc} = \exp\left[\frac{-\Delta\chi^2}{T}\right] \tag{10}$$

3. Accept any guess that improves the χ^2 (where $P_{acc} > 1$), but also accept some guesses with worse χ^2 by generating a uniform random variate between 0 and 1, U, and accepting a guess if U is less than P_{acc}:

$$\text{Accept}(X_{i+1}|U, P_{acc}) = \begin{cases} \text{True} & U < min[1, P_{acc}] \\ \text{False} & U \geq min[1, P_{acc}] \end{cases} \tag{11}$$

4. Repeat until convergence criteria are satisfied (see Section B).

To improve the sampling of parameter space we use 5 independent chains started at random points in our sampling space, with each chain run at the same temperature. The multiple chains help us find the global maximum in the likelihood surface as well as test convergence (see Appendix). During initial sampling, both the temperature of the chains and the covariance matrix used for sampling are adjusted at various intervals (see Section 4.2). The covariance matrix is computed considering all samples in the latter half of all chains, and thus the resulting best-fit parameters come from the converged samples of all 5 chains.

[9] It is often stated that assumption of initial parameters and adjustment of covariance during fitting adds bias to the fit, however it can be shown that the assumption of zero covariance between parameters and equal gaussian variance is a bias in itself(Joachimi & Taylor 2014)

4.2. *Acceptance Optimization*

Acceptance optimization involves tuning the acceptance rate such that we maximize our chances of both isolating the true global best-fit parameter vector while efficiently sampling enough of the surrounding likelihood space to allow the parameter uncertainties to be determined. Johnson et al. (2013) and Dunkley et al. (2005) cite an ideal acceptance rate of 25%. Both are achieved during an initial "burn-in" period which is discarded in the parameter uncertainty estimation.

We implement a simple regression algorithm to adjust the temperature of the Metropolis-Hastings sampler to maintain the acceptance rate within the desired range (20-30%). We compute the average acceptance rate at regular intervals (defaulted to every 10 MC steps in the burn-in period), adjusting the temperature according to the steepest-gradient descent step rule

$$T = T(1 - \alpha(A_{calculated} - A_{ideal})) \tag{12}$$

where T is the temperature and A is the acceptance rate. The additional parameter α controls the rate of change in T; it is a user-definable parameter, which controls the efficiency of the process, but does not affect its accuracy. This adjustment scheme will compensate for the changing magnitude of the covariance matrix, which will produce more or less scatter as the MCMC progresses, as well as for variance in RMS chi-square values between different samples, which will affect overall acceptance.

Table 1. Best-fit parameters[a]

Model	χ^2	σ	$\log(L_0^*/L_\odot)$	$\log(\Phi_0^*)$	z_{bp}	z_{bq}	p_1	q_1	p_2	q_2
			$f_{AGN,0}$	t_1	t_2	z_{bt}	f_{comp}[b]			
SPIRE 250μm-selected sample										
PL	0.11	4.7	$10.91^{+0.01}_{-0.09}$	$-3.25^{+0.07}_{-0.04}$	$1.38^{+1.33}_{-0.25}$	$1.85^{+0.87}_{-0.72}$	$-1.48^{+3.48}_{-1.47}$	$3.76^{+1.45}_{-1.06}$	$-1.87^{+1.96}_{-2.82}$	$3.31^{+0.23}_{-6.10}$
			$0.26^{+0.11}_{-0.10}$	$0.84^{+2.58}_{-0.89}$	$1.19^{+2.87}_{-3.12}$	$2.10^{+0.91}_{-0.15}$	$0.18^{+0.51}_{-0.01}$			
MS	0.12	4.2	$10.08^{+0.06}_{-0.05}$	$-2.36^{+0.10}_{-0.01}$	$2.16^{+0.56}_{-1.05}$	$1.20^{+1.31}_{-0.27}$	$-1.42^{+2.06}_{-2.82}$	$3.15^{+1.58}_{-1.53}$	$-2.27^{+3.65}_{-1.72}$	$4.27^{+0.40}_{-5.80}$
			$0.32^{+0.09}_{-0.12}$	$2.78^{+1.74}_{-1.97}$	$-0.12^{+3.22}_{-2.90}$	$2.56^{+0.43}_{-0.60}$	$0.77^{+0.02}_{-0.50}$			
MIPS 24μm-selected sample										
PL	0.11	7.1	$10.89^{+0.02}_{-0.08}$	$-3.23^{+0.04}_{-0.07}$	$1.79^{+0.95}_{-0.59}$	$2.68^{+0.13}_{-1.40}$	$-0.48^{+3.49}_{-2.17}$	$2.14^{+2.63}_{-2.49}$	$-2.62^{+3.78}_{-1.82}$	$-1.54^{+3.52}_{-2.76}$
			$0.28^{+0.12}_{-0.09}$	$0.07^{+2.66}_{-0.65}$	$-2.37^{+5.31}_{-1.06}$	$2.72^{+0.34}_{-0.71}$	$0.24^{+0.48}_{-0.05}$			
MS	0.09	6.8	$10.10^{+0.05}_{-0.05}$	$-2.27^{+0.03}_{-0.08}$	$1.56^{+0.99}_{-0.54}$	$1.42^{+1.21}_{-0.37}$	$-1.91^{+3.41}_{-2.06}$	$3.05^{+1.43}_{-3.73}$	$-0.95^{+2.59}_{-3.14}$	$0.12^{+3.55}_{-2.97}$
			$0.30^{+0.10}_{-0.11}$	$-0.20^{+2.70}_{-0.75}$	$-3.09^{+5.69}_{-0.49}$	$1.98^{+0.78}_{-0.25}$	$0.43^{+0.30}_{-0.21}$			
Joint results										
PL	/	/	$10.89^{+0.04}_{-0.09}$	$-3.24^{+0.07}_{-0.08}$	$2.02^{+0.99}_{-1.19}$	$1.83^{+1.28}_{-0.97}$	$-0.31^{+3.07}_{-2.88}$	$3.78^{+1.49}_{-1.65}$	$-2.16^{+2.41}_{-3.45}$	$-1.40^{+5.60}_{-3.60}$
			$0.26^{+0.13}_{-0.12}$	$0.71^{+2.14}_{-1.42}$	$0.34^{+4.53}_{-3.79}$	$2.25^{+0.97}_{-0.52}$	$0.37^{+0.37}_{-0.29}$			
MS	/	/	$10.08^{+0.07}_{-0.06}$	$-2.33^{+0.10}_{-0.05}$	$1.36^{+1.33}_{-0.60}$	$1.57^{+1.02}_{-0.89}$	$-2.13^{+3.35}_{-2.25}$	$3.30^{+1.55}_{-2.60}$	$-2.16^{+3.58}_{-2.58}$	$2.85^{+2.64}_{-5.38}$
			$0.30^{+0.16}_{-0.15}$	$0.29^{+3.07}_{-1.14}$	$-0.30^{+4.85}_{-3.79}$	$2.11^{+0.82}_{-0.47}$	$0.49^{+0.38}_{-0.34}$			

[a] The errors quoted represent the 68% confidence intervals.

[b] f_{comp} corresponds to the fraction of composites with respect to the whole AGN+composite population.

5. RESULTS

5.1. *Model setup*

We consider 2 models: one with a double power law luminosity function, "PL" hereafter; and one with a modified Schechter luminosity function "MS" hereafter. For both models, we fix the shape parameters, α and β or σ. We experimented with allowing those to vary as well, but there was too much degeneracy in the solutions for the results to be illuminating. For the PL model we adopt $\alpha = 2.6$, but did a grid search with different β values. We found that while the solutions were degenerate (you can reach formally equally good fits with a wide range of β values), the

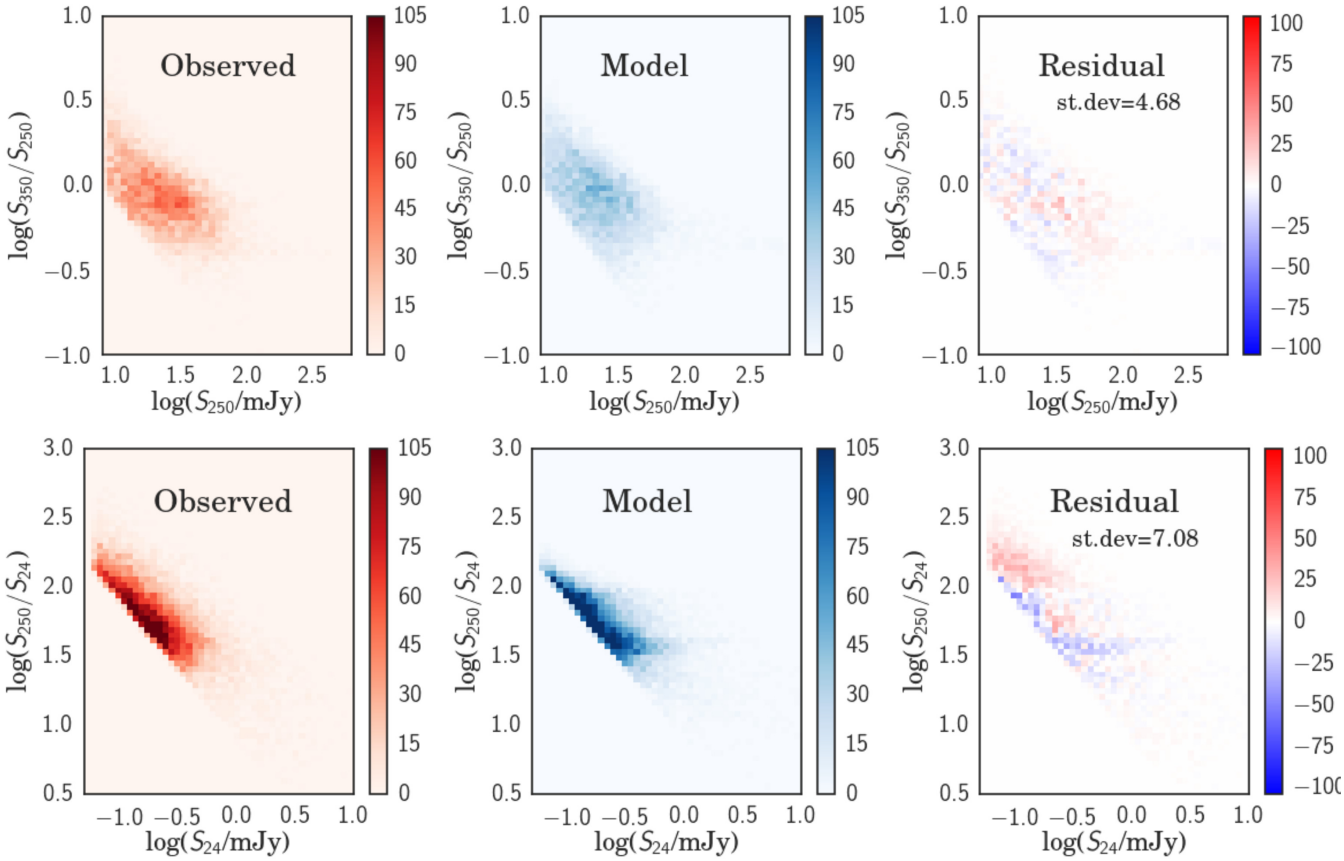

Figure 4. *Top:* Observed data, model and residual for the SPIRE 250μm-selected sample fit. *Bottom:* Observed data, model and residual for the SPIRE 24μm-selected sample. In both cases, we show the results for the PL model. In the residual panel, we also show the standard deviation of the non-zero pixels.

comparison with external constraints (see Section 5.4) showed $\beta = 0.6$ to work best. These values are consistent with the local LF measurements (see Figure 1). For the MS model we adopt $\alpha = 1.15$ and $\sigma = 0.52$ (as in Gruppioni et al. 2013). We left $log(\Phi_0^*)$ and $log(L_0^*)$ as free parameters whose allowed ranges were restricted to be consistent with the dispersion in observed local LFs (see Figure 1). The local AGN fraction of $10^{12} L_\odot$ galaxies, $f_{AGN,0}$, was restricted to the range 0.1-0.5, consistent with the uncertainty in Figure 3.

Our philosophy here is to input a relatively restricted model for the $z \sim 0$ LFs, but assume no knowledge of how the luminosity function evolves at higher redshifts. Obviously this is not typically necessary since we do have numerous studies in the literature that do give constraints on the later, but in this first SurveySim paper we aimed to test how close we get to these prior studies with as little restriction on our models as possible. To that effect, the evolutionary parameters are sampled in the range -6 to 6, allowing for both strong negative and positive luminosity/density evolution. The break redshifts, z_{bp}, z_{bq} are also very allowed a wide range from 0.5-3.5. The composite fraction, f_{comp} is allowed to vary from 0 to 1. We found that similarly wide ranges on the AGN fraction evolutionary parameters while allowing good overall fits to the total LF, consistently underestimated the AGN LF. To alleviate this we applied still wide, but more restricted ranges. Specifically, t_1 is allowed for vary from -1 to 6, t_2 varies from -6 to 6 and z_{bt} from 1.5-3.5.

5.2. *Simulating the Herschel/SPIRE Sources*

The models described in the previous section were fit individually to the two HerMES datasets described in Section 2. For the SPIRE-selected sample we used the $log(S_{350}/S_{250})$ vs. $log(S_{250})$ diagnostic plot and for the MIPS-selected sample we used $log(S_{250}/S_{24})$ vs. $log(S_{24})$ plot[10]. The goal of this exercise was to test the degree to which our conclusions depend on the specific diagnostic and/or sample selection applied.

Figure 4 shows examples of the observed, model, and residual images for each of the two samples, based on the PL

[10] We also tested that the results remain qualitatively the same when the $log(S_{350}/S_{250})$ vs. $log(S_{250})$ plot is used for the MIPS-selected sample as well.

model. We find that the residuals are fairly low, at roughly 10% level. The residual sub-structure, particularly clear in the bottom-middle panel in Figure 4, is due to discontinuous jumps in the SED library (see Figure 2), while some of the residuals are unaccounted for populations (Section 6.1). We examine the degree to which these fits are consistent with the models being a good representation of the data and its uncertainties in Section 5.3.

Table 4.2 presents the best-fit parameters and their 68% confidence levels associated to our two models and for each of the HerMES samples. Given the typically significant $1\,\sigma$ uncertainties, we obtain consistent evolutionary parameters for all models.

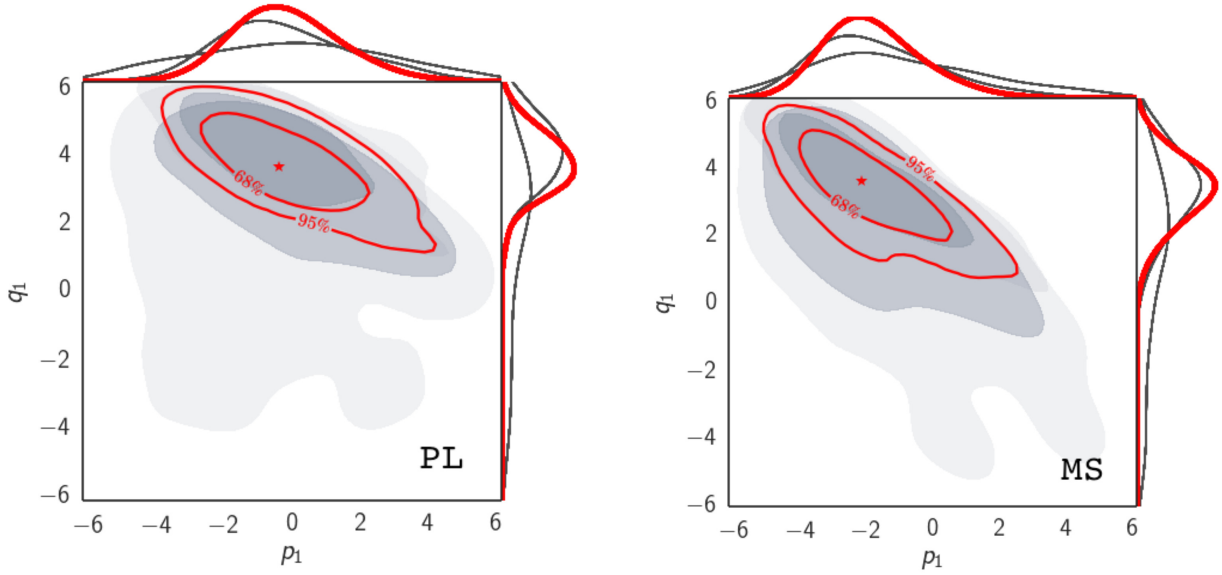

Figure 5. Joint MCMC posterior probability distributions for the evolutionary parameters p_1 and q_1. Individual runs with the SPIRE-selected and MIPS-selected samples are shown in greyscale. The joint probability distributions are marked with thick red contours. The same colors and linewidths apply to the 1d marginalized distributions shown on top and to the left of each panel. In all cases, the inner and outer contours represents respectively 68% and 95% confidence levels. The best joint value is plotted as a star. The key conclusion is that without any explicit redshift input, SurveySim favors strong luminosity evolution coupled with weak or negative density evolution until $z \sim 1-2$, where the best-fit break redshifts are found. As expected, including more data narrows the probability distributions, i.e. improves the constraints on the model parameters.

The MCMC chains for the two samples can be combined in order to find the joint probability distributions for the various parameters. This is given by $P_{joint} = \prod_{i=1} P_i$, where i is a given sample. Figure 5 illustrates this idea. This procedure highlights how in the future we can obtain even tighter constraints on the evolution of the luminosity function by combining fits to many different samples, especially ones selected in different parts of the IR SED (Bonato et al. 2016 in prep.).

Lastly, we examined each pair of parameters for each model for signs of correlations/degeneracy. By far the strongest correlation is found between the luminosity and density evolution parameters, p_1 and q_1 (shown in Figure 5).

5.3. *Goodness-of-fit*

Our χ^2-like statistic (see Table 4.2) is difficult to directly interpret in terms of goodness-of-fit because the data points here are neither uncorrelated nor necessarily obey Gaussian statistics. Therefore, we tested the goodness-of-fit by making 500 Monte Carlo realizations of the $log(S_{24})$ vs $log(S_{250}/S_{24})$ diagnostic plot. In each realization, we add Gaussian errors of $\sigma_{24} = 16\,\mu$Jy and $\sigma_{250} = 3.2$ mJy to the catalog fluxes, re-apply the selection criteria and re-generate the density plot. The same procedure is used to estimate the χ^2 between each realization and the original image as is used by SurveySim. The distribution in resulting χ^2 values is shown in Figure 6. This plot suggests that our model realization is consistent with the data within the errors given. However, the median $250\mu m$ error in the HerMES-COSMOS Spitzer-prior catalog is ~ 2 mJy or 50% lower that we have adopted here. Given the large uncertainties in the contribution of confusion and the effect of flux de-boosting to the total error budget, the larger error seems justified. Of course, this can also indicate the level of systematic uncertainty associated with the model, which is expected to be non-zero (see Section 6.1 for some likely systematics).

We performed the same test in the case of the SPIRE-diagnostic (i.e. $log(S_{250})$ vs $log(S_{350}/S_{250})$). Using the median catalog errors, we obtained a χ^2 distribution that peaked at 0.22 whereas the model best-fit has χ^2 of 0.11 (see

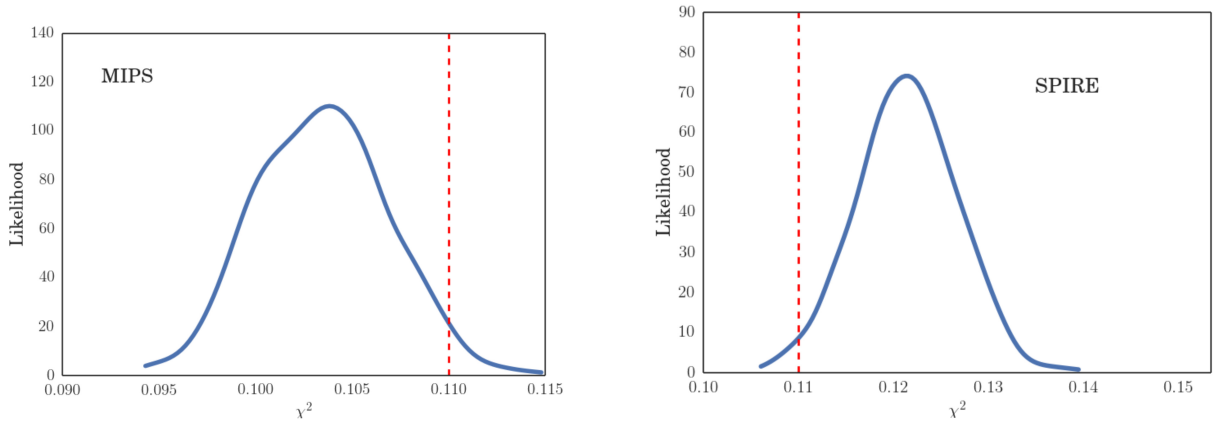

Figure 6. Goodness-of-fit estimates for the MIPS-diagnostic (*Left*) and the SPIRE-diagnostic (*Right*). In both cases, the distribution is the result of 500 Monte Carlo realizations applying Gaussian errors to the catalog photometry, regenerating the given diagnostic plot, and computing our χ^2-like statistic between each realization and the original diagnostic plot. The vertical dashed lines show the best-fit model χ^2 values (Table 4.2). See Section 5.3 for further details.

Table 4.2). Therefore, in contrast to the MIPS-diagnostic cases, here we are seeing evidence of overestimated errors. We needed to lower the nominal SPIRE errors substantially (by a factor of 2.5) to bring our Monte Carlo simulations in agreement with the model best-fit values (this is what is plotted in Figure 6 *Right*). The bulk of the error in these SPIRE catalogues is confusion errors. A likely cause of this apparent error reduction is the fact that confusion errors on S_{250} and S_{350} are correlated, simply because boosting one due to the underlying faint source distribution would inevitably boost the other as well. Therefore, the error on the SPIRE color would be less than implied if the errors on each flux density are assumed to be independent. Ultimately however, we emphasize that this is a χ^2-like statistic that should be treated with caution as it is also affected by the amount of "white space" in the diagnostic as well as how the model and observations compare in the pixels that are populated. As we discuss in Section 6, we plan to expand SurveySim to allow for other statistics so that the effect thereof can be explicitly tested.

5.4. *Comparison with external data*

In Figures 7-13, we compare our best-fit model results, based on the joint probability distributions (see Table 4.2), with a wide range of observational constraints. In all cases, the 68% uncertainties are shown as greyscale bands. These are derived very conservatively simply by looking at the parameter posterior probability distributions in the MCMC chains where the first half of the chain is removed (i.e. the burn-in period). The errorbars can be significantly reduced by making additional cuts in χ^2 which effectively only keep chain links close to the best-fit solution. However, this is not going to be representative of the full probability distribution. Therefore, the substantial uncertainties shown are genuine and encompass all possible solutions, rather than representing the uncertainty on the best-fit model alone. The later can be seen as the spread among the best-fit solutions of multiple SurveySim runs, and as expected this is substantially smaller than the nominal 68% uncertainty shown in the plots. While for simplicity we only show this once (in Figure 7), we tested that this is the case for all the comparison plots shown. The following sub-sections describe the different comparisons.

5.4.1. *Luminosity functions*

In Figure 7*top*, we compare our best-fit models with direct measurements of the total IR LF at three representative redshifts. Here, corrections for slight differences in $L_{\rm TIR}$ definitions or cosmology are not applied, but these are very small based on our experience with Figure 1 where they are applied. The significant scatter is therefore dominated by selection biases. We find that our models, especially the PL model, achieve excellent agreement with these literature values. Note that while the nominal 68% uncertainties are quite large, the best-fit solutions are robust. This can be seen in Figure 7*middle* where the better fitting PL model was run 6 times (similar results are found for the MS model as well). The scatter between these solutions is significantly smaller than the nominal uncertainty per run. Lastly, in Figure 7*bottom*, we compare the best-fit AGN LFs with the observed AGN IR luminosity function of Lacy et al. (2015). Both models are reasonably consistent with the data from Lacy et al. (2015). Again the spread between multiple runs here is much smaller than the nominal uncertainty per run.

5.4.2. *Number counts*

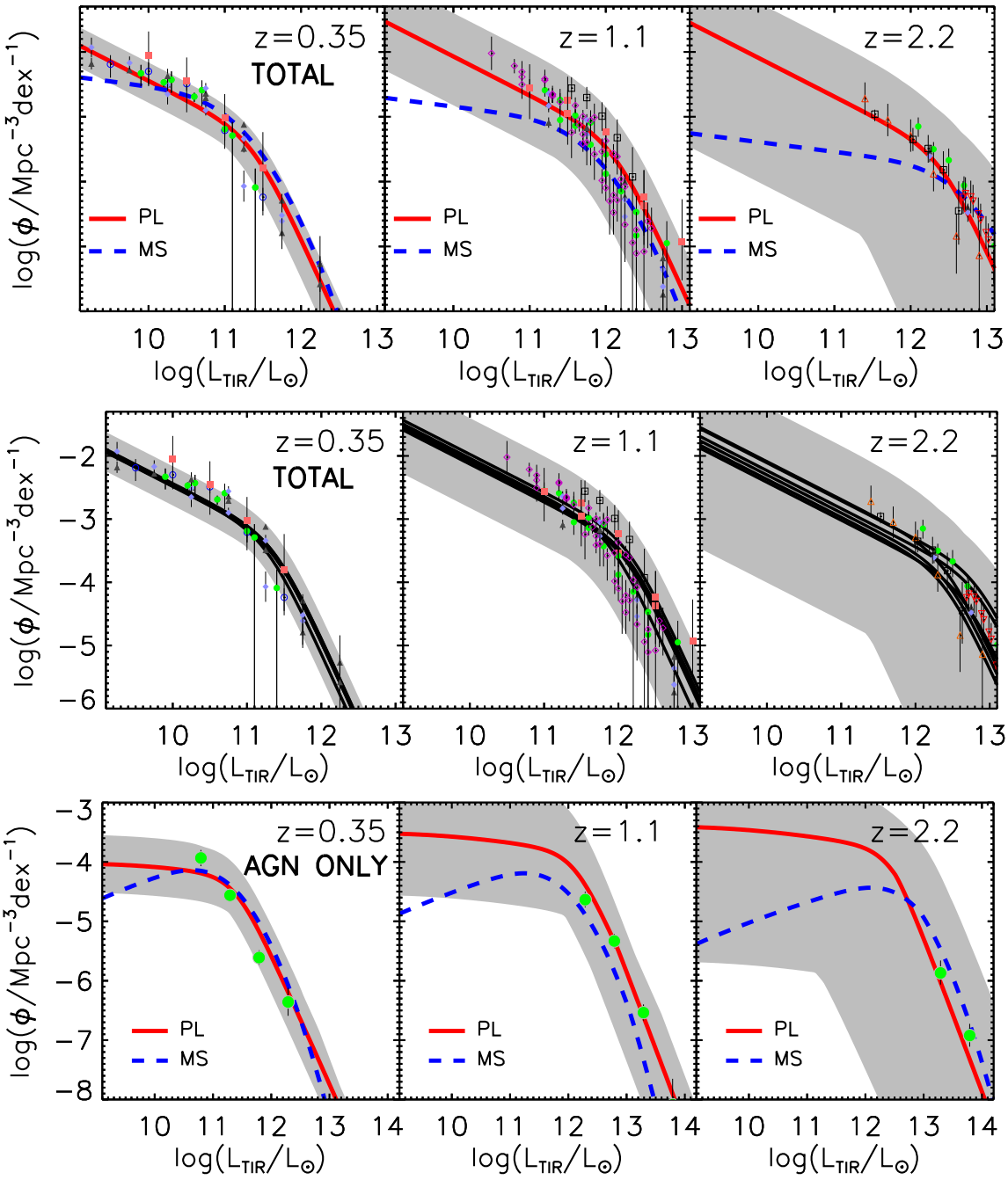

Figure 7. *Top:* Comparison, at 3 different redshifts, between our model total IR LFs and observational determinations taken from the datasets: Magnelli et al. 2009 (violet open diamonds), Magnelli et al. 2011 (orange open upward triangles), Magnelli et al. 2013 (green filled circles), Caputi et al. 2007 (black open squares), Lapi et al. 2011 (red open downward triangles), Le Floc'h et al. 2005 (lobster filled squares), Rodighiero et al. 2010 (pale blue filled diamonds), Gruppioni et al. 2013 (grey filled upward triangles) and Vaccari et al. 2010 (blue open circles). For clarity, we only show the 68% confidence levels for the PL model. These uncertainties are large as expected given that the model has been run 'blind' – i.e. without any information on the redshift evolution of the LF. Despite using only the color-magnitude distribution of HerMES sources, SurveySim is able to arrive at a good solution. *Middle:* The PL model run 6 times. Note that the spread among the best-fit solutions here is significantly smaller than the uncertainty per run. This is seen for all comparison plots in this section, although only shown explicitly here. *Bottom:* The same for the AGN IR LFs. Here we compare to the observational determinations by Lacy et al. (2015).

Figure 8 compares our best-fit Euclidean normalized differential number counts with observational data at 9 different wavelengths (from 15 to 1100 μm). Figure 9 does the same for the AGN number counts at 15 and 24μm. To first order the agreement between model and counts is good, especially for the PL model. The MS model tends to underestimate

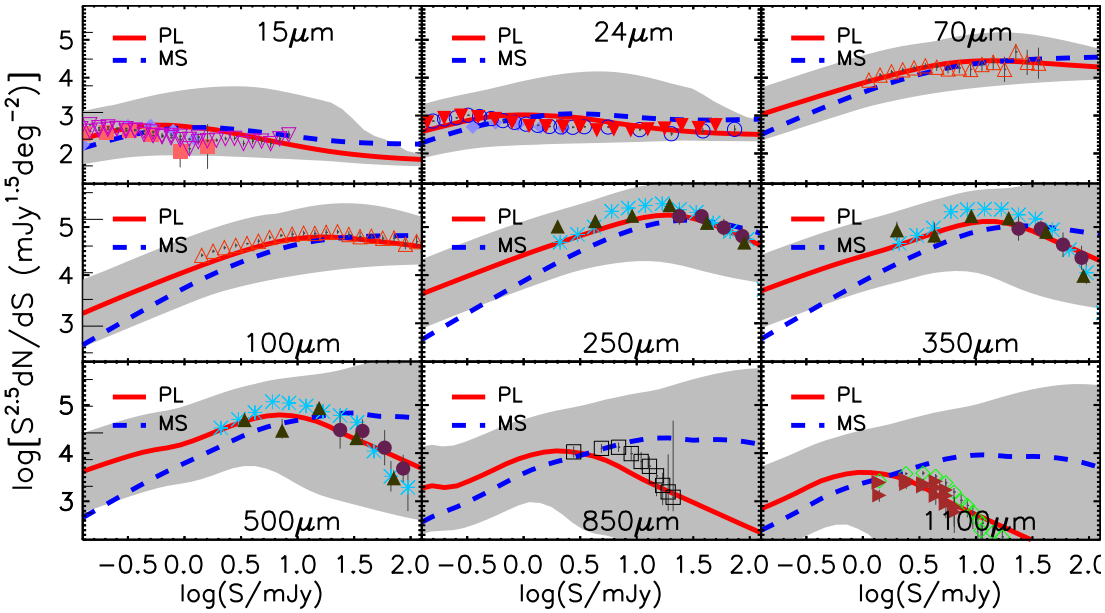

Figure 8. Model differential number counts compared with observational determinations taken from: Takagi et al. 2012 (at 15 and 24μm; pale blue filled diamonds), Hopwood et al. 2010 (15μm; lobster filled squares), Pearson et al. 2010 (15μm; violet open downward triangle), Papovich et al. 2004 (24μm; red filled downward triangle), Béthermin et al. 2010 (24μm, blue open circles), Berta et al. 2011 (70 and 100μm, orange open upward triangles), Béthermin et al. 2012b (250, 350 and 500μm, cyan stars), Oliver et al. 2010 (250, 350 and 500μm, purple filled circles), Glenn et al. 2010 (250, 350 and 500μm, olive filled upward triangles), Coppin et al. 2006 (850μm, black open squares), Scott et al. 2012 (1100μm, green open diamonds) and Hatsukade et al. 2011 (1100μm, brown filled right-facing triangles). We show the uncertainty regions associated with the PL model.

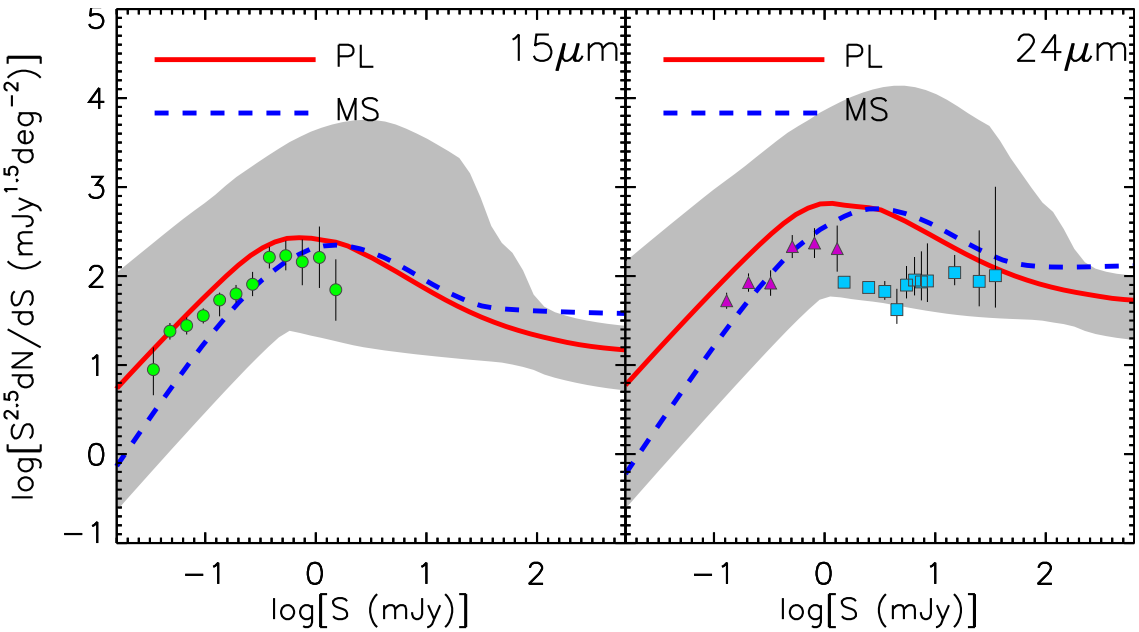

Figure 9. Our model AGN only 15 and 24μm counts compared with data from Teplitz et al. 2011 (15μm; green circles), Brown et al. 2006 (24μm; cyan squares) and Treister et al. 2006 (24μm; violet triangles). The agreement is generally good, the PL model shows slightly overestimated 24μm counts.

the counts at the faint-end, which is likely related to its adopted shallower faint-end slope (see Figure 7). Our results clearly favor a steeper faint-end slope. Even for the PL model there are some significant deviations. First, there is an indication of underestimating the peak of the SPIRE counts which may be due to the overly simplistic SED library

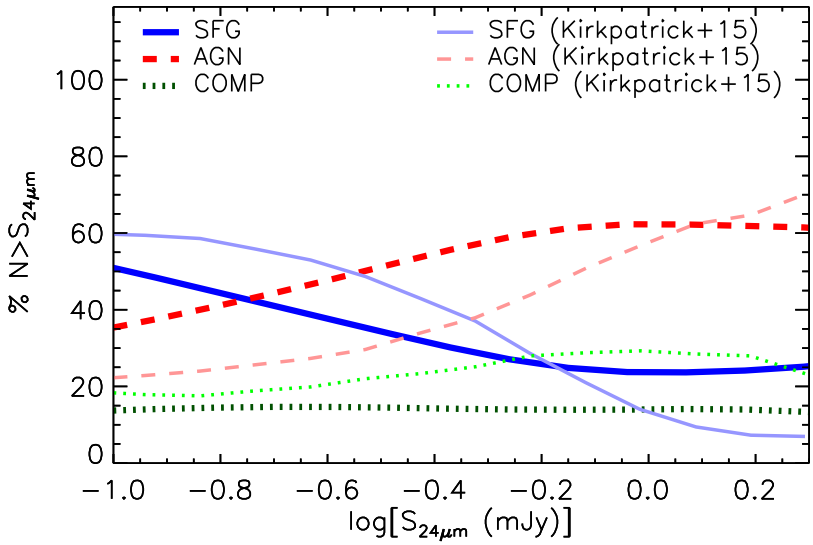

Figure 10. The relative contribution of SFGs, AGN and Composites to the the $S_{24\mu m}$ population as a function of flux. This plot mimics the same plot in Kirkpatrick et al. (2015).

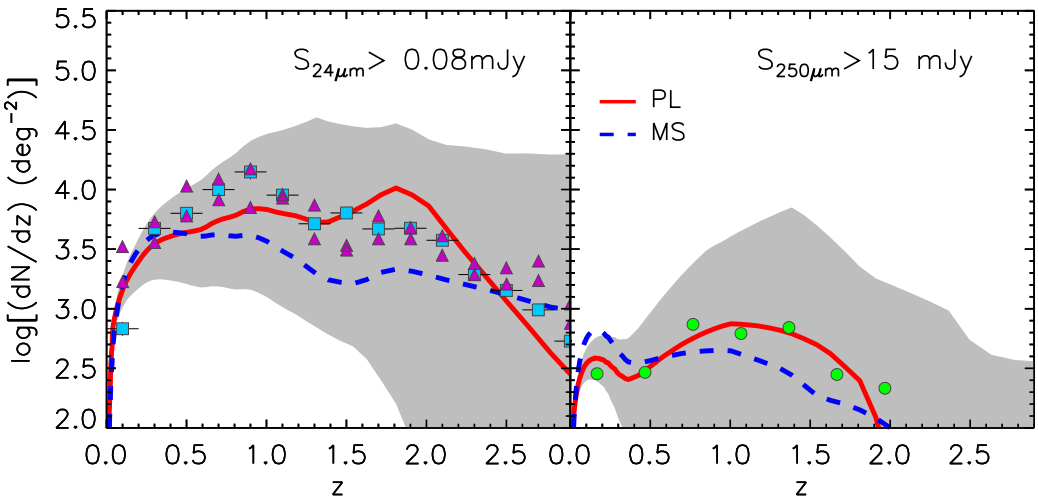

Figure 11. Model redshift distributions for 24μm-selected and 250μm-selected sources considering the flux limits given in each panel. These are compared with observational determinations taken from Le Floc'h et al. 2009 (cyan squares), Rodighiero et al. 2010 (violet triangles) and Casey et al. 2012 (green circles). Note that at 250μm, we only include star-forming galaxies in our model redshift distributions as the comparison Casey et al. 2012 sample is composed of star-forming galaxies. The Casey et al. (2012) redshift distribution is also corrected for spectroscopic completeness (as shown in their Figure 12). We show the uncertainty regions associated to the PL model.

(see Section 6.1. Second, the sub-mm counts are also underestimated. Part of this may be due to missing populations again, such as missing very cold SED sources (see Bonato et al. in prep.), but the biggest reason here is that our results are less reliable beyond $z \sim 2$ due to the paucity of sources in that regime in our samples. Because of this, we did not show luminosity functions at higher redshifts in Figure 7, but our results do underestimate the higher-z regime as shown in Section 5.4.5. The (sub-)mm counts however have a significant contribution from sources at $z > 2$ which

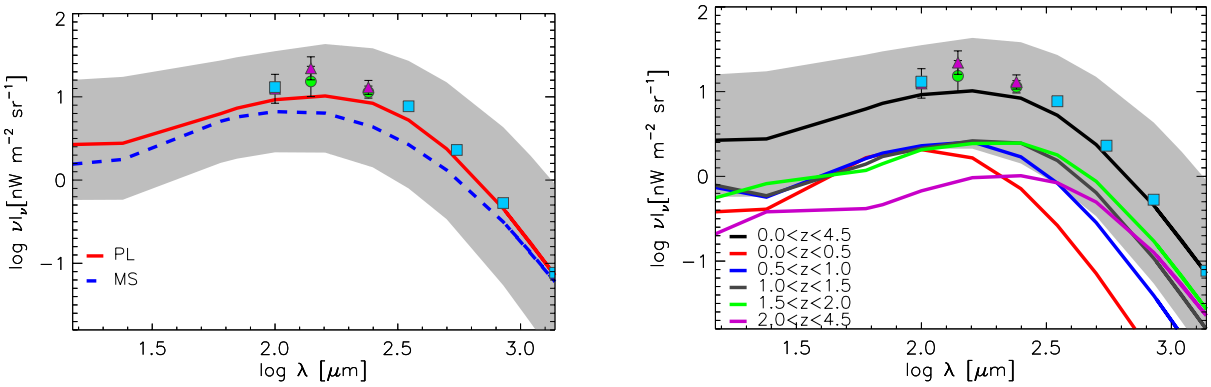

Figure 12. Comparison between our best-fit models and observed CIB determinations from Lagache et al. 1999 (green circles), Wright 2004 (violet triangles), and Planck Collaboration et al. 2014 (cyan squares). *Left* The total CIB for our PL and MS models (the uncertainty region is associated to the PL model). *Right* Contribution to the CIB by galaxies in different redshift ranges for the PL model (the uncertainty region is associated to the total).

accounts for the discrepancy seen here. The large excess in the (sub-)mm counts seen for the MS model is the result of the vastly overestimated luminosity function beyond $z > 2$ in this case. For example, see Table 4.2 where q_2 for the MS model is 2.85 vs. -1.40 for the PL model – the errorbars in both cases are huge due to the poor constraints in this regime). Lastly, we underestimate the faint-end of the 15 and 24μm counts for star-forming sources. We see this because for the AGN counts alone (Figure 9) we see a bit of an excess, which may be due to the studies besides these counts data requiring X-ray counterparts which may miss the most obscured AGN, but is likely also affected by the uncertainties in the low-L end of the AGN luminosity function.

Figure 10 shows explicitly the relative contribution of SFG, AGN and Composites to the 24μm population. Our model is in broad agreement with the empirical determination of this breakdown presented in Kirkpatrick et al. (2015). However, as indicated above, our model does tend to have a stronger AGN component relative to the SFG component. We believe this issue might also arise due to limitation of the SED template library which we speculate on in Section 6.1.

5.4.3. *Redshift distributions*

Figure 11 compares our model predictions with the observed redshift distributions at 24μm (*left-hand panel*) and at 250μm (*right-hand panel*). The 1 σ uncertainties here are very large, but again the spread between best-fit solutions from multiple runs is much smaller. The overall agreement between predicted and observed redshift distributions is reasonably good. The slight deficit in the 24μm counts at $z \sim 0.5 - 1$ is consistent with the deficit observed in the faint-end of the 15μm and 24μm counts (see Section 5.4.2).

5.4.4. *Contribution to CIB*

The total contribution to the Cosmic Infrared Background (CIB) as well as its breakdown by redshift is shown in Figure 12. We reproduce the CIB quite well. In our model, the CIB is completely dominated by star-forming galaxies. Shortward of $\sim$100μm, the CIB is largely made up of $z < 1$ galaxies, whereas at longer wavelengths $z > 1$ galaxies dominate. This is consistent with the study of Jauzac et al. (2011), who found that 81% of the 70μm CIB was emitted at $z < 1$.

5.4.5. *Star-formation rate density*

Figure 13 shows a comparison between our model predictions for the SFR density with data from the recent review of Madau & Dickinson (2014). For this comparison, we needed a few additional assumptions. Our model includes SFGs, AGN, and Composites, but each population has some star-formation rate which needs to be included in the total. Roebuck et al. (2016, in prep.) have TIR AGN fractions that include the effects of host galaxy reprocessing of the AGN light and find median fractions of 4, 66 and 39% for SFG, AGN, and Composites respectively. These fractions are used to correct for the AGN contribution to TIR of each population. The remaining $L_{\rm TIR,SF}$ is converted to SFR using the same relation used in Madau & Dickinson (2014) (their Eq.11, with $\kappa_{IR} = 4.5 \times 10^{44}$ in cgs units)[11]. Our model recovers the increase in SFR density from today until $z \sim 2$. However, there is a slight deficit at $z \sim 0.5$-1. This

[11] In order to be completely consistent with Madau & Dickinson (2014) data, we also consider a $L_{\rm TIR}$ integrated in the 8-1000μmwavelength interval and their same integration of the luminosity function (i.e. down to the relative limiting luminosity, in units of the characteristic luminosity L^*, of $L_{min} = 0.03 \times L^*$).

is the result of the slightly higher weight of AGN relative to SFG in our model as already discussed in Subsection 5.4.2. Our model strongly underestimates the SFR density past $z \sim 2$, but this is a regime that is not probed by our data therefore the results therein are not well constrained.

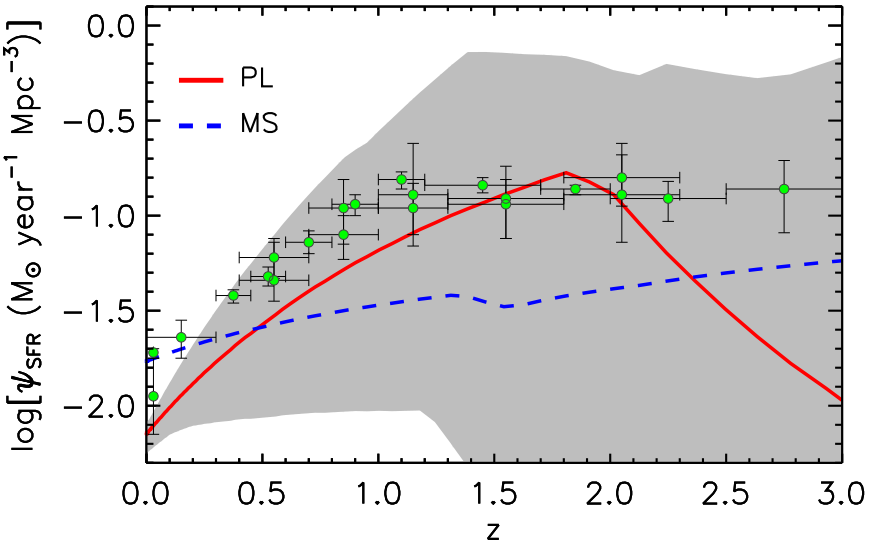

Figure 13. Comparison between our model SFR density estimates and observational average observational estimate given in the recent review Madau & Dickinson (2014). We show the uncertainty regions associated to the PL model.

6. DISCUSSION

6.1. *Need for expanded SED library*

In this paper, our approach has been to largely fix the shape of the LF and limit the range of the parameters governing it at $z \sim 0$ to be consistent with local observations, but we have allowed very wide ranges for the evolutionary parameters. For both models and both datasets, luminosity evolution is favored over density evolution until the break redshifts of $z \sim 2$. This is consistent with earlier studies (e.g. Caputi et al. 2007; Marsden et al. 2011; Gruppioni et al. 2013). Beyond these break redshifts, the evolutionary parameters are more uncertain since the data used in this paper are not significantly probing higher redshifts (see Figure 14). In the same figure, we address another key issue. The residual image for the MIPS fit shows that the faint and red tip of the distribution is underestimated by our model (see Figure 4). Figure 14*right* shows that our high redshift SFG SED templates do not quite reach this regime, i.e. they remain too blue. Examining the $z \sim 0$ SFG templates' coverage of the MIPS color-magnitude diagnostic plot (Figure 14*left*) shows these templates to be redder and fainter for a given redshift and luminosity than the corresponding $z \sim 1$ templates (see also Figure 2). Therefore the "missing" flux-color combination seen in the MIPS residual image (see Figure 4) can be reproduced by the local SFG SEDs (specifically local ULIRGs) at higher-z. However, if we adopted the local SFG templates wholesale we would achieve much worse fits overall since the entire high-z population shifts to redder colors rather than a small sub-part as observed. This difference in colors matches the observed difference between $z \sim 2$ main-sequence vs. off main-sequence star-forming galaxies presented in Elbaz et al. (2011). The galaxies behind our SED library at $z > 0$ are all on the main-sequence (for their redshifts) thus representing the more typical galaxies. Indeed, Béthermin et al. (2012a) include both galaxies on and off the main sequence in their galaxy evolution model and show that main sequence galaxies are by far the dominant component, consistent with the good fits we achieve with only this component. However, adding a small admixture of off main sequence galaxies at high-z would further reduce the observed residuals here.

An additional issue is that our library does not include Type 1 AGN, our AGN templates are more representative of obscured AGN (see Figure 2). For comparison, the SEDs of unobscured quasars (as in Richards et al. 2006) have a lower far- to mid-IR ratio i.e. less warm/cold dust relative to hot dust than our templates (see Xu et al. (2015) for further discussion on the IR SEDs of AGN). Constraining part of the AGN to have this type of SED ultimately lowers

the overall AGN component in our model and boosts the SFG galaxies (including slightly better fits to the SPIRE counts and SFR density). In Bonato et al. (in prep.) we experiment with expanding the SED library to allow for multiple star-forming galaxy SEDs and multiple AGN SEDs.

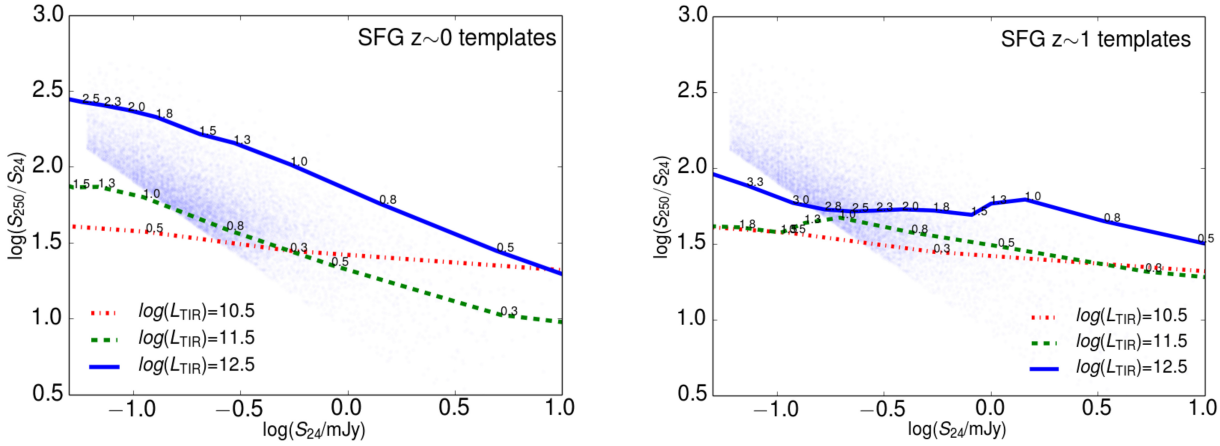

Figure 14. The $log(S_{24})$ vs. $log(S_{250}/S_{24})$ color-magnitude plot sampled by our SFG $z\sim0$ templates (*left*) vs. $z\sim1$ templates (*right*). Note that at higher redshifts/higher luminosities the $z\sim0$ templates are redder and fainter than the corresponding $z\sim1$ templates.

6.2. *The role of SurveySim*

During the last decade, many galaxy evolution models for IR galaxies and AGNs have been developed (e.g., Lagache et al. 2003;Valiante et al. 2009; Rowan-Robinson 2009;Le Borgne et al. 2009, Franceschini et al. 2010; Gruppioni et al. 2011; Marsden et al. 2011; Béthermin et al. 2011; Béthermin et al. 2012a; Cai et al. 2013). Most of these describe the evolution of multiple galaxy populations, assuming different SEDs and different evolutionary properties. A few take into account AGN (e.g. Valiante et al. 2009; Cai et al. 2013). MCMC has been used in a number of these to constrain the model uncertainties (e.g. Valiante et al. 2009; Marsden et al. 2011; Béthermin et al. 2011). Le Borgne et al. 2009 present a non-parameteric study of the IR luminosity function evolution based on inversion of the counts across the IR SED, an SED library and the constraint that the observed CIB should be matched. SurveySim builds upon this legacy with the following key features:

1) the current implementation of SurveySim does not use in the fit any direct redshift information (but we do include the redshift evolution of the SEDs encoded in our SED library), nor does it use CIB measurements to constrain the fit. It uses nothing but flux and color information. So the luminosity functions, number counts, redshift distributions, SFR density and contribution to the CIB shown in Fig. 7 represents predictions worked out by SurveySim, rather than model fits.

2) SurveySim does not hardwire either its SED template library, nor its model settings. Everything that is uncertain and that we are likely to want to update is fed into the MCMC-core externally. Even the focus on the IR regime here is external to the code – SurveySim adopts its wavelength array from the SED library it is fed.

3) SurveySim is written as a survey simulator, which means it models the selection function of a particular dataset, including the completeness function. In principle, if all parameters are fixed, SurveySim does not require a dataset input, but can be run in pure simulation mode allowing for predictions for future surveys, including redshift/galaxy type distributions etc.

4) SurveySim fits in color-magnitude space rather than the more common counts fits. This makes it easier to distinguish different populations. For example, the lack of sufficiently red high-z galaxies in the $log(S_{250}/S_{24})$ color is now easily apparent (see Figure 14), whereas such limitations can only be speculated on given discrepancies in the counts (e.g. Le Borgne et al. 2009).

5) SurveySim's MCMC structure allows for the constraints on the evolution of luminosity functions to be improved by combing the posterior probability distributions for the model parameters based on fitting different datasets. In Bonato et al. (2016 in prep.) we test our model on a wider set of IR-selected populations (from mid-IR to mm-selected) which allows us to both better constrain the evolution of the IR luminosity function, but also explicitly understand the selection biases inherent in each of these selections.

6.3. *Future improvements*

This is the first paper on SurveySim and presents version 1 of the code demonstrating its use in an application to the HerMES blind and *Spitzer*-prior SPIRE catalogs. SurveySim reproduces reasonably well the color-magnitude diagnostic plots for these catalogs, and the corresponding best-fit models are consistent with a wide range of external data. However, there are several areas in which we plan to improve the code as well as expand the range of its applicability. How different populations are included and how this mix evolves with redshift is currently the greatest uncertainty (one aspect of it is discussed above). Generalizing SurveySim beyond the 3 populations it currently assumed is the first step to allow SurveySim to operate in other parts of the spectrum including optical and near-IR on one hand, and radio on the other. SurveySim currently does not include the effects of lensing which are significant especially at brighter flux-densities in the (sub)mm regime. These can be included in a manner similar to the completeness curve. There are a couple of more technical improvements we would like to implement as well. For parameters that are known to be highly degenerate, a transformation can be made to diagonalize them in the correlation matrix. For example, p_1 and q_1 are found to be highly degenerate, but we can replace q_1 with new variable which is linear in p and has an adjustable slope and offset, trading an extra parameter for a much less degenerate space, which will allow for more efficient sampling and better defined parameter uncertainties. Lastly, we are planning to implement a choice of statistic (currently a 2-d χ^2) to test the results thereof on the performance of the code.

7. SUMMARY AND CONCLUSIONS

In this paper, we present the technical aspects of our new galaxy luminosity function simulation and fitting code, SurveySim. We demonstrate the use of the code by fitting the diagnostic plots – $log(S_{24})$ vs. $log(S_{250}/S_{24})$ and $log(S_{250})$ vs. $log(S_{350}/S_{250})$, which allows us to constrain the evolution of the IR luminosity function. Our results are consistent with previously published luminosity functions, number counts, redshift distributions, cosmic infrared background (CIB) total SED and redshift break-down, and the cosmic star-formation rate density. This demonstrates both that our statistical approach and SED library are working and provides a potential avenue for the study of large populations of galaxies when spectroscopic follow-up or even individual galaxy SED fitting are impractical.

We have made SurveySim public and welcome input from the community both in reporting any bugs founds, and suggestions for making it even more general in nature. The code, including the python interface and User's guide can be downloaded from http://cosmos2.phy.tufts.edu/~asajina/SurveySim.html.

REFERENCES

Allison, R., & Dunkley, J. 2014, MNRAS, 437, 3918

Bagaud, B., Martin, K., Abouelfath, A., et al. 2005, European Journal of Epidemiology, 20, 213

Berta, S., Magnelli, B., Nordon, R., et al. 2011, A&A, 532, A49

Béthermin, M., Dole, H., Beelen, A., & Aussel, H. 2010, A&A, 512, A78

Béthermin, M., Dole, H., Lagache, G., Le Borgne, D., & Penin, A. 2011, A&A, 529, A4

Béthermin, M., Daddi, E., Magdis, G., et al. 2012a, ApJL, 757, L23

Béthermin, M., Le Floc'h, E., Ilbert, O., et al. 2012b, A&A, 542, A58

Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, ApJ, 592, 819

Brooks, S. P., & Gelman, A. 1998, Journal of Computational and Graphical Statistics, 7, 434

Brown, M. J. I., Brand, K., Dey, A., et al. 2006, ApJ, 638, 88

Cai, Z.-Y., Lapi, A., Xia, J.-Q., et al. 2013, ApJ, 768, 21

Caputi, K. I., Lagache, G., Yan, L., et al. 2007, ApJ, 660, 97

Casey, C. M., Narayanan, D., & Cooray, A. 2014, PhR, 541, 45

Casey, C. M., Berta, S., Béthermin, M., et al. 2012, ApJ, 761, 140

Christlein, D., Gawiser, E., Marchesini, D., & Padilla, N. 2009, MNRAS, 400, 429

Coppin, K., Chapin, E. L., Mortier, A. M. J., et al. 2006, MNRAS, 372, 1621

Cowles, M. K., & Carlin, B. P. 1996, JASA, 91, pp. 883

de Jong, J. T. A., Yanny, B., Rix, H.-W., et al. 2010, ApJ, 714, 663

Dunkley, J., Bucher, M., Ferreira, P. G., Moodley, K., & Skordis, C. 2005, MNRAS, 356, 925

Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119

Franceschini, A., Rodighiero, G., Vaccari, M., et al. 2010, A&A, 517, A74

Glenn, J., Conley, A., Béthermin, M., et al. 2010, MNRAS, 409, 109

Goto, T., Arnouts, S., Inami, H., et al. 2011, MNRAS, 410, 573

Goulding, A. D., & Alexander, D. M. 2009, MNRAS, 398, 1165

Gruppioni, C., Pozzi, F., Zamorani, G., & Vignali, C. 2011, MNRAS, 416, 70

Gruppioni, C., Pozzi, F., Rodighiero, G., et al. 2013, MNRAS, 432, 23

Hastings, W. K. 1970, Biometrika, 57, 97

Hatsukade, B., Kohno, K., Aretxaga, I., et al. 2011, MNRAS, 411, 102

Hauser, M. G., & Dwek, E. 2001, ARA&A, 39, 249

Hogg, D. W., Baldry, I. K., Blanton, M. R., & Eisenstein, D. J. 2002, ArXiv Astrophysics e-prints, arXiv:astro-ph/0210394

Hopwood, R., Serjeant, S., Negrello, M., et al. 2010, ApJL, 716, L45

Jauzac, M., Dole, H., Le Floc'h, E., et al. 2011, A&A, 525, A52

Joachimi, B., & Taylor, A. 2014, in IAU Symposium, Vol. 306, IAU Symposium, 99–103

Johnson, S. P., Wilson, G. W., Tang, Y., & Scott, K. S. 2013, MNRAS, 436, 2535
Kirkpatrick, A., Pope, A., Sajina, A., et al. 2015, ApJ, 814, 9
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
Lacy, M., Ridgway, S. E., Sajina, A., et al. 2015, ApJ, 802, 102
Lagache, G., Abergel, A., Boulanger, F., Désert, F. X., & Puget, J.-L. 1999, A&A, 344, 322
Lagache, G., Dole, H., & Puget, J.-L. 2003, MNRAS, 338, 555
Lapi, A., González-Nuevo, J., Fan, L., et al. 2011, ApJ, 742, 24
Le Borgne, D., Elbaz, D., Ocvirk, P., & Pichon, C. 2009, A&A, 504, 727
Le Floc'h, E., Papovich, C., Dole, H., et al. 2005, ApJ, 632, 169
Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, ApJ, 703, 222
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Magnelli, B., Elbaz, D., Chary, R. R., et al. 2009, A&A, 496, 57
—. 2011, A&A, 528, A35
Magnelli, B., Popesso, P., Berta, S., et al. 2013, A&A, 553, A132
Marsden, G., Chapin, E. L., Halpern, M., et al. 2011, MNRAS, 417, 1192
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, JCP, 21, 1087
Negrello, M., Clemens, M., Gonzalez-Nuevo, J., et al. 2013, MNRAS, 429, 1309
Oliver, S. J., Wang, L., Smith, A. J., et al. 2010, A&A, 518, L21
Oliver, S. J., Bock, J., Altieri, B., et al. 2012, MNRAS, 424, 1614
Papovich, C., Dole, H., Egami, E., et al. 2004, ApJS, 154, 70
Pearson, C. P., Oyabu, S., Wada, T., et al. 2010, A&A, 514, A8
Petric, A. O., Armus, L., Howell, J., et al. 2011, ApJ, 730, 28
Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A30
Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81
Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, ApJS, 166, 470
Rieke, G. H., Alonso-Herrero, A., Weiner, B. J., et al. 2009, ApJ, 692, 556
Rodighiero, G., Vaccari, M., Franceschini, A., et al. 2010, A&A, 515, A8
Ross, S. 2013, in Simulation, Fifth edn., ed. S. M. Ross (Academic Press), 271 – 302
Rowan-Robinson, M. 2009, MNRAS, 394, 117
Saunders, W., Rowan-Robinson, M., Lawrence, A., et al. 1990, MNRAS, 242, 318
Scott, D. W. 1979, Biometrika, 66, pp. 605
Scott, K. S., Wilson, G. W., Aretxaga, I., et al. 2012, MNRAS, 423, 575
Takagi, T., Matsuhara, H., Goto, T., et al. 2012, A&A, 537, A24
Teplitz, H. I., Chary, R., Elbaz, D., et al. 2011, AJ, 141, 1
Treister, E., Urry, C. M., Van Duyne, J., et al. 2006, ApJ, 640, 603
Vaccari, M., Marchetti, L., Franceschini, A., et al. 2010, A&A, 518, L20
Valiante, R., Matteucci, F., Recchi, S., & Calura, F. 2009, NewA, 14, 638
Veilleux, S., Kim, D.-C., & Sanders, D. B. 1999, ApJ, 522, 113
Wang, L., Viero, M., Clarke, C., et al. 2014, MNRAS, 444, 2870
Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, ApJS, 154, 1
Wright, E. L. 2004, NewAR, 48, 465
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Xu, L., Rieke, G. H., Egami, E., et al. 2015, ApJS, 219, 18
Yuan, T.-T., Kewley, L. J., & Sanders, D. B. 2010, ApJ, 709, 884

ACKNOWLEDGEMENTS

AS and JM acknowledge support through NSF AAG#1313206. MB and AK are supported by NASA-ADAP13-0054. AP and AK acknowledge support from NSF AAG #1312418. MB thanks G. De Zotti and M. Negrello for their valuable comments and suggestions. NK thanks the Tufts Summer Scholars program, as well as Steven J. Eliopoulos and the Eliopoulos family, for their generous support.

APPENDIX

A. FITTING 2D HISTOGRAMS

The major hurdle to working with the 2D histograms discussed in Section 3.4 is the abundance of zero-valued histogram bins. We employ the bin size selection method outlined in Scott (1979), where the ideal bin size for a sample of size N with standard deviation σ is $\Delta\alpha = \frac{3.49\sigma}{\sqrt[3]{N}}$. This formula was experimentally found by Scott (1979) to best balance the need to minimize whitespace and statistical error while maintaining as much resolution as possible (see Figure A1).

The goodness-of-fit statistic we employ is the 2-D reduced χ^2, although altered to eliminate the "0" problem described above. These alterations mean good fits will always drop below unity in this statistic. To begin with, we assign the histogram bins standard Poisson errors, using tabulated values for bins less than 100, and otherwise assigning the bins $\sigma = \sqrt{N} + 1$ (Bagaud et al. 2005), where N is the number of galaxies in the bin. The Poisson bin uncertainties for both the model and observed 2-d histogram are added in quadrature leading to a 2-D χ^2 of:

$$\chi^2 = \frac{1}{n}\sum_{i,j=1}^{n,n}\frac{(O_{i,j} - M_{i,j})^2}{\sigma^2_{O_{i,j}} + \sigma^2_{M_{i,j}}} \tag{A1}$$

where i and j refer to column and row indices, and n is the side length in number of bins of our two-dimensional histogram. Here "O" and "M" refer to the observed and model histograms respectively. By employing Poisson errors, which are non-zero for bins with zero values, we ensure that the χ^2 formula is valid for all integer bin values. On the

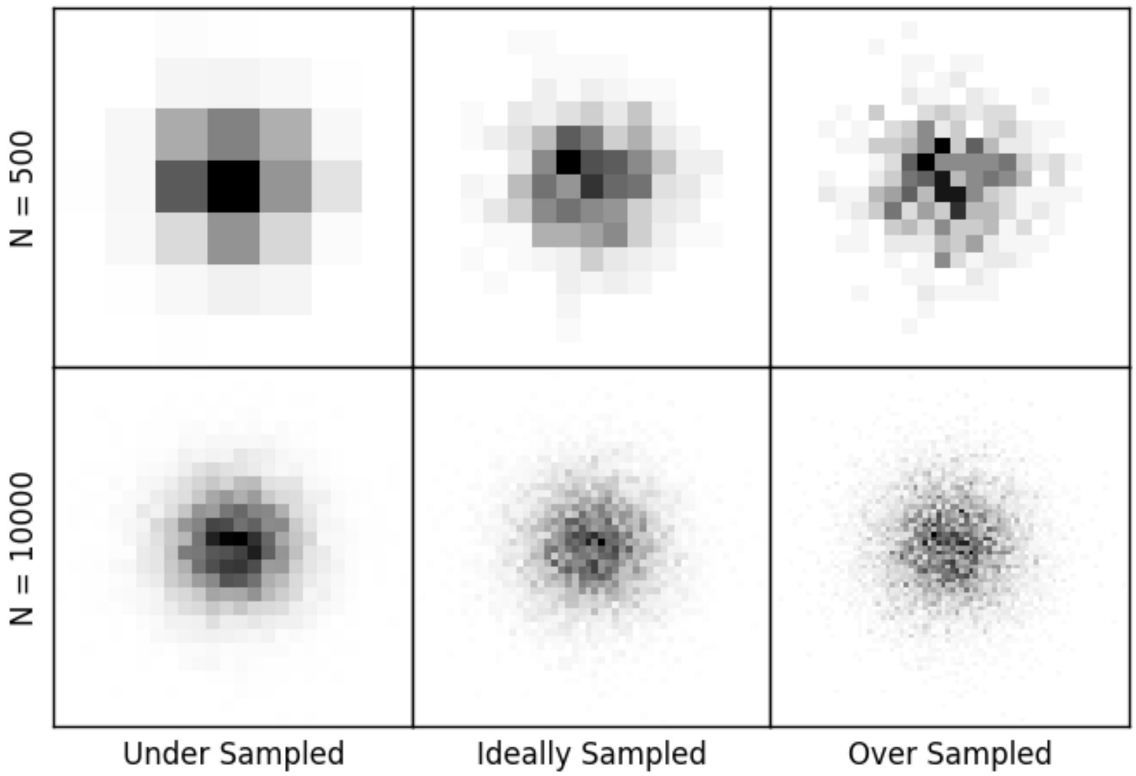

Figure A1. The effect of varying bin-width and sample size on the quality of the density histogram produced. The middle column shows histograms sized according to the Scott (1979) formula (which we adopt), while the left hand column shows histograms with twice this bin width, and the right hand column half this binwidth. The histograms on the right are too fragmented and contain too much whitespace in what should be high density areas. The histograms on the left solve this problem, but sacrifice detail, reducing the resolution with which models can be compared against one another.

other hand, this approach gives more weight to areas of higher density, as they have lower comparable error to those bins with only a few sources.

B. CONVERGENCE TESTING

Convergence testing is the least standardized aspect of MCMC implementations, with much more variance than sampling algorithms (see e.g. Cowles & Carlin 1996; Dunkley et al. 2005). If we define the ratio $r = \frac{\sigma_x^2}{\sigma_0^2}$ where σ_0^2 is the parameter variance of the prior distribution, and σ_x^2 is the variance achieved by the MCMC code, the code is said to have converged when r reaches 1 or slightly above depending on the precision to which a given parameter(s) are desired.

However, this ideal metric requires knowledge of the prior variance, which is one of the unknowns we are trying to determine. Various convergence tests represent means of approximating this ratio without knowledge of the underlying distribution variance (e.g. Dunkley et al. 2005; Allison & Dunkley 2014). Here we follow Brooks & Gelman (1998), who test convergence based on m independent chains of length $2n$ and a desired confidence limit α. The specific steps are:

1. Calculate the confidence intervals, $\mathrm{CI_m} = 100(1-\alpha)\%$, for each of the m chains from the last n chain links.
2. Calculate the total interval, $\mathrm{CI_T}$ for the combined n links of all m chains.
3. Calculate the convergence parameter, $\mathrm{R} \sim 1 + r$, as:

$$R = \frac{\mathrm{CI_T}}{\mathrm{CI_m}} \tag{B2}$$

The advantage of this method is that no assumption is made about the shape of the prior distribution, whereas other methods generally assume a Gaussian form (Brooks & Gelman 1998).

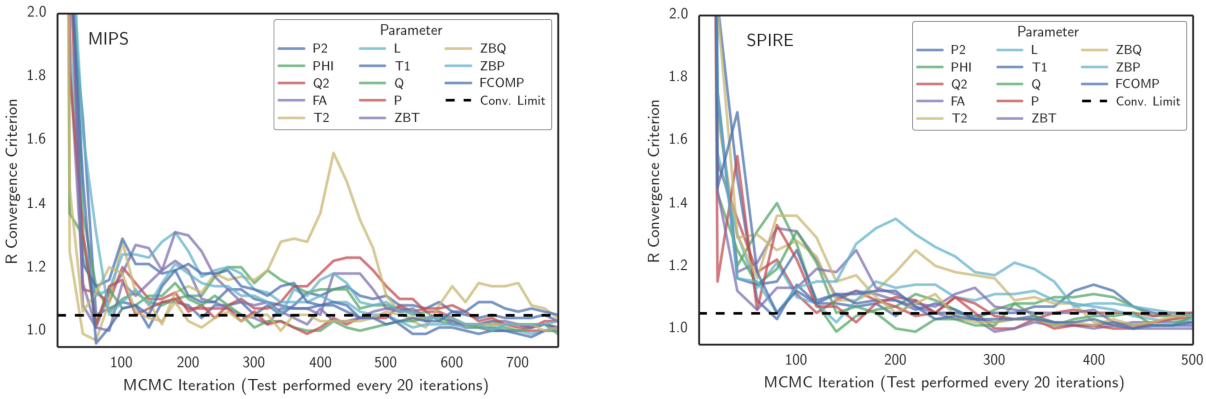

Figure B2. Convergence R criterion as a function of convergence test number, showing initial variance due to sampling change, and then steady decrease until all parameters lie below the convergence threshold. The left and right plots show similar trends for MIPS (left) and SPIRE (right) fields, with convergence time increasing with data volume. This demonstrates that convergence is robust to dimensionality and that convergence time is dominated by data volume. In addition, we found it is mainly dominated by a few of the high-variance, high-degeneracy parameters, as removing a few of the less degenerate parameters did not change the convergence time for either sample.

For example, if we want a set of chains which approximate the prior distribution to within a 95% confidence level, we set $r = 0.05$ and accordingly $R = 1.05$ as our convergence limit. The simulation is considered converged if all parameters have an R values below this limit. Figure B2 shows the R criterion as a function of time, where convergence was evaluated once for every 20 MCMC iterations. The initial learning of the covariance matrix and temperature can be seen in the initially higher variability, and we see the gradual descent as the sampling algorithm stabilizes.

For strict convergence ($r = 0.01$), Dunkley et al. (2005) found that optimal convergence can be expected in $\sim 220D$ steps, where D is the dimensionality of the sampling space. Our results suggest that at the $r = 0.05$ level, we can expect convergence to be proportional to the degeneracy of the space, as introducing additional parameters here which are non-degenerate does not significantly increase convergence time. In this paper, we adopt $R = 1.05$, i.e. we require 95% convergence confidence.