

HELP: XID+, The Probabilistic De-blender for Herschel SPIRE maps

P.D. Hurley,^{1*} S. Oliver,¹ and other HELP team members

¹*Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Falmer, Brighton BN1 9QH, UK*

Released 2002 Xxxxx XX

ABSTRACT

As part of the Herschel Extragalactic Legacy Project (HELP), we have developed a new prior based source extraction tool, *XID+*, to carry out prior source extraction on the Herschel SPIRE maps covered by HELP. *XID+* is developed using a probabilistic Bayesian framework which provides a natural framework in which to introduce prior information, and uses the Bayesian inference tool *Stan* to obtain the full posterior probability distribution on flux estimates. In this paper, we discuss the details of *XID+* and demonstrate the basic capabilities and performance by running it on simulated SPIRE maps resembling the COSMOS field, and comparing to the current prior based source extraction tool *DESPHOT*. We show that not only does *XID+* perform better on metrics such as flux accuracy and flux density error, we illustrate how obtaining the posterior probability distribution can help overcome some of the inherent issues associated with maximum likelihood based source extraction routines. We also discuss how additional work within HELP will be utilised by *XID+*. Software available at https://github.com/pdh21/XID_plus/.

Key words: galaxies: statistics – infrared: galaxies

1 INTRODUCTION

Ever since the discovery of the far-infrared background with the *Cosmic Background Explorer* (COBE) (Puget et al. 1996), surveys have aimed to observe and detect the sources responsible for the emission. Most of those sources are galaxies, with the far-infrared emission coming from dust.

While ground based observatories such as SCUBA (Holland et al. 1999), and more recently SCUBA-2 (Holland et al. 2013) and ALMA can make use of gaps in the infrared window to observe at the Rayleigh-Jeans tail of the far-infrared background, only space based facilities can observe at the peak. The first infrared space telescope, the InfraRed Astronomical Satellite (IRAS; Neugebauer et al. (1984)), observed the whole sky in four bands centred at 12, 25, 60 and 100 μm and revealed new populations of galaxies which were optically faint but luminous in the infrared (Soifer et al. 1984).

While the Infrared Space Observatory (ISO; Kessler et al. (1996)) and the Spitzer Space Telescope (Werner et al. 2004) have provided deep near and mid-infrared photometry over small fields, other smaller space based facilities such as AKARI (Murakami et al. 2007) and the Wide-field Infrared Survey Explorer (WISE, Wright et al. (2010)) have

surveyed the entire sky at mid to far-infrared and near to mid infrared wavelengths respectively. The most recent advancement in infrared astronomy has been made with the ESA Herschel Space Observatory (Pilbratt et al. 2010). Photometry from the Photoconductor Array Camera and Spectrometer (PACS Poglitsch et al. 2010) and Spectral and Photometric Imaging Receiver (SPIRE Griffin et al. 2010) instruments have given us an unprecedented view of the far-infrared Universe by providing observations that measure across the peak of the far-infrared background and at greater sensitivity and resolution than has been achieved at these wavelengths.

With surveys such as the Herschel Multi-Tiered Extragalactic Survey (HerMES; Oliver et al. (2012)) and the Herschel ATLAS survey (H-ATLAS; Eales et al. (2010)), over 1000 square degrees of the sky has been observed by the SPIRE instrument. However, the limited resolution of the SPIRE photometer and high source density at the SPIRE wavelengths introduces two problems when trying to combine SPIRE data with existing multi wavelength data.

The large beam size and low signal to noise of the SPIRE data results in large positional uncertainties for SPIRE detected sources. This makes matching to catalogues at other wavelengths difficult, as there may be more than one potential counterpart within the positional errors of the SPIRE detected source.

* Email: p.d.hurley@sussex.ac.uk

The second problem is that of source confusion. At the wavelengths observed by SPIRE, there is a high source density. This can often result in multiple sources being located within the SPIRE beam and, in extreme cases, lead to multiple astronomical sources being associated with one SPIRE detected source.

To obtain accurate photometry from the SPIRE maps, overcoming the source confusion problem is essential. One way to solve the problem is to use prior information to accurately separate out the flux in the SPIRE maps to the underlying astronomical objects. For example, if we know the precise location of a galaxy (e.g. from an optical image where resolution is better), we expect a galaxy to be found in the SPIRE maps at the same location.

Several techniques have been developed that utilise the positions of sources detected at other wavelengths, usually 24 μm and 1.4 GHz, to disentangle the various contributions from discrete sources to the SPIRE flux in a given beam element (e.g. Roseboom et al. 2010, 2011; Chapin et al. 2011). This process is made possible by the high correspondence between the 24- μm and 1.4-GHz populations and those observed at far-IR wavelengths; >80 per cent of the cosmic IR background at SPIRE wavelengths can be accounted for by 24- μm sources with $S_{24} > 25 \mu\text{Jy}$ (Marsden et al. 2009; Pascale et al. 2009; Elbaz et al. 2010, e.g.), while the strong correlation between the far-IR and radio luminosity is known to hold across a wide range in redshift and luminosity (e.g. Ivison et al. 2010). Up to the present day, most of these techniques have used a maximum likelihood optimisation approach. However, Safarzadeh et al. (2015) have recently used an MCMC based approach to fit PACS simulated maps, using prior information on position and flux.

The 24 μm XID algorithm, used by HerMES (Roseboom et al. 2010, 2011) used a source list taken from the 24- μm source catalogue to introduce positional priors for sources in the 250, 350 and 500 μm maps. The same method was also used to produce HerMES DR2 source catalogues, with 250 μm sources as priors (Wang et al. 2014, DESPHOT).

As part of the *Herschel Extragalactic Legacy Project* (HELP), we have developed an alternative prior based approach for source extraction in confusion dominated maps. Our new method, *XID+*, is built upon a Bayesian probabilistic framework which provides a natural framework in which to introduce additional prior information. By using the Bayesian inference tool, (*Stan, Team 2015b,a*) to sample full posterior distribution, we are also able to provide more accurate flux density error estimates, whilst avoiding some of the issues associated with the maximum likelihood and L1-constrained fitting approach used by *DESPHOT*. In this paper, we demonstrate that *XID+* outperforms *DESPHOT* when using just positional information. In Section 2 we discuss the algorithm, show how the software performs on simulated SPIRE maps in section 3 and discuss how *XID+* can make use of flux prior information, delivered by the HELP project (Hurley et al. in prep) in section 4.

2 XID+ ALGORITHM

The basic goal of *XID+* is to use the SPIRE maps to infer the likely SPIRE flux of sources we already know about. Bayesian inference is well suited to these requirements. It

allows the use of prior information and provides a posterior distribution of the parameter(s) after taking into account the observed data.

As discussed previously, we also want to provide a framework to do science directly with the maps rather than adding the additional step of first creating catalogues, which in essence is a form of data compression.

We therefore adopt a Bayesian probabilistic modelling approach for our *XID+* algorithm. It aims to:

- map out the posterior rather than the traditional maximum likelihood point estimate, thereby providing a precise measure of uncertainty.
- Extend the use of prior information beyond just using positional information about sources.

In the following section, we describe our *XID+* algorithm. As this algorithm builds upon knowledge gained from the original XID (a.k.a *DESPHOT*) algorithm used by HerMES (Roseboom et al. 2010, 2011; Wang et al. 2014), we describe *XID+* in the context of how it differs from *DESPHOT*.

2.1 Basic Model

Our data \mathbf{d} are maps with $n_1 \times n_2 = M$ pixels. Our model assumes the map are formed from n known point sources, with an unknown flux density \mathbf{f} . The point response function (PRF) tells us the contribution each source makes to each pixel in the map. As the SPIRE maps are mean subtracted, our model requires a global background level (B) and some unknown noise term (δ). Our map can therefore be described as follows:

$$\hat{\mathbf{d}} = \sum_{i=1}^M \mathbf{P}_i f_i + B + \delta \quad (1)$$

where \mathbf{d} is the image, \mathbf{P}_i is the PRF for source i , f_i is the flux density for source i , B is a global background estimate and δ is the noise term.

We can rewrite the above equation in the linear form:

$$\hat{\mathbf{d}} = \mathbf{A}\mathbf{f} + \delta \quad (2)$$

Where A is the pointing matrix, an $M \times n + 1$ matrix giving the contribution of sources and background to each pixel.

Our map \mathbf{d} will have an associated, and measurable, variance and possibly covariance between the pixels, which we define here as $\mathbf{N}_d = \langle \delta \delta^T \rangle$. We define the likelihood as the Gaussian probability function for the data given the flux densities

$$L(\mathbf{f}) = p(\mathbf{d}|\mathbf{f}) \propto |\mathbf{N}_d|^{-1/2} \exp \left\{ -\frac{1}{2}(\mathbf{d} - \hat{\mathbf{d}})^T \mathbf{N}_d^{-1} (\mathbf{d} - \hat{\mathbf{d}}) \right\} \quad (3)$$

The maximum likelihood solution to this equation can be found by setting $\chi = (\mathbf{d} - \hat{\mathbf{d}})^T \mathbf{N}_d^{-1} (\mathbf{d} - \hat{\mathbf{d}})$, finding the minimum and rearranging such that:

$$\hat{\mathbf{f}} = (\mathbf{A}^T \mathbf{N}_d^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{N}_d^{-1} \mathbf{A} \quad (4)$$

Equation 4 can be solved directly, either by brute-force matrix inversion or via other linear methods. As discussed in Roseboom et al. (2010, 2011); Wang et al. (2014), linear

approaches ignore prior knowledge that fluxes cannot have negative flux density. They are also incapable of discriminating between real and spurious sources, which can result in overfitting. To overcome these issues, Roseboom et al. (2011) used the non-negative weighted LASSO algorithm (Tibshirani 1996; Zou 2006; ter Braak et al. 2010).

LASSO is a shrinkage and selection method for linear regression and works by treating sources either ‘inactive’ and flux density set to zero, or ‘active’. It switches sources on one at a time, with the order determined by reduction in chi-squared gained by turning them on. The process continues until some tolerance is reached.

For XID+, we want to map out the entire posterior, $p(\mathbf{f}|\mathbf{d})$, rather than find the maximum likelihood solution. This has the benefit that it gives us more accurate information about how certain we are about the predicted fluxes. The posterior can be defined as:

$$p(\mathbf{f}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{f}) \times p(\mathbf{f}) \quad (5)$$

where $p(\mathbf{d}|\mathbf{f})$ is our likelihood, defined in equation 3 and $p(\mathbf{f})$ is our prior on the fluxes.

In our probabilistic framework, we can illustrate our model for the map, defined in equation 2 via a probabilistic graphical model (PGM). Figure 1 shows our PGM for our basic XID+ model, where boxes represent dimensions, open circles as random variables, dots as deterministic or (fixed) variables. For our simplest model, the sky co-ordinates of our sources are treated as deterministic (i.e. known). The PRF is also assumed to be known, and modelled as a Gaussian, with full-width half-maximum (FWHM) of 18.15, 25.15 and 36.3 arcsec for 250, 350 and 500 μm respectively. Both these deterministic variables are used to make the pointing matrix A which gives the contribution each source makes to each pixel j in the map. Each source has its own flux f which is a random variable. By multiplying f , A and adding our global estimate for the background B , we can make our model for the map, M .

2.1.1 Stan

Now we have our probabilistic model, we need a Bayesian inference tool capable of sampling from it to obtain the posterior. We use the Bayesian inference tool, *Stan*, which is ‘a probabilistic programming language implementing full Bayesian statistical inference with MCMC sampling’. *Stan* uses the adaptive Hamiltonian Monte Carlo (HMC) No-U-Turn Sampler (NUTS) of Hoffman & Gelman (2014) to efficiently sample from the posterior. It does this by using the gradient information, allowing fast traversing of high dimensional and highly correlated joint posterior distributions.

Stan has its own modelling language, in which one constructs probabilistic models. Our model for *Stan* can be found in Appendix A.

2.1.2 Estimating Convergence

As with all MCMC routines, one needs to run enough chains and run them long enough to be confident the global minimum has been found and that it has been thoroughly sampled.

As default, we run four separate chains from different

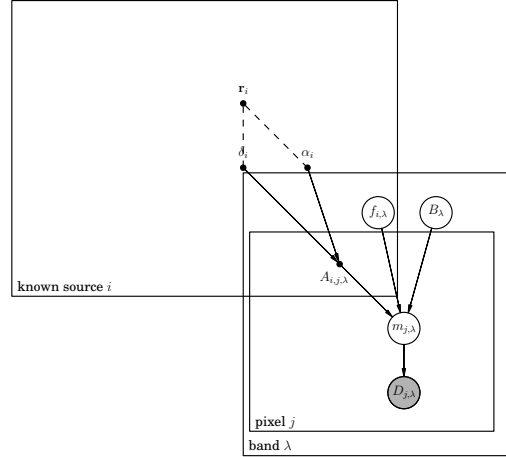


Figure 1. Our probabilistic model for XID+. Boxes represent dimensions, open circles as variables, dots as deterministic or (fixed) variables. Created with DAFT (<http://daft-pgm.org/>)

initial positions. We also discard the first half of the chain as ‘warm up’ to ensure the chains have converged to the posterior distribution. We then assess the convergence of each parameter by comparing the variation between and within chains using the diagnostics described in Gelman et al. (2013) which can be summarised as follows: Each chain is split in two and the between (B) and within chain (W) variance is calculated. B and W are then used to calculate the marginal posterior variance. This in turn can be used to estimate the potential scale reduction \hat{R} , which reduces to 1 as the number of iterations tends to infinity. An \hat{R} value > 1.2 suggests chains require more samples. We provide \hat{R} for each parameter.

Due to the nature of MCMC, samples from MCMC routines are correlated. Inference from correlated samples is less precise than from the same number of independent draws. In order to check there are enough independent draws we estimate the effective number of samples n_{eff} , defined in Gelman et al. (2013). We require n_{eff} to be 5 times the number of chains and provide the estimate for each parameter.

2.2 Map segmentation

The survey fields in HELP vary in size from 0.3 to 290 square degrees. Ideally, source photometry and background estimation would be done on the full image. In practice this will be computationally infeasible. *DESPHOT* segmented the map by locating islands of high SNR pixels enclosed by low SNR pixels.

The most optimum and transparent way of segmenting the map would use distinct tiles but fitting for sources within and beyond the tiling region. Using the fact that, conditional on the fluxes, the data from each tile is independent, the

posterior from the different tiles can be combined to give an overall posterior for the whole map using the following formula:

$$P(\mathbf{F}|D_1, \dots, D_n, D_A) = a \frac{\prod_{i=1}^n P(\mathbf{F}|D_i, D_A)}{P(\mathbf{F}|D_A)^{n-1}} \quad (6)$$

where n is number of tiles and $P(\mathbf{F}|D_A)$ is our prior on the flux, given some ancillary data. As the number of dimensions is highly dimensional, the only feasible way of multiplying highly dimensional probability distribution functions is to model them as multivariate Gaussians. We investigated fitting the fluxes in both normal and log space, however in neither case was the multivariate Gaussian approximation appropriate.

We have settled on a tiling scheme that has tiles that overlap in the map. This makes sure sources that lie near the edge of a tile, will be nearer the centre of the adjacent tile. As before, for each tile, we fit to sources both in and beyond. For our final posterior, for each source, we take the posterior from the tile that is optimum, i.e. the tile in which the source lies closest to the centre.

2.3 Uncertainties and Covariances

DESPHOT provides an estimate of the covariance matrix associated with the fluxes (\mathbf{N}_f) from $(\mathbf{A}^T \mathbf{N}_d^{-1} \mathbf{A})^{-1}$. Due to the Cramer-Rao inequality, this estimate is a lower limit. It also assumes the *DESPHOT* algorithm is linear, which is not strictly true having introduced LASSO and non-negative priors. As a result, the uncertainties are unreliable. For *XID+*, we have the full posterior, allowing the true variance to be properly characterised. This not only gives us a better estimate for marginalised uncertainty for each source, but it also provides the covariance information between sources (as seen in Figure 6).

3 SIMULATIONS

In order to test and quantify *XID+*, we use simulated SPIRE maps of the COSMOS field, a good example of a deep map i.e. where $\sigma_{conf.} \gg \sigma_{inst.}$. In order to get realistic clustering, we use the mock catalogues from the latest version of the Durham semi-analytic model, *GALFORM*. (Lacey et al. in prep., also described in Cowley et al. (2015)). The model is designed to populate Millennium-class, dark matter only, N-body simulations with a WMAP7 cosmology and minimum halo mass of $1.9 \times 10^{10} h^{-1} M_\odot$. The dust model is motivated by the radiative transfer code GRASIL (Silva et al. 1998) and can accurately reproduce the predictions for rest-frame wavelengths $\lambda_{rest} > 70 \mu\text{m}$.

A mock 100 μm input catalogue, similar to that expected from a PACS catalogue, is generated by taking the mock catalogue and making a cut at a flux limit of 50 mJy. We use this as our prior input catalogue for both *XID+* and *DESPHOT*. In order to compare performance, we look at three measures: precision, flux accuracy, and flux error density. For *XID+*, we only consider sources whose output median flux is above 1mJy. Likewise, with *DESPHOT*, we only consider sources which have a maximum likelihood flux of greater than 1mJy.

3.1 Flux Precision

Precision is a measure of how well the flux is believed to be constrained. For our posterior sample, this relates to how spread out the sample is and so we use the interquartile range (IQR) as our measure of precision. Figure 2 shows the 50th, 84th and 16th percentile (i.e. median and median $\pm \sigma$) for 7 bins in true flux. IQR is normalised and as a function of input flux for both *XID+* and *DESPHOT*. As one would expect, IQR/S_{True} decreases as a function of input flux, indicating a higher precision is achieved for the brighter sources. While 250 and 350 μm outputs achieve a similar level of precision, the outputs for 500 μm do not reach the same level of precision. In comparison to *DESPHOT*, *XID+* is marginally worse for all three bands, though as we show later, this is linked to a better estimate of uncertainty.

3.2 Flux Accuracy

Flux accuracy is a measure of how far away the estimated flux is from the truth. We use the difference between our median flux estimate from our posterior and the true flux from the simulation, normalised by true flux, as our estimate of flux accuracy. Figure 3 shows how flux accuracy changes as a function of input flux for all three bands. As before, we show the 50th, 84th and 16th percentile for 7 bins in true flux. For 250 μm , *XID+* reaches a zero level by $\approx 5\text{mJy}$, where as *DESPHOT* underestimates the flux for all but the very brightest sources. For 350 μm and 500 μm , a zero level is reached by $\approx 10\text{mJy}$, where as *DESPHOT* continues to underestimate for all but the very brightest sources.

3.3 Flux Error density

Flux error density measures how well accurate the estimated uncertainty is correct. Estimated flux values should be within one sigma of the true value 68.27% of the time and within 2 sigma 95.45% of the time. Figure 4 shows the flux error density (or Z score) as a function of input flux. For *DESPHOT*, uncertainties are assumed to have a normal distribution, truncated at zero. With *XID+*, we have the full posterior and no longer have to make an assumption on the shape of the uncertainty distribution. Instead, we calculate error density by taking the percentile at which the true flux value falls within the posterior, and convert the percentile to a corresponding sigma level.

For the 250 μm band, and sources under $\approx 25\mu\text{Jy}$, *XID+* produces a Z score distribution which that matches that expected if uncertainties are correctly estimated (i.e. distribution is centred around zero, with width ≈ 1). Above 25 μJy , the median Z score increases, indicating flux and uncertainty is being under estimated. In comparison, for all fluxes, the uncertainty distribution from *DESPHOT* is above 1, and increases to a Z score of over 2 for the brighter sources. This indicates that the majority of sources in *DESPHOT* lie over 1 σ away from their true flux. The flux error densities for 350 μm and 500 μm show a similar behaviour, though not as severe.

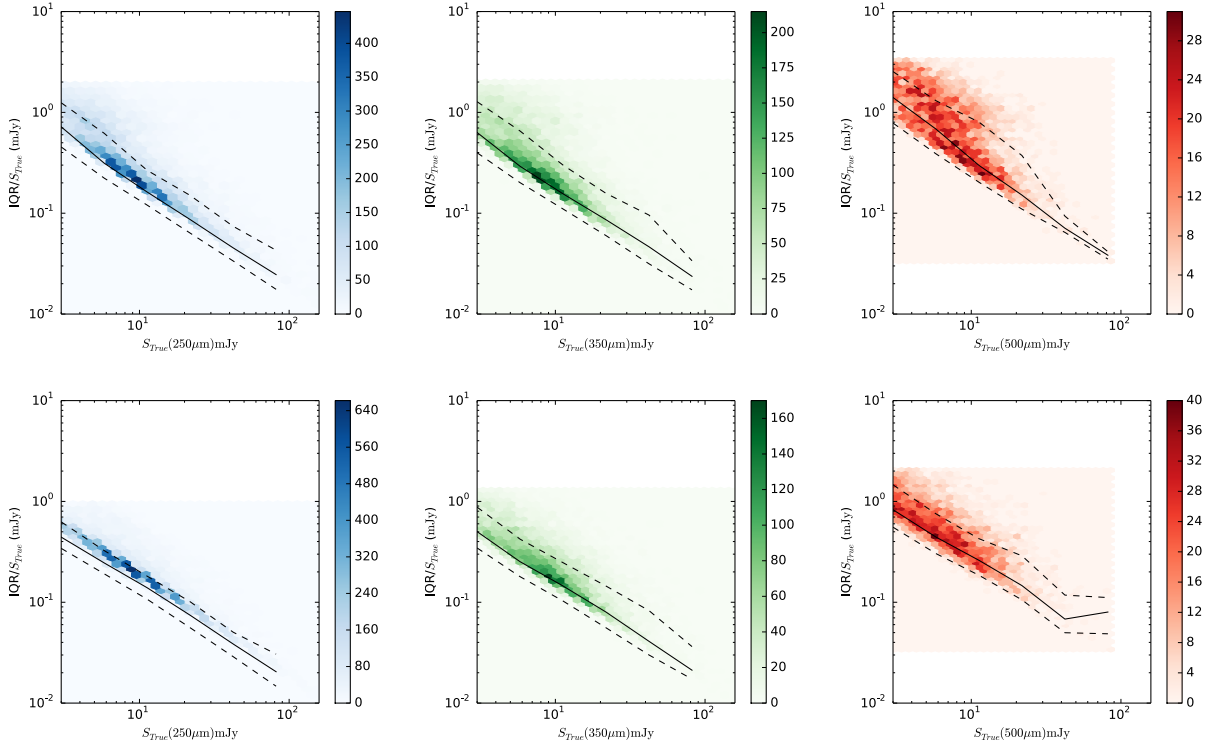


Figure 2. Inverse precision, or IQR of *XID+* (top) and *DESPHOT* (bottom) as a function of true flux, for the 250 (blue), 350 (green) and 500 (red) μm SPIRE bands. Black solid and dashed lines shows median and standard deviation respectively. Density of objects is illustrated via underlying colour.

3.4 Convergence

As described in Section 2.1.2, we provide \hat{R} as an estimate of convergence and \hat{n}_{eff} as a measure of independence within the sample. Figures 5 show the histogram for \hat{R} and \hat{n}_{eff} for the three bands, with our thresholds for the statistics shown by dotted lines. In our fit to the simulated SPIRE maps, we use four chains, each with 1500 iterations (half of which are discarded as warm up). This leads to over 99.99% of the sources having an \hat{R} and \hat{n}_{eff} within the threshold for all three bands, indicating our solution is well converged. In cases where convergence has not been reached, the number of iterations can be increased.

3.5 Remaining Noise

The purpose of using prior based source extraction is to overcome the confusion noise in the SPIRE maps, which, as measured by Nguyen et al. (2010) is 5.8, 6.3 and 6.8 mJy at 250, 350 and 500 μm , respectively. By looking at the residual map (i.e. $\mathbf{d} - (\mathbf{A}\mathbf{f} + \mathbf{B})$), we can estimate the remaining residual confusion noise in the maps.

We do this by measuring the standard deviation of the pixels in the residual map convolved with the PSF. This provides an estimate of the total noise remaining in the residual map ($\sigma_{tot.}$). In order to obtain an estimate of the confusion noise ($\sigma_{conf.}$), an estimate of the instrumental noise is made from the uncertainty map, and then removed in quadrature, such that $\sigma_{conf.}^2 = \sigma_{tot.}^2 - \sigma_{inst.}^2$. Table 1 shows the total and confusion noise from the residual Lacey simulated maps, for

	250 μm		350 μm		500 μm	
	$\sigma_{tot.}$	$\sigma_{conf.}$	$\sigma_{tot.}$	$\sigma_{conf.}$	$\sigma_{tot.}$	$\sigma_{conf.}$
<i>XID+</i>	1.9	1.3	2.2	1.9	2.6	1.9
<i>DESPHOT</i>	2.4	1.9	3.5	3.3	3.4	2.9

Table 1. Table showing the remaining total and confusion noise for the three SPIRE bands (in mJy), from running *XID+* and *DESPHOT* on the Lacey simulated COSMOS maps.

both *DESPHOT* and *XID+*, and for all three SPIRE bands. For *XID+*, we calculate a residual map (and remaining noise estimate) for each sample in the posterior, Table 1 shows the median from the sample.

As demonstrated by the values in Table 1, *XID+* reaches a confusion noise level substantially lower than the present in the maps. It is also lower than that achieved by *DESPHOT*. This is could be due to the fact we fit the background simultaneously with the sources and/or the LASSO algorithm in *DESPHOT* restricting the fit of close-by sources, hence increasing the flux variations in the residual map.

3.6 Correlated Sources

For sources that are close together, the uncertainty on the flux estimates can be correlated. One of the advantages of obtaining the full posterior is we get a proper estimate of uncertainty and its correlation. This is particularly apparent

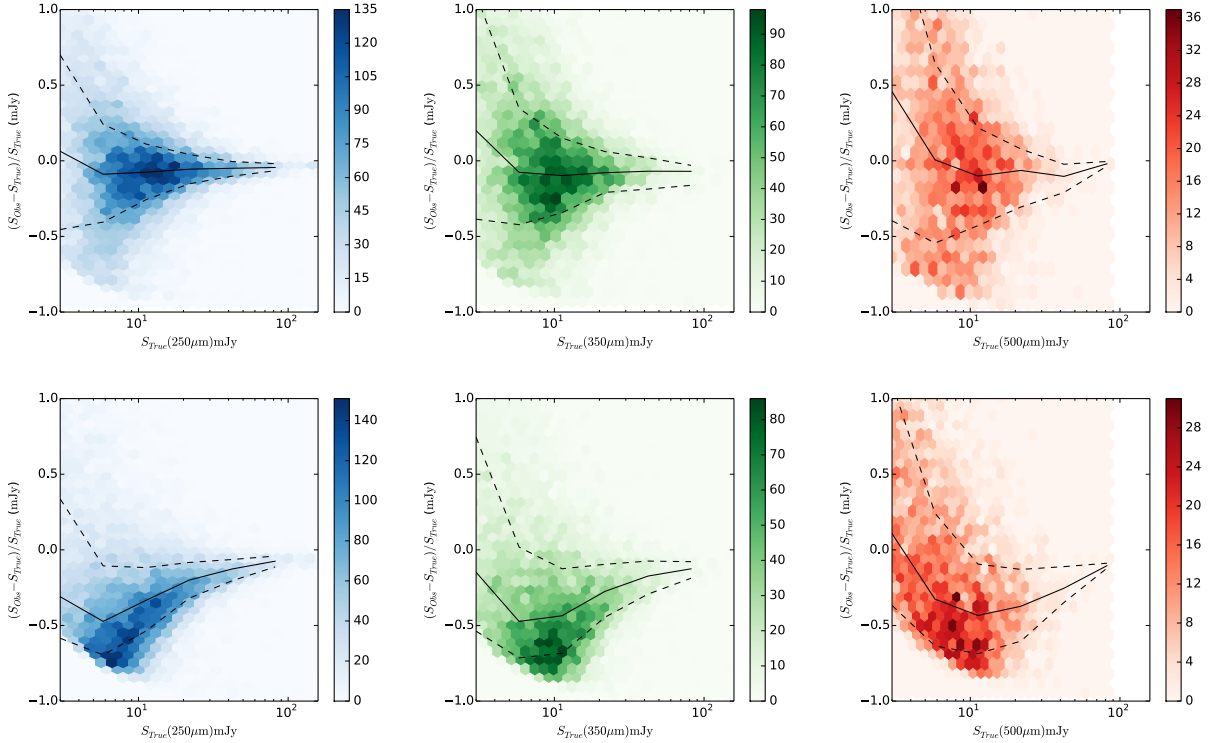


Figure 3. Accuracy of *XID+* (top) and *DESPHOT* (bottom) as a function of true flux, for the 250 (blue), 350 (green) and 500 (red) μm SPIRE bands. Black solid and dashed lines shows median and standard deviation respectively. Density of objects is illustrated via underlying colour.

when comparing flux estimates for with *DESPHOT*, which, by using the LASSO algorithm forces one source to have all the flux and the other nearby source to zero. Figure 6 shows an example of two correlated sources from our simulated map, and the estimate from both *XID+* and *DESPHOT*. The posterior provided by *XID+*, fully captures the correlated uncertainty, whereas the ‘winner takes all’ approach from *DESPHOT* clearly fails to estimate the true flux for both sources.

4 CONCLUSIONS

In this paper we have introduced the prior based source detection software, *XID+*. By using the Bayesian inference tool *Stan*, we are able to fully sample the posterior probability, which in turn gives a better understanding of the uncertainty associated with the source flux.

Having run *XID+* on simulated maps, we have shown this is extremely advantageous for maps that are confusion limited, such as the *Herschel* observations that are part of *HerMES*. In comparison to the current maximum likelihood based software *DESPHOT*, *XID+* performs far better in all the three main metrics; flux accuracy, precision and uncertainty accuracy.

By using a probabilistic approach, we have a framework where we can introduce prior information on the source fluxes in a transparent manner. For this basic version, we use a simple \log_{10} uniform prior, with bounds at 10^{-2} and 10^3 mJy .

As can be seen in Figure 6, where sources are correlated, if we have prior knowledge on the flux of one of the sources, it can help us determine a more precise flux for the other.

As demonstrated by Safarzadeh et al. (2015), these priors could come from fitting spectral energy distribution models to multi-wavelength ancillary data. Another alternative is to use machine-learning algorithms to ‘learn’ the expected flux from the statistical population. As part of HELP, the testing and benchmarking of suitable methods for deriving SPIRE flux priors will be presented in Hurley et al. (in prep). With *XID+*, we now have the tool to be able to use that prior information in a clear and transparent way to obtain better SPIRE flux estimates, which is essential if we are to extract the wealth of information contained within in the *Herschel* maps.

ACKNOWLEDGEMENTS

REFERENCES

- Chapin E. L., et al., 2011, Monthly Notices of the Royal Astronomical Society, 411, 505
- Cowley W. I., Lacey C. G., Baugh C. M., Cole S., 2015, MNRAS, 446, 1784
- Eales S., et al., 2010, PASP, 122, 499
- Elbaz D., et al., 2010, A&A, 518, L29
- Foreman-Mackey D., Price-Whelan A., Ryan G., Emily, Smith M., Barbary K., Hogg D. W., Brewer B. J., 2014, triangle.py v0.1.1
- Gelman A., Carlin J., Stern H., Dunson D., Vehtari A., Rubin

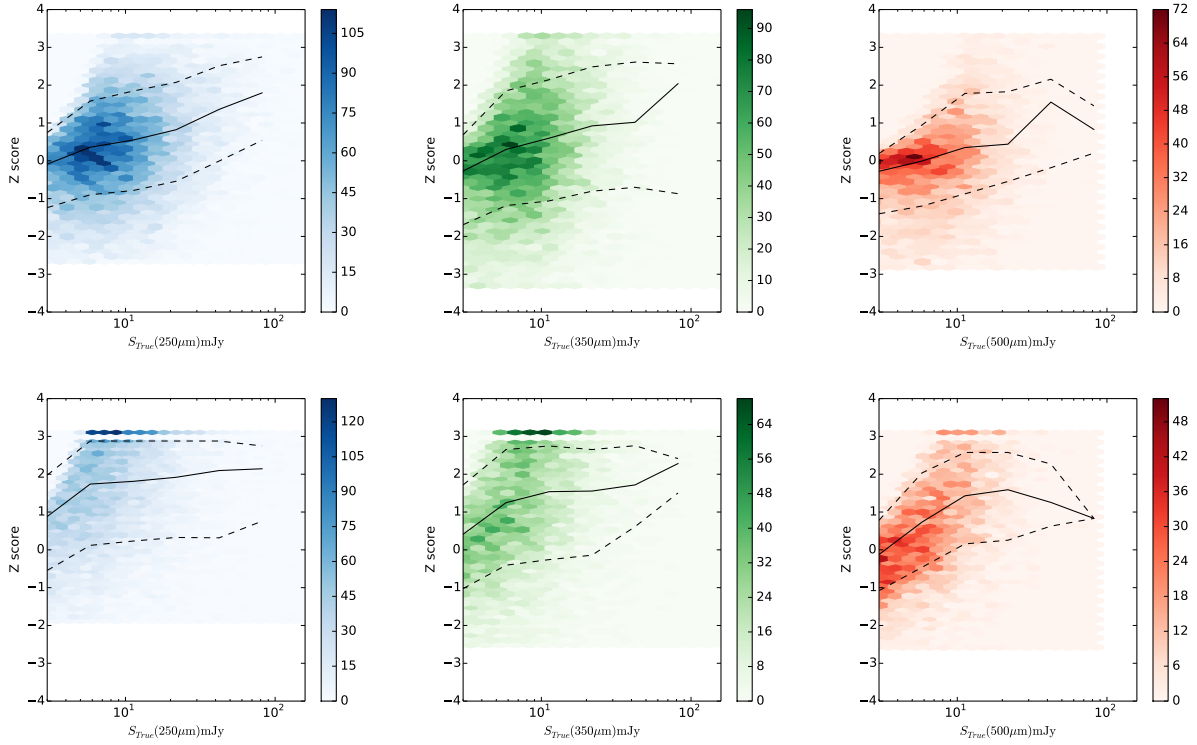


Figure 4. Z score, or flux density error for *XID+* (top) and *DESPHOT* (bottom) as a function of true flux, for the 250 (blue), 350 (green) and 500 (red) μm SPIRE bands. Black solid and dashed lines shows median and standard deviation respectively. Density of objects is illustrated via underlying colour.

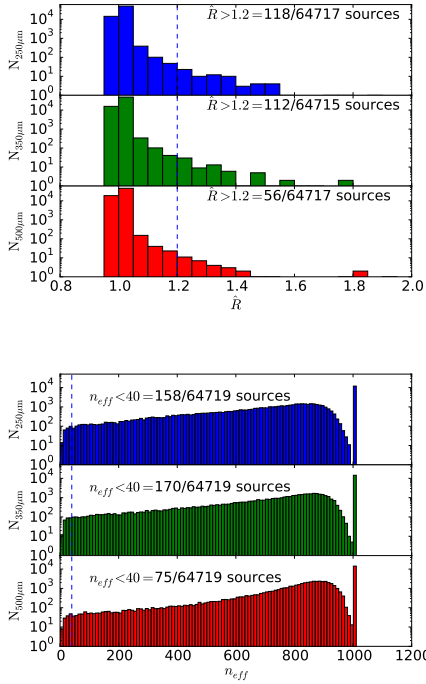


Figure 5. \hat{R} and \hat{n}_{eff}

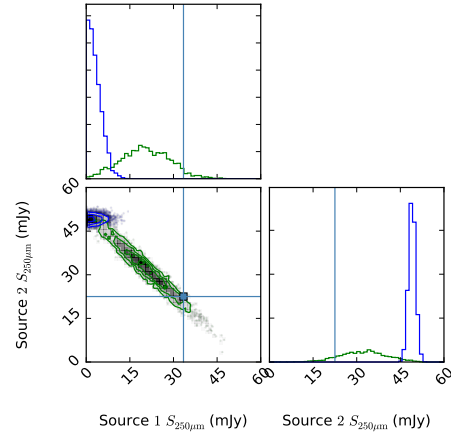


Figure 6. Posterior triangle plot of two correlated sources that are $2''$ apart, with true fluxes indicated by the vertical and horizontal line. *DESPHOT* (blue) assigns all the flux to one source, whereas as with *XID+* we get the full uncertainty information from the posterior. Plot created with [Foreman-Mackey et al. \(2014\)](#).

D., 2013, Bayesian Data Analysis, Third Edition, Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis
 Griffin M. J., et al., 2010, Astron. Astrophys., 518, L3
 Hoffman M. D., Gelman A., 2014, Journal of Machine Learning Research, 15, 1593
 Holland W. S., et al., 2013, MNRAS, 430, 2513
 Holland W. S., et al., 1999, MNRAS, 303, 659

- Iverson R. J., et al., 2010, *A&A*, 518, L31
 Kessler M. F., et al., 1996, *Astron. Astrophys.*, 315, L27
 Marsden G., et al., 2009, *ApJ*, 707, 1729
 Murakami H., et al., 2007, *PASJ*, 59, 369
 Neugebauer G., et al., 1984, *Astrophysical Journal*, 278, L1
 Nguyen H. T., et al., 2010, *A&A*, 518, L5
 Oliver S. J., et al., 2012, *Monthly Notices of the Royal Astronomical Society*, 424, 1614
 Pascale E., et al., 2009, *ApJ*, 707, 1740
 Pilbratt G. L., et al., 2010, *Astronomy and Astrophysics*, 518, L1
 Poglitsch A., et al., 2010, *Astron. Astrophys.*, 518, L2
 Puget J.-L., Abergel A., Bernard J.-P., Boulanger F., Burton W. B., Desert F.-X., Hartmann D., 1996, *Astron. Astrophys.*, 308, L5
 Roseboom I. G., et al., 2011, *Monthly Notices of the Royal Astronomical Society*, 419, 2758
 Roseboom I. G., et al., 2010, *arXiv*
 Safarzadeh M., Ferguson H. C., Lu Y., Inami H., Somerville R. S., 2015, *Astrophys. J.*, 798, 91
 Silva L., Granato G. L., Bressan A., Danese L., 1998, *Astrophys. J.*, 509, 103
 Soifer B. T., et al., 1984, *Astrophysical Journal*, 278, L71
 Team S. D., 2015a, *Pystan: the python interface to stan*, version 2.7.0
 Team S. D., 2015b, *Stan: A c++ library for probability and sampling*, version 2.7.0
 ter Braak C. J., Boer M. P., Totir L. R., Winkler C. R., Smith O. S., Bink M. C., 2010, *Genetics*, 185, 1045
 Tibshirani R., 1996, *Journal of the Royal Statistical Society (Series B)*, 58, 267
 Wang L., et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 2870
 Werner M. W., et al., 2004, *The Astrophysical Journal Supplement Series*, 154, 1
 Wright E. L., et al., 2010, *AJ*, 140, 1868
 Zou H., 2006, *Journal of the American Statistical Association*, 101, 1418

APPENDIX A


```
//Full Bayesian inference fit XID
```

```
data {
  int<lower=0> nsrc; //number of sources
  //-----PSW-----
  int<lower=0> npix_psw; //number of pixels
  int<lower=0> nnz_psw; //number of non neg entries in A
  vector[npix_psw] db_psw; //flattened map
  vector[npix_psw] sigma_psw; //flattened uncertainty map (assuming no covariance between pixels)
  real bkg_prior_psw; //prior estimate of background
  real bkg_prior_sig_psw; //sigma of prior estimate of background
  vector[nnz_psw] Val_psw; //non neg values in image matrix
  int Row_psw[nnz_psw]; //Rows of non neg values in image matrix
  int Col_psw[nnz_psw]; //Cols of non neg values in image matrix
  //-----PMW-----
  int<lower=0> npix_pmw; //number of pixels
  int<lower=0> nnz_pmw; //number of non neg entries in A
  vector[npix_pmw] db_pmw; //flattened map
  vector[npix_pmw] sigma_pmw; //flattened uncertainty map (assuming no covariance between pixels)
  real bkg_prior_pmw; //prior estimate of background
  real bkg_prior_sig_pmw; //sigma of prior estimate of background
  vector[nnz_pmw] Val_pmw; //non neg values in image matrix
  int Row_pmw[nnz_pmw]; //Rows of non neg values in image matrix
  int Col_pmw[nnz_pmw]; //Cols of non neg values in image matrix
  //-----PLW-----
  int<lower=0> npix_plw; //number of pixels
  int<lower=0> nnz_plw; //number of non neg entries in A
  vector[npix_plw] db_plw; //flattened map
  vector[npix_plw] sigma_plw; //flattened uncertainty map (assuming no covariance between pixels)
  real bkg_prior_plw; //prior estimate of background
  real bkg_prior_sig_plw; //sigma of prior estimate of background
  vector[nnz_plw] Val_plw; //non neg values in image matrix
  int Row_plw[nnz_plw]; //Rows of non neg values in image matrix
  int Col_plw[nnz_plw]; //Cols of non neg values in image matrix
}
parameters {
  vector<lower=-2.0,upper=3.0> [nsrc] src_f_psw; //source vector
  real bkg_psw; //background
  vector<lower=-2.0,upper=3.0> [nsrc] src_f_pmw; //source vector
  real bkg_pmw; //background
  vector<lower=-2.0,upper=3.0> [nsrc] src_f_plw; //source vector
  real bkg_plw; //background
}

model {
  vector[npix_psw] db_hat_psw; //model of map
  vector[npix_pmw] db_hat_pmw; //model of map
  vector[npix_plw] db_hat_plw; //model of map

  vector[nsrc+1] f_vec_psw; //vector of source fluxes and background
  vector[nsrc+1] f_vec_pmw; //vector of source fluxes and background
  vector[nsrc+1] f_vec_plw; //vector of source fluxes and background

  //Prior on background
  bkg_psw ~ normal(bkg_prior_psw, bkg_prior_sig_psw);
  bkg_pmw ~ normal(bkg_prior_pmw, bkg_prior_sig_pmw);
  bkg_plw ~ normal(bkg_prior_plw, bkg_prior_sig_plw);
```

```

//Prior on flux of sources (not being used yet)
//src_f_psw ~normal(-1,2.2);
//src_f_pmw ~normal(-1,2.2);
//src_f_plw ~normal(-1,2.2);

//background is now contribution from confusion only!!
f_vec_psw[nsrc+1] <-bkg_psw;
f_vec_pmw[nsrc+1] <-bkg_pmw;
f_vec_plw[nsrc+1] <-bkg_plw;

// Transform to normal space. As I am sampling variable then transforming I don't need a Jacobian
for (n in 1:nsrc) {
  f_vec_psw[n] <- pow(10.0,src_f_psw[n]);
  f_vec_pmw[n] <- pow(10.0,src_f_pmw[n]);
  f_vec_plw[n] <- pow(10.0,src_f_plw[n]);
}

// Create model maps (i.e. db_hat = A*f) using sparse multiplication
for (k in 1:npix_psw) {
  db_hat_psw[k] <- 0;
}
for (k in 1:nnz_psw) {
  db_hat_psw[Row_psw[k]+1] <- db_hat_psw[Row_psw[k]+1] + Val_psw[k]*f_vec_psw[Col_psw[k]+1];
}

for (k in 1:npix_pmw) {
  db_hat_pmw[k] <- 0;
}
for (k in 1:nnz_pmw) {
  db_hat_pmw[Row_pmw[k]+1] <- db_hat_pmw[Row_pmw[k]+1] + Val_pmw[k]*f_vec_pmw[Col_pmw[k]+1];
}

for (k in 1:npix_plw) {
  db_hat_plw[k] <- 0;
}
for (k in 1:nnz_plw) {
  db_hat_plw[Row_plw[k]+1] <- db_hat_plw[Row_plw[k]+1] + Val_plw[k]*f_vec_plw[Col_plw[k]+1];
}

// likelihood of observed map|model map
db_psw ~ normal(db_hat_psw, sigma_psw);
db_pmw ~ normal(db_hat_pmw, sigma_pmw);
db_plw ~ normal(db_hat_plw, sigma_plw);

// As actual maps are mean subtracted, requires a Jacobian adjustment
//db_psw <- db_obs_psw - mean(db_obs_psw)
//increment_log_prob(log((size(db_obs_psw)-1)/size(db_obs_psw)))
//db_pmw <- db_obs_pmw - mean(db_obs_pmw)
//increment_log_prob(log((size(db_obs_pmw)-1)/size(db_obs_pmw)))
//db_plw <- db_obs_plw - mean(db_obs_plw)
//increment_log_prob(log((size(db_obs_plw)-1)/size(db_obs_plw)))
}

```