

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

P.D. Hurley,<sup>1\*</sup> S. Oliver,<sup>1</sup> J. Scudder,<sup>1</sup> M. Griffin,<sup>2</sup> S. Duivenvoorden,<sup>1</sup> M. Sargent<sup>1</sup>

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

<sup>2</sup>*Cardiff School of Physics and Astronomy, Cardiff University, Queens Buildings, The Parade, Cardiff CF24 3AA*

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 uncertainty accuracy, 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 run XID+ on the COSMOS SPIRE maps from HerMES, using a 24  $\mu\text{m}$  catalogue as a prior and show the marginalised SPIRE colour-colour plot and marginalised contribution to the cosmic infrared background at the SPIRE wavelengths. For the first time, we provide the posterior probability samples for the COSMOS field as a data product, alongside the more standard catalogue. We also discuss how additional work within HELP on providing prior information on fluxes can and will be utilised by XID+. Software available at [https://github.com/pdh21/XID\\_plus/](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. 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 infrared atmospheric transmission windows to observe at the tail of the cosmic infrared background, only space-borne 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  $\mu\text{m}$  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-borne 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 advance 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) 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, due to the large beam size of

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

the SPIRE, and the galaxy density ( $\approx 30$  per SPIRE beam for optical sources with  $B < 28$ ), multiple galaxies can be located within the SPIRE beam. This is referred to as the problem of source confusion,

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 assign the flux in the SPIRE maps to the underlying astronomical objects. For example, if we know the location of a galaxy to a reasonable tolerance (e.g. from an optical image where resolution is better), we may 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  $24\ \mu\text{m}$  and  $1.4\ \text{GHz}$ , 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  $24\text{-}\mu\text{m}$  and  $1.4\text{-GHz}$  populations and those observed at far-IR wavelengths;  $>80$  per cent of the cosmic IR background at SPIRE wavelengths can be accounted for by  $24\text{-}\mu\text{m}$  sources with  $S_{24} > 25\ \mu\text{Jy}$  (e.g. Marsden et al. 2009; Pascale et al. 2009; Elbaz et al. 2010; Béthermin et al. 2012), 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, which suffers from two major issues. The first is that variance and co-variance of source fluxes can not be properly estimated. The second is the inability to discriminate between real and spurious sources, which can result in overfitting. The list driven algorithm developed for HerMES (DESPHOT Roseboom et al. 2011; Wang et al. 2014) tried to overcome this by using the non-negative weighted LASSO algorithm (Tibshirani 1996; Zou 2006; ter Braak et al. 2010), a shrinkage and selection method which introduces an additional penalty term in an attempt to reduce the number of sources needed to fit the map. Although successful in reducing the number of sources, when multiple sources are located close-by (i.e. within the SPIRE beam), the method has been found to wrongly assign all the flux to one source.

The solution to both of these problems is to fully explore the posterior probability distribution with Bayesian inference techniques such as Markov Chain Monte Carlo (MCMC) methods. By fully exploring the posterior, the variance and covariance between sources can be properly estimated. Also, by considering the covariance between sources (i.e. how the flux of sources affect each other), the probability of sources being spurious or real is taken into account, removing the need for methods such as LASSO.

Up until the present day, use of MCMC techniques has been computationally unfeasible. However, advances in computational technology and algorithms such as Hamiltonian Monte Carlo now make this sort of approach a viable alternative, as demonstrated by Safarzadeh et al. (2015), who used an MCMC based approach to fit PACS simulated maps.

As part of the *Herschel Extragalactic Legacy Project* (HELP; Oliver in prep.), 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

way in which to introduce additional prior information. By using the Bayesian inference tool, (*StanStan Development Team 2015a,b*) to sample the 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 LASSO fitting approach used by DESPHOT. In this paper, we show that XID+ outperforms DESPHOT when using just positional information. In Section 2 we discuss the algorithm, and show how the software performs on simulated SPIRE maps in Section 3. In Section 4 we apply XID+ on the HerMES COSMOS SPIRE maps, using a  $24\ \mu\text{m}$  catalogue as a prior and show the resulting marginalised SPIRE colour-colour plot and contribution to the cosmic infrared background. We discuss how XID+ can make use of flux prior information, delivered by the HELP project (Hurley et al. in prep) in Section 5 and make final conclusions in 6.

## 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.

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 lossy 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 full 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 maps are formed from  $S$  known point sources, with 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 ( $B$ ). Our map can therefore be described as follows:

$$\mathbf{d} = \sum_{i=1}^S \mathbf{P}_i \mathbf{f}_i + B \quad (1)$$

where  $\mathbf{d}$  is our model of the map,  $\mathbf{P}_i$  is the PRF for source  $i$ ,  $f_i$  is the flux density for source  $i$ ,  $B$  is a global background estimate.

We can rewrite the above equation in the linear form:

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

Where  $\mathbf{A}$  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

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

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

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

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).

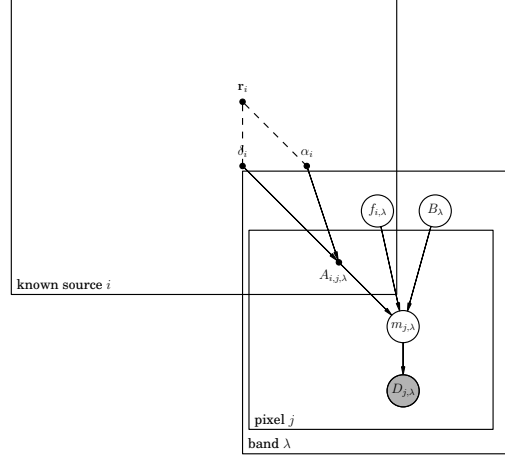
LASSO is a shrinkage and selection method for linear regression and works by treating sources either as ‘inactive’ with flux density set to zero, or ‘active’. It switches sources on one at a time, with the order determined by reduction in chi-squared gained by turning them on. The process continues until 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 complete information on 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 a plate diagram (Bishop 2006) for our PGM of the basic XID+ model, where boxes indicate repeated values such as source ( $i$ ), pixel ( $j$ ) and band ( $\lambda$ ). Open circles correspond to random variables and dots are deterministic (or fixed) variables. For our simplest model, the sky co-ordinates ( $\alpha_i$  and  $\delta_i$ ) of our sources are treated as deterministic (i.e. known). The PRF is assumed to be a Gaussian, with full-width half-maximum (FWHM) of 18.15, 25.15 and 36.3 arcsec for 250, 350 and 500  $\mu\text{m}$  respectively (Griffin et al. 2010). Both these deterministic variables are used to make the pointing matrix  $A_{i,j,\lambda}$  which gives the contribution source  $i$  makes to each pixel  $j$  in the map at wavelength  $\lambda$ . Each source has its own flux  $f_{i,\lambda}$  which is a random variable. By multiplying  $f$ ,  $A$  for all sources and pixels, and adding our global estimate for the background  $B$ , we can make our model for the map,  $m$ .



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

### 2.1.1 Stan

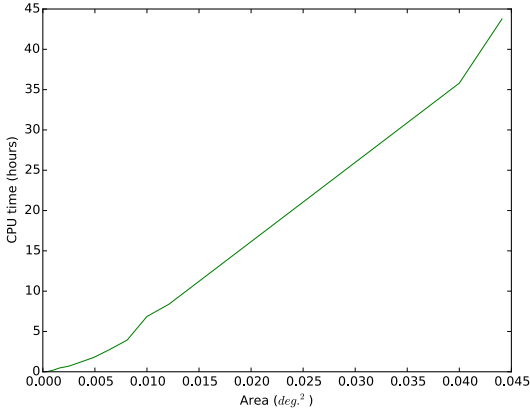
Now that we have our probabilistic model, we need to sample from it to obtain the posterior. We use the Bayesian inference tool, *Stan*, which is ‘a probabilistic programming language implementing full Bayesian statistical inference with MCMC sampling’. *Stan* uses the adaptive Hamiltonian Monte Carlo (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 initial positions in parameter space. 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-chain ( $B$ ) 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



**Figure 2.** CPU time as a function of tile area (in degrees) for XID+.

$\hat{R}$ , 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 10 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 unfeasible. DESPHOT segmented the map by locating islands of high SNR pixels enclosed by low SNR pixels.

We adopt a simpler tiling scheme that splits the map data into square tiles, with a user defined tile length,  $t_l$ . The tiles overlap with each other by 25% in both Right Ascension and Declination. This makes sure no source is less than 12.5% of  $t_l$  away from its optimum tile. So that we consider all sources that could foreseeably contribute to the map data in a tile, we fit to sources that are within 0.01 degrees (i.e. FWHM of SPIRE 500  $\mu$ m beam) of the tile boundary. For our final combined posterior, for each source, we take the posterior from the optimum tile, i.e. the tile in which the source lies closest to the centre.

The choice of  $t_l$  effects the computational time of XID+. The required CPU time is found to scale linearly with the amount of data (i.e. number of pixels). Figure 2 shows the CPU time as a function of tile area.

## 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 7).

## 3 SIMULATIONS

In order to test and quantify the performance of XID+, we use simulated SPIRE maps of the COSMOS field, a good example of a deep map i.e. where confusion noise ( $\sigma_{conf.}$ ) is  $\gg$  than instrumental noise ( $\sigma_{inst.}$ ). 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  $\lambda_{rest} > 70 \mu$ m.

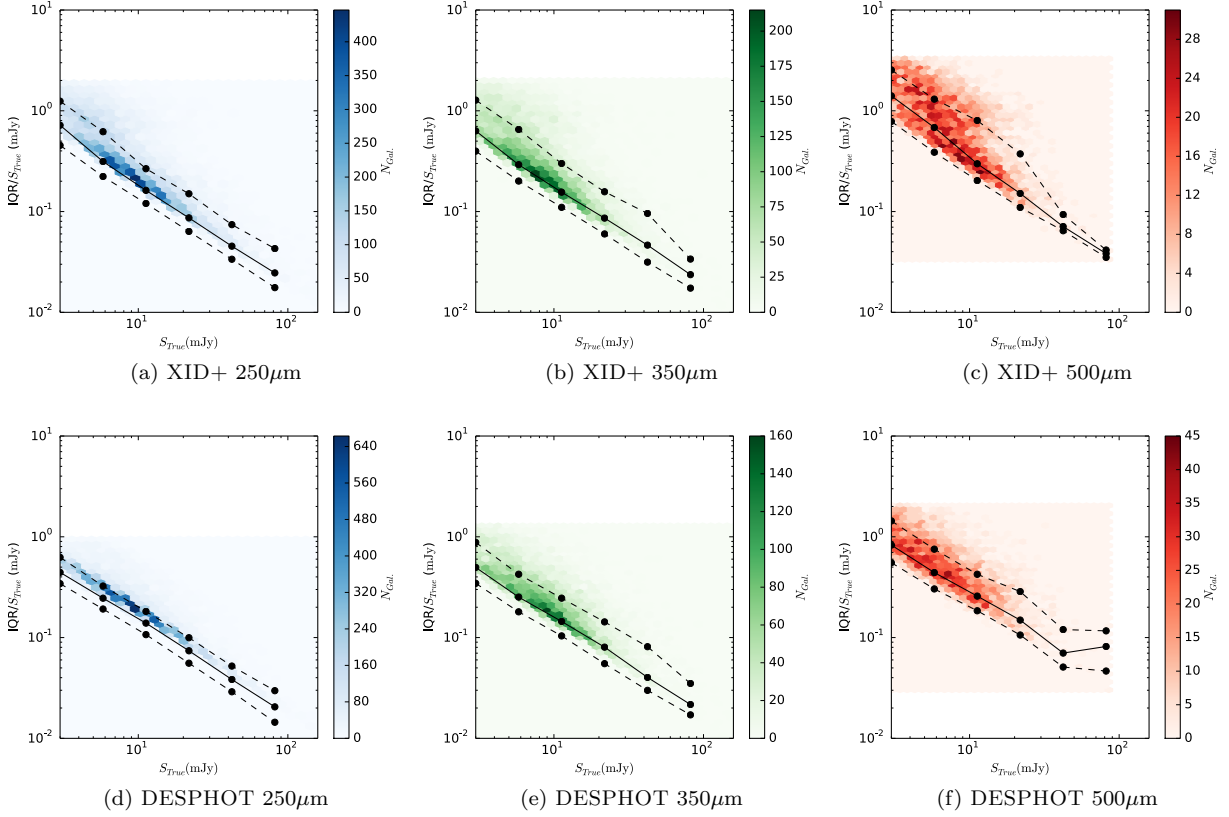
A mock 100  $\mu$ m input catalogue, similar to what would be expected of a PACS catalogue, is generated by taking the mock catalogue and making a cut at a flux limit of 50  $\mu$ Jy, giving a total of 64,719 sources over 3.4 square degrees. We use this as our prior input catalogue for both XID+ and DESPHOT. In order to compare performance, we look at three measures: precision, flux accuracy, and flux uncertainty accuracy. For XID+, we only consider sources whose output median flux is above 1 mJy. Likewise, with DESPHOT, we only consider sources which have a maximum likelihood flux of greater than 1 mJy.

### 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 the spread of the sample and so we use the interquartile range (IQR) as our measure of precision. Figure 3 shows the 16th, 50th and 84th percentile (i.e. median and median  $\pm \sigma$ ) of the IQR for 6 bins in true flux. IQR is normalised 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  $\mu$ m outputs achieve a similar level of precision, the outputs for 500  $\mu$ m do not reach the same level of precision. In comparison to DESPHOT, XID+ is marginally less precise 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 the median flux estimate from our posterior and the true flux from the simulation, normalised by true flux, as our estimate of flux accuracy. Figure 4 shows how flux accuracy changes as a function of input flux for all three bands. As before, we show the 16th, 50th, and 84th percentile for 6 bins in true



**Figure 3.** 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)  $\mu\text{m}$  SPIRE bands. Black solid and dashed lines shows median and standard deviation respectively. Density of objects is illustrated via underlying colour.

flux. For 250  $\mu\text{m}$ , XID+ there is no discernible offset from the truth by  $\approx 5\text{mJy}$ , whereas DESPHOT underestimates the flux for all but the very brightest sources. For 350  $\mu\text{m}$  and 500  $\mu\text{m}$ , the offset from the truth reaches zero by  $\approx 10\text{mJy}$ , whereas DESPHOT continues to underestimate for all but the very brightest sources.

### 3.3 Flux Uncertainty accuracy

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 5 shows the flux uncertainty accuracy (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 do not have to make an assumption on the shape of the uncertainty distribution. If we assume the posterior has good frequentist coverage, then we can calculate uncertainty accuracy by taking the percentile at which the true flux value falls within the posterior, and convert the percentile to a corresponding sigma level.

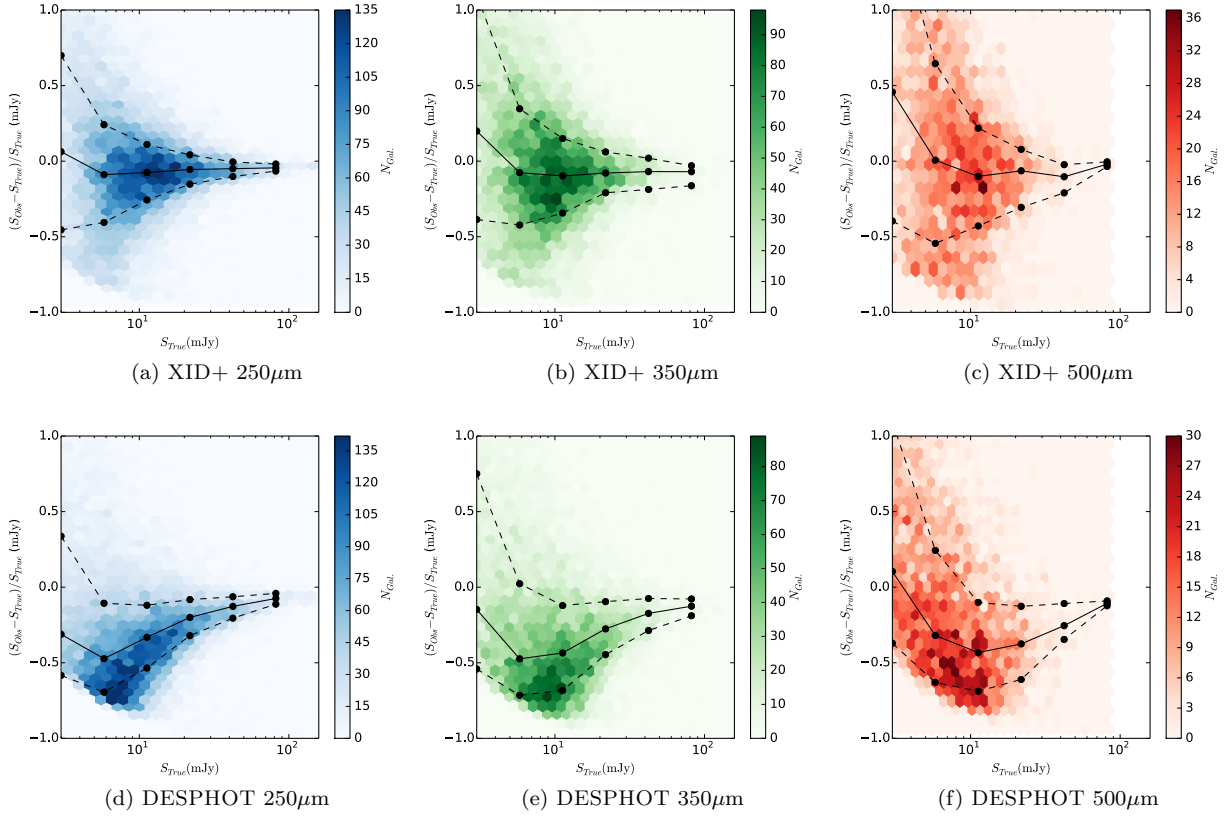
For the 250  $\mu\text{m}$  band, and sources  $\lesssim 25\text{mJy}$ , XID+ produces a Z score distribution that matches that expected if uncertainties are correctly estimated (i.e. distribution is centred around zero, with width  $\approx 1$ ). Above 25mJy, the median Z score increases, indicating that flux and uncertainty are being under estimated. In comparison, for all fluxes, the

uncertainty distribution from DESPHOT are above 1 and increases to over 2 for the brighter sources. There are also a large number of sources with a Z score greater than 3 (as seen by the higher density in bins at a Z score of 3). This indicates that DESPHOT is a poor estimator with the majority of sources in DESPHOT lying more than 1  $\sigma$  away from their true flux. The flux uncertainty accuracies for 350  $\mu\text{m}$  and 500  $\mu\text{m}$  show a similar behaviour, though not as severe.

### 3.4 Convergence

As described in Section 2.1.2, we provide  $\hat{R}$  as an estimate of convergence and  $n_{eff}$  as a measure of independence within the sample. Figures 6 show the histogram for  $\hat{R}$  and  $n_{eff}$  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  $n_{eff}$  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.





**Figure 4.** Accuracy of XID+ (top) and DESPHOT (bottom) as a function of true flux, for the 250 (blue), 350 (green) and 500 (red)  $\mu\text{m}$  SPIRE bands. Black solid and dashed lines shows median and standard deviation respectively. Density of objects is illustrated via underlying colour.

### 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  $\mu\text{m}$ , 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 PRF<sup>1</sup>. 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 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 that present in the maps. It is also lower than that achieved by DESPHOT. This could be due to the fact we fit the

	250 $\mu\text{m}$		350 $\mu\text{m}$		500 $\mu\text{m}$	
	$\sigma_{tot.}$	$\sigma_{conf.}$	$\sigma_{tot.}$	$\sigma_{conf.}$	$\sigma_{tot.}$	$\sigma_{conf.}$
XID+	1.9	1.3	2.2	1.9	2.6	1.9
DESPHOT	2.4	1.9	3.5	3.3	3.4	2.9

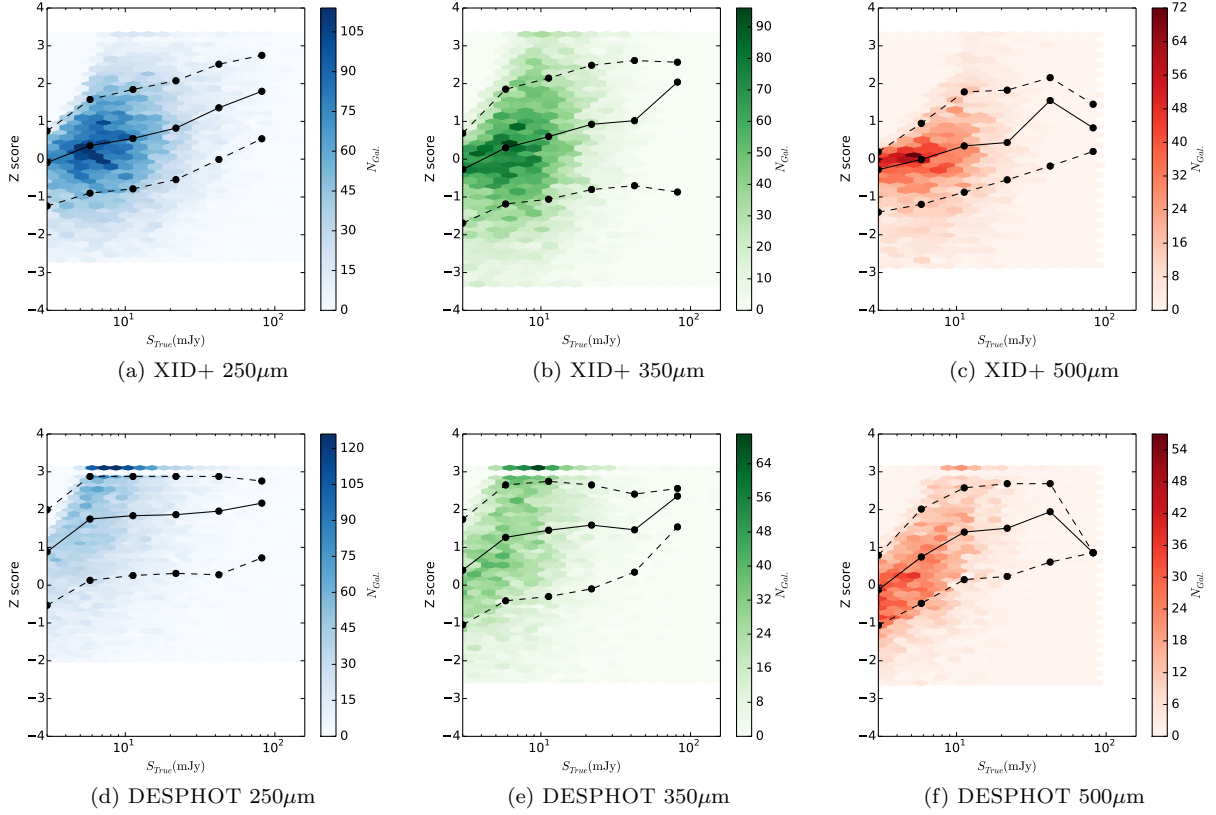
**Table 1.** Table showing the remaining total and confusion noise for the three SPIRE bands (in mJy), from running XID+ and DESPHOT on the Lacey simulated COSMOS maps.

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.

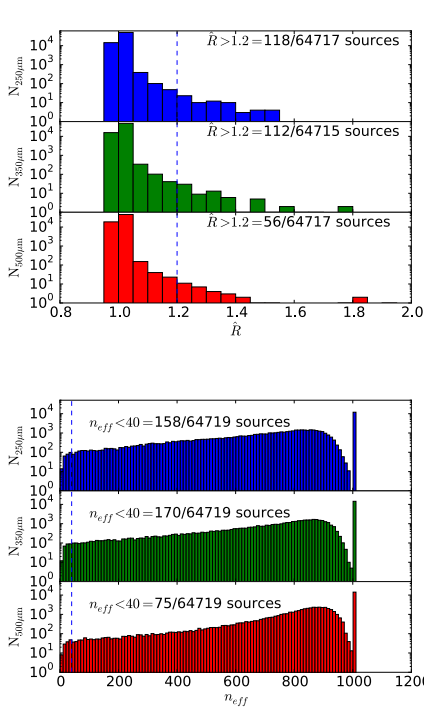
### 3.6 Correlated Sources

For sources that are close together (i.e. within FWHM of the PRF), 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 when comparing flux estimates with DESPHOT, which, by using the LASSO algorithm forces one source to have all the flux and the other nearby source to zero. Figure 7 shows an example of two sources which are 2'' apart. The 250  $\mu\text{m}$  flux estimate from

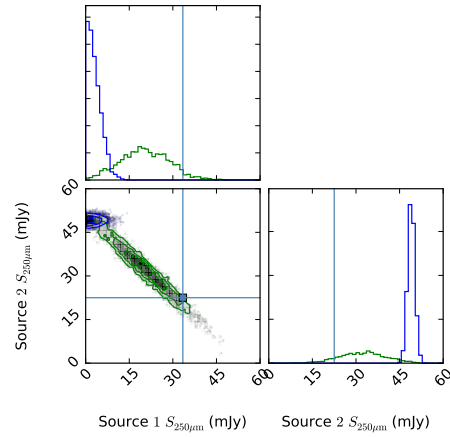
<sup>1</sup> The best way to find point sources in a noisy map is to convolve with the PRF, as described in [Chapin et al. \(2011\)](#)



**Figure 5.** 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)  $\mu\text{m}$  SPIRE bands. Black solid and dashed lines shows median and standard deviation respectively. Density of objects is illustrated via underlying colour.

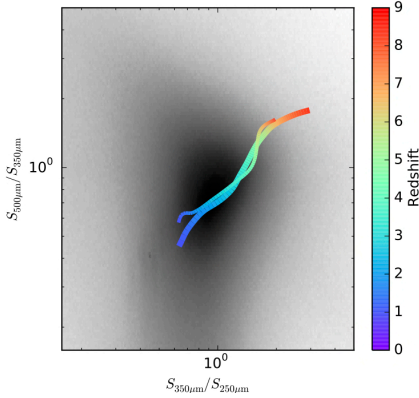


**Figure 6.**  $\hat{R}$  and  $n_{\text{eff}}$



**Figure 7.** 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. \(2014\)](#).

both XID+ and DESPHOT are shown in green and blue respectively. 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.



**Figure 8.** The marginalised SPIRE colour-colour probability density (in black) from the MIPS 24  $\mu\text{m}$  sources. Over plotted are the redshifted spectral energy distributions for Arp220 (thick line) and M82 (thin line) (Polletta et al. 2007).

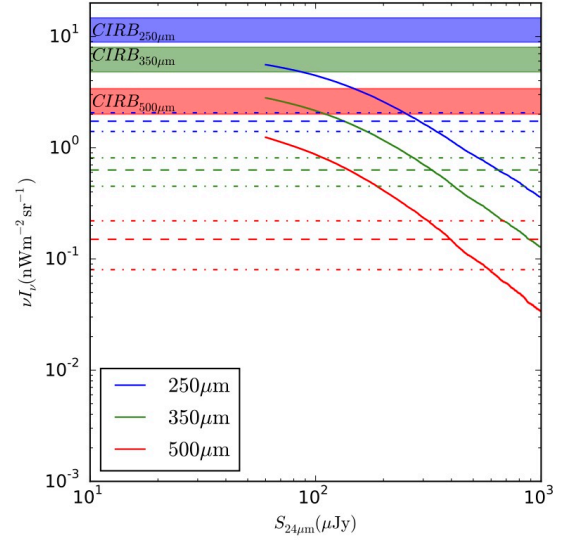
#### 4 COSMOS FIELD

Having satisfactorily demonstrated the performance on simulations, we have run XID+ on the HerMES COSMOS SPIRE maps from the 2nd Data release. As a prior, we take the MIPS 24  $\mu\text{m}$  catalogue (Floc'h et al. 2009), which covers an area of 2.265 square degrees and includes 52,092 sources with a 24  $\mu\text{m}$  flux density above 60  $\mu\text{Jy}$ . Figure 8 shows the marginalised probability density of our MIPS 24  $\mu\text{m}$  prior catalogue in SPIRE colour-colour space and is constructed by combining the 1500 samples from the posterior, for all 52,092 sources. The redshift tracks for Arp220 and M82 are over-plotted and run through the highest density region at redshifts of around 2-3. Figure 9 shows the cumulative contribution to the cosmic infrared background (CIRB) at 250, 350 and 500  $\mu\text{m}$  from our MIPS 24  $\mu\text{m}$  prior catalogue, as a function of 24  $\mu\text{m}$  flux. The filled bands illustrate the FIRAS absolute measurements from Lagache & Puget (2000), while dashed and dashed dotted lines indicate the contribution to the CIRB from SPIRE resolved sources (Oliver et al. 2010). By going to a depth of 60  $\mu\text{Jy}$  in the 24  $\mu\text{m}$  catalogue, we reach  $5.573 \pm 0.003$ ,  $2.805 \pm 0.002$  and  $1.24 \pm 0.002$   $\text{nWm}^{-2}\text{sr}^{-1}$  or 47, 44 and 46% of the nominal measured values at 250, 350 and 500  $\mu\text{m}$  (Lagache & Puget 2000). As a comparison, at the  $(40 \text{ beams})^{-1}$  depth, (Oliver et al. 2010) resolve  $1.73 \pm 0.33$ ,  $0.63 \pm 0.18$ ,  $0.15 \pm 0.07$  or 15, 10, 6%.

Our final data product consists of a catalogue, summarising the SPIRE fluxes via the 16th, 50th and 84th percentile (i.e. median and median  $\pm \sigma$ ), the median background and the convergence statistics. We also make available the 3000 samples from the posterior probability distribution, each of which can be thought of as a possible catalogue in probability space.

#### 5 DISCUSSION

We have run XID+ on simulated maps of the COSMOS field and compared it to DESPHOT using three main metrics; flux accuracy, precision and uncertainty accuracy. On accuracy, XID+ performs significantly better in all three bands, and although appearing marginally less precise, the loss of



**Figure 9.** The cumulative contribution to the cosmic infrared background (CIRB) from our MIPS 24  $\mu\text{m}$  sources, as a function of 24  $\mu\text{m}$  flux, at 250 (blue), 350 (green) and 500 (red)  $\mu\text{m}$ . The one  $\sigma$  confidence regions for the CIRB at SPIRE wavelengths, as measured by Lagache & Puget (2000) are plotted with blue, green and red bands. The mean and one  $\sigma$  uncertainty for the CIRB resolved by SPIRE (Oliver et al. 2010) are shown with dashed and dashed-dotted lines.

precision relates to far better estimates for flux uncertainties.

The higher performance gained by XID+ comes from fully capturing the posterior probability distribution on flux estimates. By exploring the posterior, we get a proper handle on uncertainties and no longer have to employ penalisation techniques such as LASSO, which are known to behave erratically.

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. However, as can be seen in Figure 7, 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).

More generally, the probabilistic model used in XID+ can easily be expanded, allowing distributions such as the flux distribution (or number counts) to be modelled explicitly. In principle, the probabilistic model could become detailed enough to simultaneously fit luminosity functions and the main sequence distributions and we will explore these expansions in future papers.



## 6 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 distribution, 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 three main metrics; flux accuracy, precision and uncertainty accuracy.

We have run XID+ on the HerMES COSMOS SPIRE maps from the 2nd Data release, using the MIPS 24  $\mu\text{m}$  catalogue (Floc'h et al. 2009) as our prior. Using the full posterior, we have created a marginalised SPIRE colour-colour plot, illustrating the probability distribution of our MIPS 24  $\mu\text{m}$  catalogue in SPIRE colour-colour space. We have also shown that the MIPS 24  $\mu\text{m}$  sources contribute 47, 44 and 46% to the cosmic infrared background at 250, 350 and 500  $\mu\text{m}$ . We provide the catalogue and posterior probability distribution samples as a data product at .... As far as we are aware, this is the first time the full posterior probability distribution is made available as a data product for list driven photometry.

## ACKNOWLEDGEMENTS

## REFERENCES

- B  thermin M., et al., 2012, *Astron. Astrophys.*, 542, A58  
 Bishop C. M., 2006, *Pattern Recognition and Machine Learning*. Springer  
 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  
 Floc'h E. L., et al., 2009, *The Astrophysical Journal*, 703, 222  
 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 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  
 Ivison R. J., et al., 2010, *A&A*, 518, L31  
 Kessler M. F., et al., 1996, *Astron. Astrophys.*, 315, L27  
 Lagache G., Puget J. L., 2000, *Astron. Astrophys.*, 355, 17  
 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., in prep., *Herschel extragalactic legacy project*  
 Oliver S. J., et al., 2012, *Monthly Notices of the Royal Astronomical Society*, 424, 1614  
 Oliver S. J., et al., 2010, *Astron. Astrophys.*, 518, L21  
 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  
 Polletta M., et al., 2007, *Astrophys. J.*, 663, 81  
 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  
 Stan Development Team, 2015a, *Pystan: the python interface to stan*, version 2.7.0  
 Stan Development Team, 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);
}

```