Mon. Not. R. Astron. Soc. **000**, 1–14 (2010) Printed 12 July 2011 (MN LATEX style file v2.2)

Herschel–ATLAS: counterparts from the UV–NIR in the science demonstration phase catalogue⋆

D.J.B. Smith[1,2]† L. Dunne[1], S.J. Maddox[1], S. Eales[3], D.G. Bonfield[2], M.J. Jarvis[2], W. Sutherland[4], S. Fleuren[4], E.E. Rigby[1], M.A. Thompson[2], I.K. Baldry[5], S. Bamford[1], S. Buttiglione[6], A. Cava[28], D.L. Clements[8], A. Cooray[9], S. Croom[10], A. Dariush[3], G. de Zotti[6,11], S.P. Driver[12,29], J.S. Dunlop[13], J. Fritz[14], D.T. Hill[12], A. Hopkins[15], R. Hopwood[16], E. Ibar[17], R.J. Ivison[17,13], D.H. Jones[15], L. Kelvin[12], L. Leeuw[18], J. Liske[19], J. Loveday[20], B.F. Madore[21], P. Norberg[13], P. Panuzzo[22], E. Pascale[3], M. Pohlen[3], C.C. Popescu[23], M. Prescott[5], A. Robotham[12], G. Rodighiero[6], D. Scott[24], M. Seibert[21], R. Sharp[15], P. Temi[25], R.J. Tuffs[26], P. van der Werf[27,13], E. van Kampen[19]

[1]*Centre for Astronomy and Particle Theory, The School of Physics & Astronomy, Nottingham University, University Park Campus, Nottingham, NG7 1HR, UK*
[2]*Centre for Astrophysics Research, Science & Technology Research Institute, University of Hertfordshire, Hatfield, Herts, AL10 9AB, UK*
[3]*School of Physics & Astronomy, Cardiff University, Queen Buildings, The Parade, Cardiff, CF24 3AA, UK*
[4]*School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London, E1 4NS, UK*
[5]*Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead, CH41 1LD, UK*
[6]*INAF – Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, I-35122, Padova, Italy*
[7]*Instituto de Astrofísica de Canarias (IAC) and Departamento de Astrofísica de La Laguna (ULL), La Laguna, Tenerife, Spain*
[8]*Astrophysics Group, Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*
[9]*University of California, Ivine, Department of Physics & Astronomy, 4186 Frederick Reines Hall, Irvine, CA 92697-4575, USA*
[10]*Sydney Institute for Astronomy, School of Physics, University of Sydney, NSW 2006, Australia*
[11]*SISSA, Via Bonomea 265, I-34136 Trieste, Italy*
[12]*SUPA, School of Physics and Astronomy, University of St. Andrews, North Haugh, St. Andrews, KY16 9SS, UK*
[13]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, UK*
[14]*Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, B-9000 Gent, Belgium*
[15]*Anglo-Australian Observatory, PO Box 296, Epping, NSW 1710, Australia*
[16]*Department of Physics and Astronomy, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK*
[17]*Uk Astronomy Technology Centre, Royal Observatory, Edinburgh, EH9 3HJ, UK*
[18]*SETI Institute, 515 N. Whisman Avenue, Mountain View, CA 94043, USA*
[19]*European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748, Garching bei München, Germany*
[20]*Astronomy Centre, University of Sussex, Falmer, Brighton, BN1 9QH, UK*
[21]*Observatories of the Carnegie Institution, 813 Santa Barbera St., Pasadena, CA 91101, USA*
[22]*CEA, Laboratoire AIM, Irfu/SAp, Orme des Merisiers, F-91191 Gif-sur-Yvette, France*
[23]*Jeremiah Horrocks Institute, University of Central Lancashire, Preston, PR1 2HE, UK*
[24]*Department of Physics and Astronomy, 6224 Agricultural Road, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada*
[25]*Astrophysics Branch, NASA Ames Research Center, Mail Stop 2456, Moffett Field, CA 94035, USA*
[26]*Max Planck Institut für Kernphysik (MPIK), Saupfercheckweg, 69117 Heidelberg, Germany*
[27]*Leiden Observatory, Leiden University, PO Box 9513, NL - 2300 RA Leiden, The Netherlands*
[28]*Departamento de Astrofísica, Facultad de CC. Físicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain*
[29]*International Centre for Radio Astronomy Research, The University of Western Australia, 7 Fairway, Crawley, Perth, Western Australia, WA6009*

12 July 2011

ABSTRACT

We present a technique to identify optical counterparts of 250 μm-selected sources from the *Herschel*-ATLAS survey. Of the 6621 250 μm > 32 mJy sources in our science demonstration catalogue we find that ~ 60 percent have counterparts brighter than $r = 22.4$ mag in the Sloan Digital Sky Survey. Applying a likelihood ratio technique we are able to identify 2423 of the counterparts with a reliability $R > 0.8$. This is approximately 37 percent of the full 250 μm catalogue. We have estimated photometric redshifts for each of these 2423 reliable counterparts, while 1099 also have spectroscopic redshifts collated from several different sources, including the GAMA survey. We estimate the completeness of identifying counterparts as a function of redshift, and present evidence that 250 μm-selected *Herschel*-ATLAS galaxies have a bimodal redshift distribution. Those with reliable optical identifications have a redshift distribution peaking at $z \approx 0.25 \pm 0.05$, while sub-mm colours suggest that a significant fraction with no counterpart above the r-band limit have $z > 1$. We also suggest a method for selecting populations of strongly-lensed high redshift galaxies. Our identifications are matched to UV–NIR photometry from the GAMA survey, and these data are available as part of the *Herschel*-ATLAS public data release.

Key words: Galaxies: Local, Galaxies: Infrared, Galaxies: Star-forming, Methods: Statistical, Submillimetre: Galaxies

1 INTRODUCTION

One of the key problems to overcome when conducting multi-wavelength surveys is determining which sources are associated with one another in different wave-bands, and which are unrelated. When multiple observations have been conducted at similar wavelengths and with similar resolution and sensitivity, this problem can be reliably addressed by using a simple nearest–neighbour match. However, in the situation where the two distinct sets of observations to be matched have considerably different resolution – for example matching far-infrared or sub-millimetre survey data to an optical catalogue (e.g. Sutherland et al., 1991, Clements et al., 1996, Serjeant et al., 2003, Clements et al., 2004, Ivison et al. 2005, 2007, Wang & Rowan-Robinson, 2009, Biggs et al. 2010) – the large positional uncertainties in the longer-wavelength data can make it much more difficult to find reliable associations between sub–millimetre sources and their optical/near–infrared counterparts.

One method which can be used to identify the most likely counterpart to a low-resolution source, is the Likelihood Ratio technique (hereafter LR), first suggested by Richter (1975), and expanded by Sutherland & Saunders (1992) and Ciliegi et al. (2003). The crucial advantage of the LR technique over other methods is that it not only uses the positional information contained within the two catalogues, but also includes brightness information (both of the individual potential counterparts, and of the higher resolution catalogue as a whole) to identify the most reliable counterpart to a low-resolution source.

The *Herschel* Astrophysical Terahertz Large Area Survey (*Herschel*–ATLAS, Eales et al., 2010) is the largest open-time key project that will be carried out with the *Herschel Space Observatory* (Pilbratt et al., 2010). The *Herschel*–ATLAS will survey in excess of 550 deg^2 in five channels centred on 100, 160, 250, 350 and 500 μm, using the PACS (Poglitsch et al., 2010) and SPIRE instruments (Griffin et al., 2010). This makes *Herschel*–ATLAS currently the largest area extragalactic *Herschel* survey. The *Herschel*–ATLAS observations consist of two scans in parallel mode reaching 5σ point source sensitivities of 132, 126, 32, 36 and 45 mJy in the 100 μm, 160 μm, 250 μm, 350 μm and 500 μm channels respectively, with beam sizes of approximately 9, 13, 18, 25 and 35 arcsec in the same five bands. The SPIRE and PACS map-making procedures are described in the papers by Pascale et al. (2010) and Ibar et al. (2010), while the catalogues are described in Rigby et al. (2010). One of the primary aims of the *Herschel*–ATLAS was to obtain the first unbiased survey of the local Universe at sub-mm wavelengths, and as a result the survey was designed to overlap with existing large optical and infrared surveys.

In this paper, we present a discussion of our implementation of the LR technique to identify the most reliable counterparts to 250 μm–selected sources in the *Herschel*–ATLAS science demonstration phase (SDP) data field (Eales et al., 2010). This field was chosen in order to take advantage of multi–wavelength data from the Sloan Digital Sky Survey (SDSS – York et al., 2000), and the UK Infrared Deep Sky Survey Large Area Survey (UKIDSS-LAS – Lawrence et al., 2007). This field also overlaps with the 9 hour field of the Galaxy And Mass Assembly survey (GAMA – Driver et al., 2010). The GAMA catalogue (Hill et al., 2011), comprises not only thousands of redshifts (for galaxies selected as described in Baldry et al., 2010, and observed with the maximum possible tiling efficiency – Robotham et al., 2010), but also r–band–defined aperture–matched photometry in the $ugrizYJHK$ bands. In addition, the GAMA fields are being systematically observed using the *Galaxy Evolution EXplorer (GALEX)* satellite (Martin et al., 2005) at Medium Imaging Survey depth to provide aperture–matched FUV and NUV counterparts to the catalogued GAMA sources (the *GALEX*–GAMA survey; Seibert et al., in prep). These counterparts will potentially be of great scientific value once the most reliable optical counterpart can be established for each *Herschel*–ATLAS source.

In section 2 we present the specific LR method that we have used to identify counterparts to 250 μm–selected sources from the *Herschel*–ATLAS SDP catalogue in an r–band catalogue of model magnitudes derived from the SDSS DR7. In section 3 we present the redshift properties of our catalogue, which covers ~ 16 deg^2 over the GAMA 9 hour field. Section 4 contains some basic results

⋆ *Herschel* is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA

† E-mail: daniel.j.b.smith@gmail.com

based on our reliable catalogue, and in section 5 we present some concluding remarks about the likelihood ratio technique and the resulting catalogue.

2 CALCULATING THE LIKELIHOOD RATIO

The likelihood ratio, i.e. the ratio between the probability that the source is the correct identification and the corresponding probability for an unrelated background source, is calculated as in Sutherland & Saunders (1992):

$$L = \frac{q(m)f(r)}{n(m)}, \quad (1)$$

in which $n(m)$, and $q(m)$ correspond to the SDSS r–band magnitude probability distributions of the full r–band catalogue and of the true counterparts to the sub–millimetre sources, respectively, while $f(r)$ represents the radial probability distribution of offsets between the 250 μm positions and the SDSS r–band centroids. We will now describe how we calculate each component of this relationship in turn.

2.1 Calculating the radial dependence of the likelihood ratio, $f(r)$

Here, $f(r)$ is the radial probability distribution function of the positional errors as a function of the separation from the SPIRE 250 μm position in arcseconds (r), given by:

$$f(r) = \frac{1}{2\pi\sigma_{\rm pos}^2}\exp\left(\frac{-r^2}{2\sigma_{\rm pos}^2}\right) \quad (2)$$

where r is the separation between the 250 μm and r–band positions, and $\sigma_{\rm pos}$ is the standard positional error (which is assumed to be isotropic).

For *Herschel*–ATLAS SDP observations, it was necessary to determine the SPIRE positional uncertainties. Since this information was not available *a priori*, we empirically estimated $\sigma_{\rm pos}$ using the SDSS DR7 r–band catalogue positions, assuming that the SDSS positional errors were negligible in comparison to the SPIRE errors. To determine $\sigma_{\rm pos}$, we derived histograms of the separations between the positions in the MAD-X SPIRE catalogue (Rigby et al., 2010) of the 5σ 250 μm sources, and all of those objects in the r–band SDSS DR7 catalogue within 50 arcsec, doing this for both the North–South and East–West directions (Figure 1). These histograms can be well–described as the sum of the Gaussian positional errors plus the clustering signal for SDSS sources convolved with Gaussian errors, $G(\theta,\sigma)$, with $\sigma = \sigma_{\rm pos}$:

$$n(x) = G'(x,\sigma_{\rm pos}) + \left(\sum_y w(\theta) * G(\theta,\sigma_{\rm pos})\right), \quad (3)$$

$$n(y) = G'(y,\sigma_{\rm pos}) + \left(\sum_x w(\theta) * G(\theta,\sigma_{\rm pos})\right), \quad (4)$$

where $w(\theta) = A\theta^{-\delta}$, with θ being measured in degrees for the purposes of comparison with the literature. We determined the values of A and δ empirically based solely on galaxies in the SDSS catalogue over > 35 deg^2 centred on the *Herschel*–ATLAS SDP field (limited to $r < 22.4$), with the best fit parameters $A = 6.89 \pm 0.90 \times 10^{-3}$ and $\delta = 0.689 \pm 0.069$, in reasonable agreement with the values of Connolly et al., 2002. The effects of clustering (i.e. $w(\theta) * G(\theta,\sigma_{\rm pos})$) are shown in the top panel of Figure 1.

In order to determine the 1σ positional error of the 250 μm selected catalogue, we conducted a simple χ^2 fit of our model (equations 3 & 4) to the histograms. The results are shown in Figure 1 for the summations in the East-West and North-South directions in the middle and bottom panels respectively. The clustering signal is shown in the bottom two panels by the dotted lines, with the histograms and their Poisson error bars overlaid with the best fit model (solid lines). The 1σ positional errors were found to be 2.49 ± 0.10 arcsec and 2.33 ± 0.09 arcsec in the two directions, consistent with one another within the errors. The advantages of this method are two–fold; firstly, it is not necessary to identify the counterparts to the 250 μm sources *a priori*, and secondly, the centroids of the best fit Gaussians may be used to determine astrometric corrections in the SPIRE maps (e.g. Pascale et al., 2010). The value for $\sigma_{\rm pos}$ that we adopted was the weighted mean 2.40 ± 0.09 arcsec.

Theoretically, the positional uncertainty should depend on the signal-to-noise ratio (SNR) of the detection and on the full-width at half maximum (FWHM) of the SPIRE 250 μm beam (18.1 arcsec, Pascale et al., 2010), following the results derived in Ivison et al. (2007; $\sigma_{\rm th} = 0.6\frac{\rm FWHM}{\rm SNR}$) and assuming the case of uncorrelated noise. We use our empirical results in Figure 1 to calibrate the theoretical relation presented in Ivison et al. (2007) to our data, and assume that our results are symmetric in RA and Dec. This leads us to introduce a factor of 1.09 to give equation 6:

$$\sigma_{\rm pos} = 1.09 \times \sigma_{\rm th} \quad (5)$$

$$= 0.655\frac{\rm FWHM}{\rm SNR}. \quad (6)$$

Although the SNR of some SPIRE 250 μm sources is very high, it is unphysical to allow $\sigma_{\rm pos}$ in equation 6 to approach zero for three main reasons:

- Whilst it is acceptable to neglect the SDSS DR7 positional errors for the purposes of determining $\sigma_{\rm pos}$ (section 2.1), the astrometric precision for sources in the SDSS DR7 catalogue is non-zero (< 0.1 arcsec – Abazajian et al., 2009).
- Large sources, especially those without Gaussian surface brightness profiles (e.g. bright spiral galaxies), have considerably larger positional uncertainties associated with them.
- Confusion provides a lower limit to the positional errors of the SPIRE catalogue, although the SNR in equation 6 does include confusion noise as described in Rigby et al. (2010) and Pascale et al (2010).

Other effects that can influence the positional uncertainty include imprecise knowledge of the beam morphology and the effects of drifts and jitter in the *Herschel* pointing model.

To account for these effects, we do not allow the positional uncertainty to fall below 1 arcsec, and we also include a term which adds 5 percent of the SDSS r–band isophotal major axis in quadrature to the value determined by equation 6, for those sources with r–band model magnitudes < 20.5. Finally, $f(r)$ must be renormalised so that

$$2\pi\int_0^\infty f(r)rdr = 1. \quad (7)$$

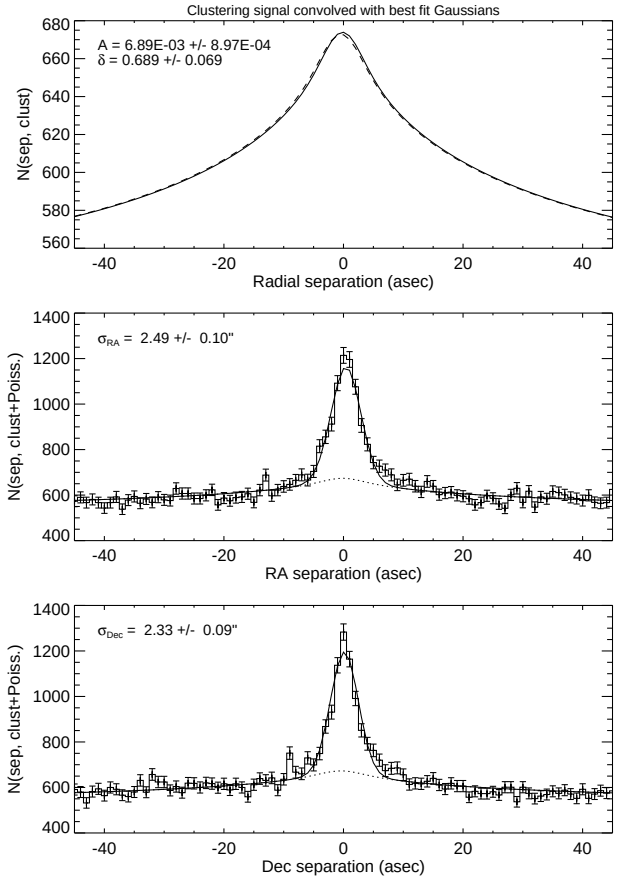


Figure 1. In order to derive the 1σ positional errors of the SPIRE 250 μm–selected sources, we produced histograms of the total number of SDSS sources within a box 50 arcsec on a side around the SPIRE 250 μm centres. After accounting for the clustering of SDSS sources (the top panel shows the signal expected for the clustering of SDSS sources in the RA and Dec directions convolved with Gaussian positional errors – the results are shown as solid and dashed lines for RA and Dec, respectively, and these results appear as the dotted lines in the bottom two panels), we can add in an appropriate Gaussian distribution of centres to account for the actual positions of the SPIRE sources (equations 3 & 4). Performing a χ^2 minimisation allows us to then empirically determine the 1σ positional uncertainty for these sources, which are shown as $\sigma_{\rm RA}$ and $\sigma_{\rm Dec}$.

2.2 Calculating the magnitude dependence of the likelihood ratio

Calculating the LR requires two further pieces of magnitude information, $n(m)$ and $q(m)$. The quantity $n(m)$ is simply the probability that a background source is observed with magnitude m. To estimate this, we calculate the distribution of SDSS DR7 r–band model magnitudes for all of the primary photometry sources in the catalogue, normalised to the total area of the catalogue (which is approximately 36.0 deg^2 for the SDSS catalogue that we use for this purpose).

The non-triviality lies in the calculation of $q(m)$ – the probability that a true counterpart to a 250μm source has a magnitude m. To estimate this we calculate the r–band magnitude distribution of the counterparts to the 250 μm sources using the method of Ciliegi et al. (2005). This method involves counting all objects in the optical catalogue within some fixed maximum search radius ($r_{\rm max}$) of the SPIRE positions. To avoid influencing the results of this analysis with erroneous deblends in the SDSS DR7 catalogue (which artificially alter the number counts), we eyeballed the SDSS r–band images of each of the 5σ 250 μm sources, removing 370 SDSS sources from the input catalogue. The magnitude distribution of the remaining objects is referred to as total(m). Here we have adopted $r_{\rm max} = 10$ arcsec, which encloses >99.996% of the real counterparts to the 250 μm sources based on our derived value for $\sigma_{\rm pos}$. The distribution total(m) is then background–subtracted to leave the magnitude distribution of excess sources around the 250 μm centres, real(m):

$$\mathrm{real}(m) = \left[\mathrm{total}(m) - \left(n(m) \times N_{\rm centres} \times \pi \times r_{\rm max}^2\right)\right], \quad (8)$$

where $N_{\rm centres}$ is the number of 250 μm sources in the catalogue. This enables us to empirically estimate $q(m)$ from the sources in our optical catalogue rather than modelling the r–band magnitude distribution of 250 μm–selected *Herschel*–ATLAS sources. The distribution $q(m)$ is given by equation 9:

$$q(m) = \frac{\mathrm{real}(m)}{\sum_m \mathrm{real}(m)} \times Q_0. \quad (9)$$

Q_0 is the fraction of true counterparts which are above the SDSS limit, and is calculated thus:

$$Q_0 = \frac{N_{\rm matches} - \left(\sum_m n(m) \times \pi r_{\rm max}^2 \times N_{\rm centres}\right)}{N_{\rm centres}}, \quad (10)$$

here $N_{\rm matches}$ represents the number of possible IDs within 10.0 arcsec of the SPIRE positions, and $N_{\rm centres}$ is defined as above. Since the value of Q_0 will be different for galaxies and unresolved sources in our catalogue, we must calculate $q(m)$, $n(m)$ and Q_0 separately for each population.

We separate resolved and unresolved sources using a slightly modified version of the GAMA colour–colour relation from Baldry et al. (2010, modified such that $\Delta_{sg,jk} > 0.40$ rather than 0.20 to avoid adding an unphysical sharp edge to the stellar locus in Figure 2). Having separated the two populations, we corrected the positions of the unresolved sources for known proper motions in the USNO/SDSS DR7 catalogue (Munn et al., 2004), precessing their co–ordinates to the epoch of the *Herschel*–ATLAS SDP observations. Only those unresolved sources with proper motions detected at a SNR $\geqslant 3$ were updated.

For our SDSS DR7 r–band catalogue, $Q_0^{\rm gal} = 0.583$, i.e. 58.3 percent of the galaxy counterparts are brighter than our magnitude limit. For the unresolved sources the value is $Q_0^{\rm unres} = 0.010$, indicating that only 1 percent of the unresolved sources in the catalogue are detected at $\geqslant 5\sigma$ in our 250μm data (although see section 2.3.1). Thus we determine that overall $Q_0 = Q_0^{\rm gal} + Q_0^{\rm unres} = 0.593$.

The distributions of $q(m)$, and $n(m)$ (as well as the magnitude dependence of the LR – $q(m)/n(m)$) are shown in Figure 3, in which the left and right columns show the values for the resolved and unresolved sources, respectively. While the $q(m)$ distribution for galaxies is well–sampled at $r \gtrsim 14$ mag, we assume that $q(m)/n(m)$ is constant for all sources brighter than this, enabling us to use our well-defined $n(m)$ to estimate $q(m)$ for the brightest galaxies.

Since the fraction of *Herschel*–ATLAS sources associated with unresolved counterparts is low (reflected in $Q_0^{\rm unres} = 0.010$), the method used to determine $q(m)$ for these sources differs. In order to ensure that the LR results for stars/QSOs are not domi-

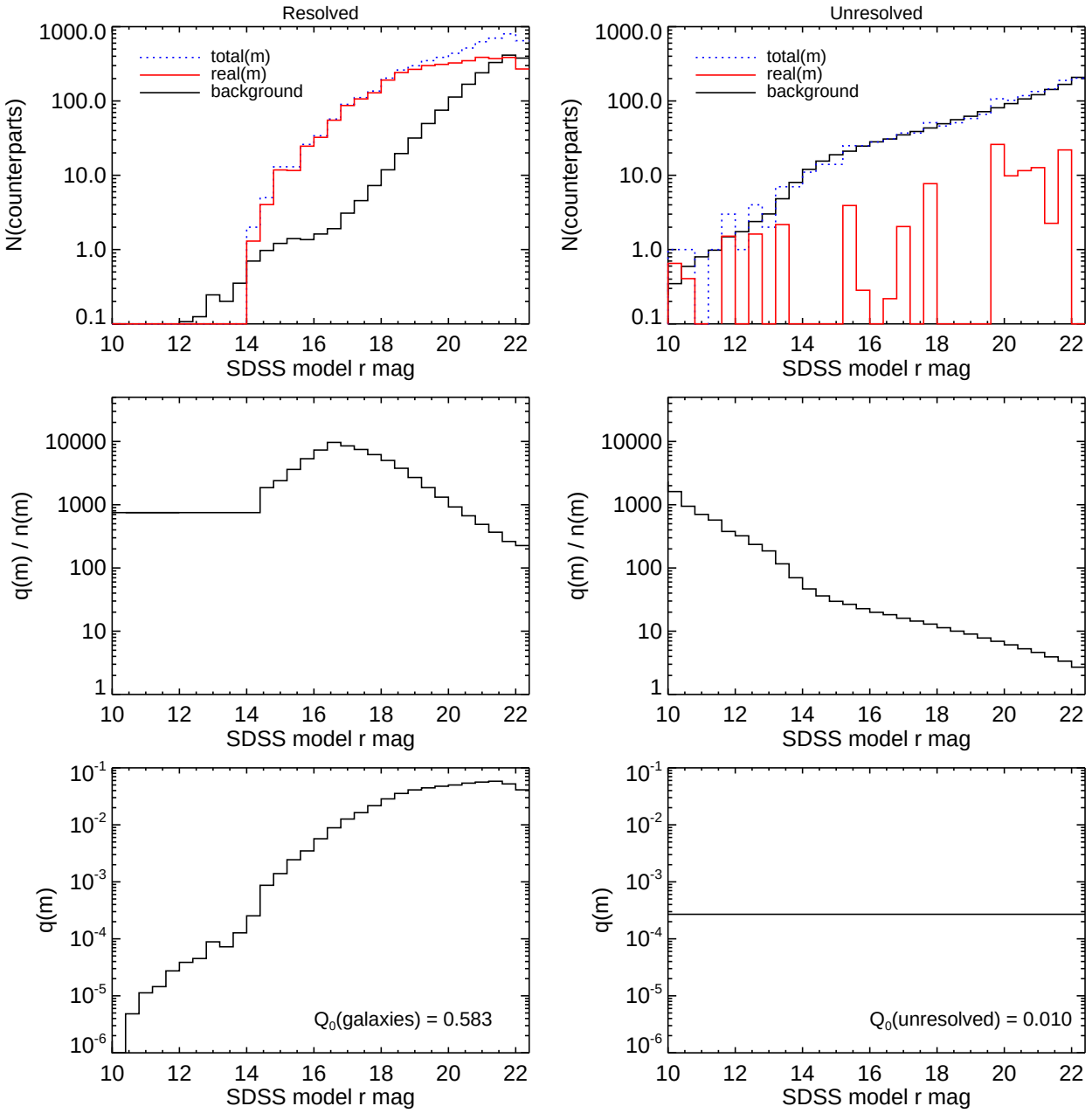


Figure 3. Deriving the magnitude dependence of the LR for the resolved and unresolved counterparts to the 250μm catalogue (left and right columns, respectively). The analysis for the **resolved** sources is discussed first, while the alternative procedure followed for the unresolved sources is described subsequently. **Top Left**: Total(m) (blue, dotted) represents the SDSS DR7 r–band model magnitude distribution of all the resolved sources that lie within 10.0 arcsec of the SPIRE 250 μm centres. The black histogram represents the number of galaxies that we would expect within these search radii due to the background SDSS number counts alone. The red histogram, dubbed real(m) as per Ciliegi et al. (2003), is the difference between the two, i.e. the SDSS DR7 r–band model magnitude distribution of the excess sources above the background. **Middle Left:** The ratio of $q(m)/n(m)$ represents the magnitude dependence of the LR. To avoid having a zero probability of a given source being the real counterpart due to our limited statistics on real(m) (and hence $q(m)$) at bright magnitudes, we use the ratio of $q(m)/n(m)$ in the brightest well-sampled bin ($r_{\rm mag} = 14.2$) to define the values of $q(m)$ for resolved sources with SDSS r–band magnitude $\leqslant 14.0$. **Bottom Left:** The resulting $q(m)$ distribution – our best estimate of the probability that a true counterpart to a 250 μm source has a magnitude m – using the $n(m)$ distribution to overcome the small number statistics at bright r–band magnitudes. To further reduce the effects of noise, we boxcar smooth the $q(m)$ distribution for resolved sources with a 3 bin kernel. Q_0 for the resolved sources is determined to be 0.583. For the **unresolved** sources (right column), which have considerably fewer excess sources (reflected in the lower value of $Q_0^{\rm unres}$), the corresponding analysis is slightly different. We define the magnitude dependence of the LR – $q(m)/n(m)$ – assuming a constant $q(m)$ (bottom panel, right column) normalised to reflect $Q_0^{\rm unres} = 0.010$. This situation will improve with the higher quality statistics that the full *Herschel*–ATLAS catalogue will produce.

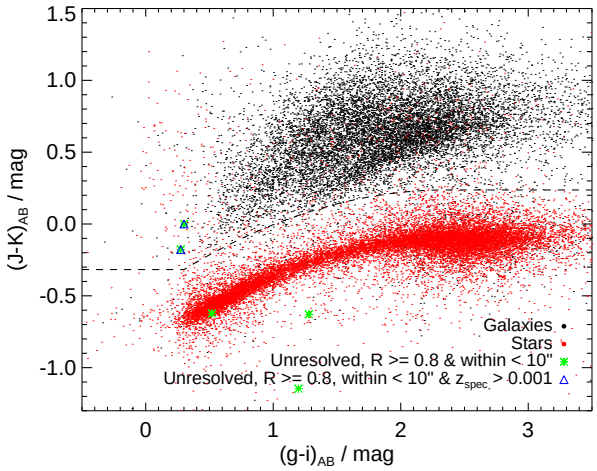


Figure 2. Colour–colour diagram for sources in the GAMA catalogue. We use the relationship of Baldry et al. (2010) to distinguish between unresolved sources (stars and QSOs) and galaxies. Those sources with stellar/QSO colours in the SDSS/LAS catalogue data over the GAMA 9 hour field are displayed in red, while those with galaxy colours are displayed in black. Objects with colours consistent with QSOs are located toward the upper left corner of this plot. Of the five $R \geqslant 0.8$ sources in our catalogue classified as unresolved, we find that three satisfy the GAMA colour selection criteria for being stellar, and so are potentially evolved stars, dust-obscured QSOs or debris disk candidates possibly indicative of a proto-planetary system (e.g. Thompson et al., 2010). The dashed line describes the first order star–galaxy separation locus (for more details see Baldry et al., 2010). The star–galaxy separation locus has been modified slightly from the Baldry et al. value due to the fainter magnitudes considered in our survey.

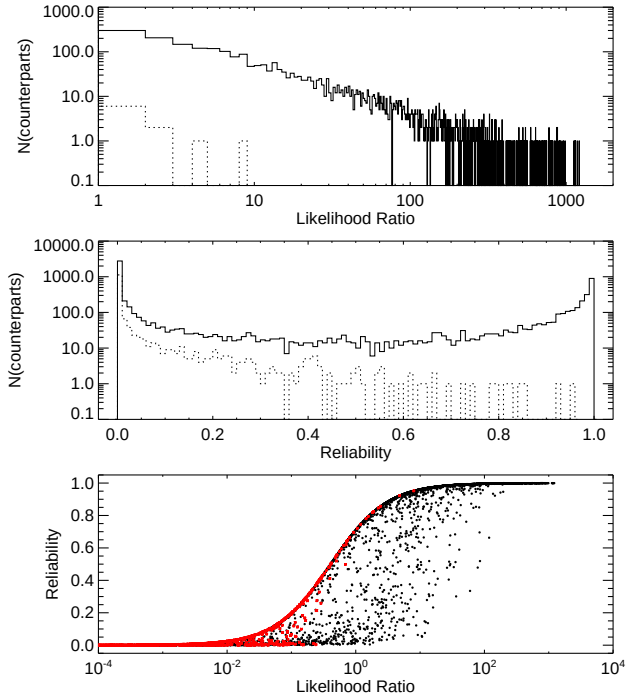


Figure 4. Top: Histogram of the Likelihood Ratio values for all of the 7230 potential counterparts to the 6621 5σ sources in the SPIRE SDP catalogue. The LR values for the resolved sources (i.e. galaxies) are shown as the solid histogram, with unresolved sources shown as the dotted histogram. Middle: Reliabilities for each counterpart. Once more, the solid histogram represents the resolved sources, while the dotted histogram represents the unresolved sources. There are a total of 2423 sources which have a reliability $\geqslant 0.8$, of which five are unresolved using the star/galaxy separation criteria of Baldry et al. (2010). Bottom: The variation of the reliability as a function of the likelihood ratio. This is not a linear relation since some sources have more than one counterpart with a high likelihood ratio. There are 263 SDSS r–band sources with reliability < 0.8 but $L > 1.63$ (the value above which $R > 0.8$ for a single counterpart within the 10.0 arcsec maximum search radius). These may be interacting systems, as discussed in section 4.3. These sources also demonstrate a possible limitation of the LR method, since the method implicitly assumes that there is only one true counterpart to a given 250 μm source.

nated by small number statistics, we assume a flat prior on $q(m)$, normalised to retain $Q_0^{\rm unres} = 0.010$ (figure 3).

We can correct our value for Q_0 for the clustering of SDSS sources by simply dividing Q_0 by $1 + \int_0^{10{\rm arcsec}} w(\theta) d\theta = 1.0008$ (remembering that θ is measured in degrees), giving a clustering–corrected value of $\tilde{Q}_0 = 0.592$. This value is broadly consistent with the recent results of Dunlop et al. (2010), who recover optical counterparts to 8 out of 20 250 μm sources brighter than 36 mJy in data from the BLAST observations of the GOODS-South field to a comparable i–band magnitude (albeit with lower angular resolution at 250 μm and much more sensitive optical, infrared and radio data), while Dye et al. (2009) found 80 counterparts to the 175 BLAST 250μm sources brighter than 55 mJy down to similar magnitude limits in r– or R–band data (S. Dye, private communication).

To account for the fact that an *Herschel*–ATLAS source may have more than one possible counterpart, we also define a reliability R_j for each object j being the correct counterpart out of all those counterparts within $r_{\rm max}$, again following Sutherland & Saunders (1992):

$$R_j = \frac{L_j}{\sum_i L_i + (1 - Q_0)}, \qquad (11)$$

where the LR values have been determined for the resolved and unresolved counterparts separately (see Figure 4). The reliability is a key statistic; we recommend using only those counterparts with reliability $R \geqslant 0.8$ for analysis, since this ensures not only that the contamination rate is low (see below), but also that only one r–band source dominates the far–infrared emission (as required for e.g. deriving spectral energy distributions for 250 μm–selected galaxies in the *Herschel*–ATLAS catalogue, Smith et al. *in prep*). This is more conservative than other works in the literature (e.g. Chapin et al., 2010), where the chosen LR limit was defined based on a 10 percent sample contamination rate.

In order to estimate the number of false IDs in our reliable sample, we calculate:

$$N({\rm false}) = \sum_{R \geqslant 0.8} (1 - R). \qquad (12)$$

As a result we expect 103 false IDs in our sample, which corresponds to a contamination rate of 4.2%. For those investigations in which it is desirable only to determine whether an optical source is associated with an *Herschel*–ATLAS object (with additional caveats about lensed sources and the de-blending efficiency in the optical catalogue), it is sufficient to use a likelihood ratio cut (e.g. $L > 5.0$, i.e. the source is 5 times more likely to be associated with the sub–millimetre object than it is to be a chance superpo-

Table 1. The distribution of the number of SDSS r–band sources within 10.0 arcsec of the 250 μm positions, and the fraction of reliable counterparts. There are 2869 sources with only one possible match within 10.0 arcsec, and yet only 1389 of these are determined to be reliable; the vast superiority of the LR technique over a simple nearest-neighbour algorithm is evident.

N(matches)	N(250 μm sources)	$N(R \geqslant 0.8)$	%
0	1865		
1	2869	1389	48.4
2	1400	782	55.9
3	400	210	52.5
4	76	38	50.0
5	9	4	44.4
6	2	0	0.0
TOTAL:	6621	2423	36.6

sition of sources). This aspect of the likelihood ratio technique is discussed in more detail in section 4.3.

In Table 1, we present the number of possible optical counterparts within 10.0 arcsec of the 250 μm sample, including the relative fractions of reliable associations. Only half of the 250 μm sources with a single optical counterpart within the search radius are deemed reliable.

To estimate the fraction of 250 μm sources with a counterpart above our detection limit recovered as having $R \geqslant 0.8$, we assume that 250 μm–selected SDSS sources cluster in the same way as SDSS r–band–selected sources (the results of Maddox et al., 2010, suggest that this assumption is reasonable). Under this assumption, we may calculate the completeness, η, of the reliable sources in our sample:

$$\eta = \frac{n(R \geqslant 0.8)}{n(250\mu m > 5\sigma)} \times \frac{1}{\tilde{Q}_0}. \quad (13)$$

We have reliably identified η =61.8 percent of the optical counterparts bright enough to be detected in the SDSS r–band catalogue. This constitutes an overall identification rate of 36.6 percent for $\geqslant 5\sigma$ 250 μm sources in the *Herschel*–ATLAS SDP observations.

2.3 Checking the identification process

By selecting sources at 250 μm rather than longer wavelengths, the negative k-correction that results in e.g. 850 μm–selected galaxies residing at a median redshift of $z \sim 2$ (Chapman et al., 2003) has a much less dramatic effect, and observations have shown that a significant fraction of 250 μm–selected galaxies reside at $z < 1$ (e.g. Chapin et al., 2010, Dunlop et al., 2010, Dye et al., 2010). As a result, their optical counterparts will be much brighter than 850 μm–selected sources, and therefore readily detectable by shallower optical/near–infrared imaging with a much lower source density. For the SDSS DR7 r–band source catalogue that we use for the purposes of this investigation, we expect only ~ 0.48 background sources within the 10.0 arcsec search radius, down to the magnitude limit of $r = 22.4$ mag (of these, $\sim$0.26/0.22 will be resolved/unresolved, respectively). Furthermore, these background sources may be expected to be evenly distributed throughout the area within the maximum search radius, unlike the true counterparts.

We performed the following simple checks to determine the effectiveness of the LR technique for the *Herschel*–ATLAS SDP catalogue.

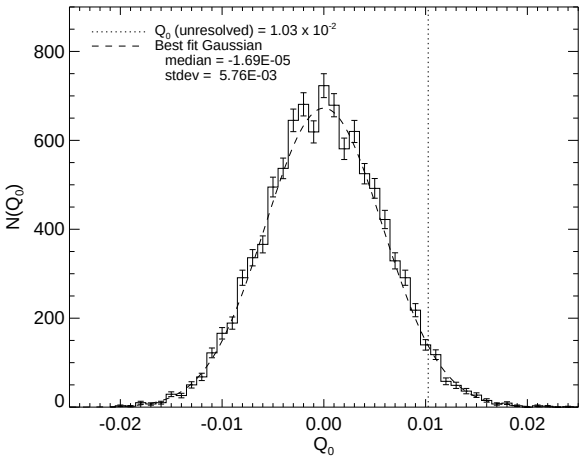


Figure 5. Histogram of the 10000 realisations of determining Q_0 for 6621 randomly–positioned *Herschel*–ATLAS sources, designed to determine the 1σ uncertainty on Q_0. A Gaussian distribution with median and standard deviation derived from the histogram is overlaid (dashed line), and the value of $Q_0^{\rm unres}$ is indicated by the vertical dotted line.

2.3.1 LR analyses of random catalogues

As a first test of whether the LR technique produces sensible results, we wanted to test the method in the absence of any true association between the 250 μm and SDSS positions. We randomised the positions of the 6621 sources and re-ran the LR analysis 10000 times, recording the derived value of Q_0 each time. The histogram of the resulting Q_0 distribution had a median of 0.000 with a 1σ uncertainty of 0.006. In these cases, where $Q_0 \approx 0$, the values for L and hence R are unreliable, since the distributions of total(m) and real(m) that we determine are almost identical, and the latter is strongly affected by noise as a result (see section 2.2). The histogram of the simulated values for Q_0 is shown in Figure 5, with a Gaussian distribution with appropriate median and standard deviation overlaid (dashed line). The derived value of $Q_0^{\rm unres}$ determined in section 2.2 is overlaid as the vertical dotted line. With $Q_0^{\rm unres}$ residing within 2σ of the median Q_0 value for these random catalogues, it is not clear that the population of unresolved sources is detected in the *Herschel*–ATLAS SDP data (see section 4.4). However, to avoid the possibility of missing real counterparts of potentially great scientific importance, we must not ignore the possibility that unresolved sources are detected in our 250 μm catalogue.

2.3.2 LR analyses of SDSS galaxies

We also performed a test in which we replaced the SPIRE positions in our LR analysis with the positions of SDSS galaxies, while retaining the 250μm fluxes and errors in order to accurately reproduce the positional uncertainties according to the method given in section 2.1. We found that we recovered 97.2% of the SDSS galaxies at $R \geqslant 0.8$, which reduced to 93.9% when the SDSS positions were varied according to a Gaussian positional offset with standard deviation appropriate for the signal-to-noise ratio of the real SPIRE sources (according to the rescaled formula given in section 2.1). Comparing this value to our overall ID rate we see that we are

not typically missing counterparts because we have underestimated their positional errors, but because of the fact that approximately 39% of 250 μm sources (the actual value is $1 - \tilde{Q}_0$) are not detected in SDSS r–band data down to the magnitude limit of our survey.

We also compared our new catalogues with the small subsets of overlapping objects in the Imperial *IRAS*-FSC Redshift Catalogue (IIFSCz, Wang & Rowan-Robinson, 2009) and FIRST surveys (Becker, White & Helfand, 1995). These comparisons are presented in detail in appendix A, but to summarize:

- We find updated positions for five IIFSCz galaxies which previously had misidentified or unidentified counterparts.
- Our catalogue is consistent with a catalogue of FIRST radio sources matched to the 250 μm sample, provided that the radio sources are detected in our optical data, and that the optical counterparts contain only single components at moderate separations from the 250 μm centroids.
- A small collection of *Spitzer Space Telescope* snapshot images taken at near infrared wavelengths reinforce our belief in the accuracy of our method.

3 REDSHIFTS IN THE *HERSCHEL*–ATLAS 9HR FIELD

3.1 Spectroscopic Redshifts

The GAMA catalogue (Driver et al., 2010) contains 12,626 new spectroscopic redshifts in the *Herschel*–ATLAS SDP region for sources satisfying the GAMA target selection criteria (including magnitude limits of $r < 19.4$, $z < 18.2$ & $K < 17.6$ – Baldry et al. 2010). In addition, there are a further 3281 redshifts available in this region from the SDSS DR7, 248 from the 2SLAQ-LRG survey (Cannon et al., 2006), 939 from the 2SLAQ-QSO survey (Croom et al., 2009) and 29 from the 6dFGS (Jones et al., 2009). 1099 spectroscopic redshifts for reliable counterparts were collated (including those from the SDSS DR7, 6dFGS, 2SLAQ-QSO/LRG surveys and the GAMA catalogue), meaning that 41.0% of our $R \geqslant 0.8$ counterparts have spectroscopic redshifts (and 15.0% of all $\geqslant 5\sigma$ 250 μm sources). We note that none of the spectroscopic catalogues (with the exception of the 6dFGS catalogue) extends to declinations less than -1 deg. The number of redshifts for reliable counterparts from each spectroscopic survey is presented in table 2, and the redshift properties of 250 μm selected galaxies are discussed in section 3.3.

3.2 Photometric Redshifts

For those sources without spectra, we estimate photometric redshifts using optical and near–infrared photometry. The *Herschel*–ATLAS SDP field has almost complete optical coverage in $ugriz$ from the SDSS DR7 and near–infrared $YJHK$ photometry from 7th data release of the UKIDSS Large Area Survey (Lawrence et al. 2007). As well as having spectroscopy from GAMA and SDSS, these very wide-area surveys overlap with several deeper spectroscopic surveys (Davis et al., 2003, Cannon et al., 2006, Davis et al., 2007, Lilly et al., 2007) which allow us to construct a spectroscopic training set with large numbers of objects (> 1000 per bin of unit magnitude or 0.1 in redshift) up to r–band magnitudes $r < 23$ and redshifts $z < 1.0$, i.e. to approximately the photometric depth of SDSS and UKIDSS-LAS.

This large and relatively complete training set allows us to use an empirical regression method to estimate the photometric redshifts. We use the well-known artificial neural network code ANNZ (Collister & Lahav, 2004) with a network architecture of $N : 2N : 2N : 1$, where N is the number of photometric bands used as inputs; although there are 9 photometric bands ($ugrizYJHK$) available in this case, we ignore bands where an object has no coverage or where the photometry is flagged as dubious, and train separate neural networks for all combinations of bands with at least three good detections. An advantage of ANNZ is that it provides redshift error estimates, σ_z, based on the photometric errors it is supplied with; we checked that these errors were distributed correctly by confirming that, for a set of validation data with spectroscopic redshifts, $(z_{\rm phot} - z_{\rm spec})/\sigma_z$ follows a Gaussian distribution centred on zero – however we found that the width of the best-fitting Gaussian was ~ 1.4, indicating that the errors were underestimated by this factor on average. To improve the accuracy of the error estimates, we used the width of this distribution to correct the error estimates individually for each trained network, with correction factors of typically 1.3 to 2.

Confirming the accuracy of empirical photometric redshifts is always difficult, since the objects for which we have spectroscopic redshifts for comparison are, by necessity, drawn from the same sample which is used to train the neural network, and may not be fully representative of the whole population of objects to which the method is applied. A particular concern in our case was that the *Herschel* counterparts are likely to have a different distribution in colour space than the training-set galaxies, and so any bias in the photometric redshift as a function of colour could cause the average redshift of *Herschel* counterparts to be systematically wrong. However, we satisfied ourselves that this was not a problem by looking for trends in the difference between photometric and spectroscopic redshift as a function of colour, and finding no significant trend with any colour. For example, the best-fitting straight line relationship between $(z_{\rm phot} - z_{\rm spec})$ and $(r - K)$, for objects in our validation dataset, has a gradient of 0.00035 – two orders of magnitude smaller than the scatter of $(z_{\rm phot} - z_{\rm spec})$, which for the same sample has a standard deviation of 0.037 overall.

In the event that the counterparts are so obscured that they are invisible in the optical data, they will clearly have unreliable photometric redshifts (this is inevitable, given the small number of detections that would be available), but this scenario will become apparent as large errors on the photometric redshifts of these sources. In any case, these sources will not pass our $r < 22.4$ mag selection criterion. With this new catalogue of photometric redshifts, all sources detected in at least three photometric bands have either a spectroscopic or photometric redshift.

3.3 Redshift distribution and completeness of reliably-identified 250 μm sources

Using a method analogous to that used in section 2.2 to calculate the r–band magnitude distribution of counterparts to 250 μm sources, we may determine the completeness of our reliably–matched 250 μm and SDSS objects as a function of redshift. Here we define the completeness as the fraction of reliably identified counterparts, compared with all of those counterparts that are detected in our data i.e. $Q_0 \times 6621 \approx 3925$ sources. First, we use the star-galaxy separation method from Baldry et al. (2010 – discussed in section 4.4) to ensure that only galaxies remain in our sample. We then use the catalogue of photometric and spectroscopic redshifts discussed in section 3.2 to determine the distribution of the galaxy redshifts in the catalogue, $n(z)$. Since we know that this catalogue covers an area of 36.0 deg^2, we can scale to the total

Table 2. The number of spectroscopic redshifts for reliable counterparts from each survey used in our final catalogue. The majority of the spectroscopic redshifts used in our follow-up studies of the reliable 250 μm associations come from GAMA. There are a total of 1099 $R \geqslant 0.80$ 250μm counterparts with spectroscopic redshifts.

Survey	Reference	Number of redshifts		Percentage of
		$R \geqslant 0.8$	Catalogue	*Herschel*–ATLAS IDs
2SLAQ-LRG	Cannon et al. (2006)	3	248	0.12
2SLAQ-QSO	Croom et al. (2009)	4	939	0.17
6dFGS	Jones et al. (2009)	12	29	0.50
GAMA	Driver et al. (2010)	766	12626	31.6
SDSS DR7	Abazajian et al. (2009)	316	3281	13.0

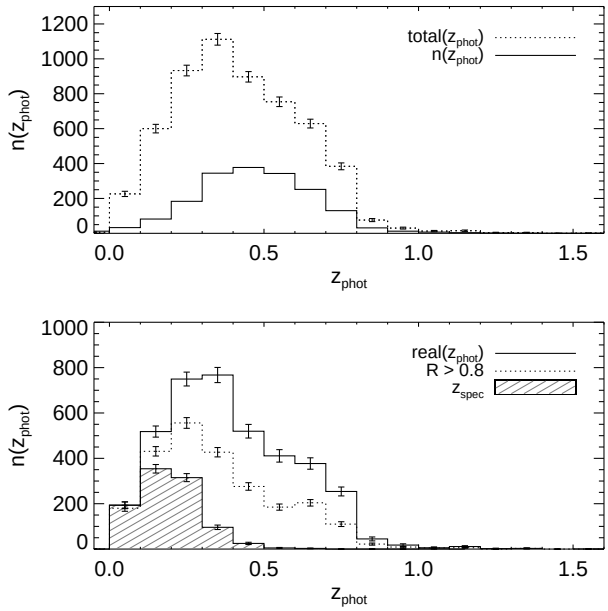


Figure 6. Top: total(z) (dotted histogram) represents the photometric redshift distribution of all of the sources within 10.0 arcsec of a 250 μm position, while $n(z)$ (solid line) represents the expected background based on the area covered and the redshift catalogue number counts. We note that the histogram of background sources is consistent with the results of Oyaizu et al. (2008). Bottom: real(z) (solid line) shows the redshift distribution of the excess sources above the background around the 250 μm centres, and is compared with the dotted line which shows the photometric redshift distribution for those counterparts with $R \geqslant 0.8$ in our LR analysis. We note that real(z) peaks at a lower redshift than the intrinsic $n(z)$ for SDSS galaxies (solid histogram, top panel). The shaded histogram shows the number of spectroscopic redshifts for galaxies (as defined by the star galaxy separation criteria in Baldry et al., 2010) in our sample. The percentage completeness in our catalogue is given in Table 3.

sky area searched around the 250 μm sources ($6621 \times \pi r^2_{\rm max}$) to determine the expected background, n(z) (see Figure 6). Subtracting total(z) (the z distribution of all sources within 10.0 arcsec of the 250 μm centres) from $n(z)$ then gives us the number of excess sources around the 250 μm positions, real(z), and by comparing this with the photometric redshift distribution for those sources with $R \geqslant 0.8$ we can estimate the completeness of our reliable catalogue as a function of z. The completeness values for our reliable catalogue in bins between $0.0 < z < 1.1$ are presented in Table 3.

Figure 6 also shows the redshift distribution of 250 μm–selected *Herschel*–ATLAS sources with reliable counterparts and

Table 3. The percentage completeness of our reliable catalogue as a function of photometric redshift. These values are derived as described in section 3.3, using a method analogous to that applied to determine the r–band magnitude distribution of 250 μm sources. The errors are determined assuming that they are dominated by the Poissonian errors on real($z_{\rm phot}$).

$z_{\rm phot}$	Completeness (%)	$\sigma_{\rm comp}$
0.0–0.1	93.2	7.5
0.1–0.2	83.2	4.8
0.2–0.3	74.2	4.2
0.3–0.4	55.6	4.8
0.4–0.5	53.1	6.0
0.5–0.6	45.0	7.4
0.6–0.7	54.1	7.0
0.7–0.8	43.3	10.0
0.8–1.1	52.7	16.7

spectroscopic redshifts in our GAMA/SDSS DR7 catalogue. The redshift distribution of reliable r–band counterparts peaks at $z_{\rm phot} = 0.25 \pm 0.05$, with a median value of $z_{\rm phot} = 0.31^{+0.28}_{-0.15}$, or $z_{\rm spec} = 0.18^{+0.10}_{-0.09}$ if only spectroscopic redshifts are included (here the errors on the peak are based on the half-width of one histogram bin, and those on the median values are derived according to the 16th and 86th percentiles of the redshift cumulative frequency distribution). The disparity between these median redshift values is to be expected since the photometric redshifts are computed to fainter magnitude limits than the spectroscopic redshifts have been measured in the GAMA 9hr catalogue.

It is also interesting to note that the photometric redshift distribution of excess 250 μm sources – real($z_{\rm phot}$) – peaks at a lower redshift than the intrinsic redshift distribution of the SDSS photometric redshift catalogue, $n(z_{\rm phot})$. This is not due to our inability to reliably identify sources at higher redshifts, since the N(z) includes statistical non-detections, and indeed the magnitude distribution of the true counterparts (calculated in section 2.2) peaks at brighter magnitudes then the background $n(m)$. This determination of real($z_{\rm phot}$) indicates that the low redshift population of *Herschel*–ATLAS galaxies in our 5σ sample is generally at lower redshift than the average SDSS galaxy, raising an interesting question about the $\sim$ 40% of *Herschel*–ATLAS sources which do not have a counterpart above the SDSS DR7 limit. The redder sub-mm colours of these blank field sources suggests that they are at much higher redshifts (see Figure 7, and Section 4.1), and furthermore, the study of *Herschel*–ATLAS colours by Amblard et al. (2010) indicates a second population of sources at $z \sim 2$. The fact that we do not see a rising $n(z)$ for *Herschel*–ATLAS sources in SDSS out to the SDSS limit suggests that the total *Herschel*–ATLAS $n(z)$ is bimodal, with a low-redshift peak at $z \approx 0.35 \pm 0.05$ (where the error

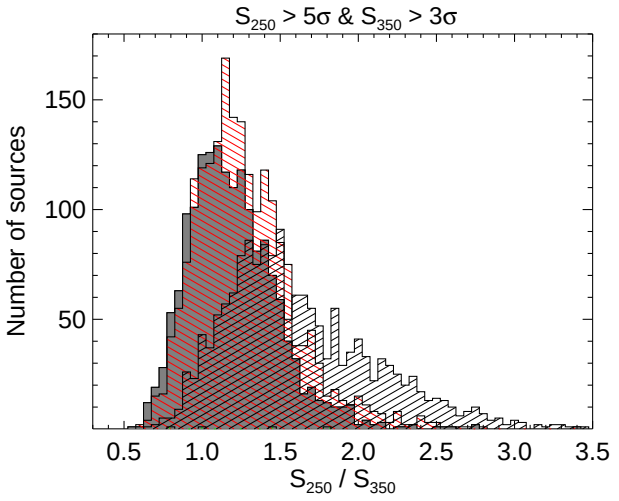


Figure 7. Histograms of the SPIRE colours for those sources with reliable galaxy counterparts ($R \geqslant 0.8$, shaded black), those for which the most reliable candidate has $R < 0.8$ (shaded red), and those remaining sources in the MAD-X catalogue without any SDSS r–band counterparts. All sources are detected at $\geqslant 5\sigma$ in the 250 μm band, and $\geqslant 3\sigma$ in the 350 μm band. Those galaxies for which we identify $R \geqslant 0.8$ counterparts have considerably bluer average S_{250}/S_{350} colours than those for which we are unable to identify a counterpart.

is once more derived according to the width of one histogram bin) and a higher-redshift peak at $z > 1$. Such behaviour is predicted in several models of sub-mm galaxy populations (e.g. Lagache et al. 2004; Negrello et al. 2007, Wilman et al., 2010) and also suggested by stronger clustering in samples of *Herschel*–ATLAS galaxies selected to have redder far–infrared colours (Maddox et al. 2010), as well as the steep upturn in the *Herschel*–ATLAS number counts at fluxes below 100 mJy (Clements et al. 2010) and the results of BLAST, which include deeper optical samples with fainter spectroscopy, albeit with smaller object samples by more than an order of magnitude (Dunlop et al. 2010 and Chapin et al. 2010).

4 RESULTS

The resulting values for the likelihood ratio and reliability for the 6621 5σ sources in the 250 μm selected catalogue are shown in Figure 4. Approximately 58% of galaxies with $S_{250\,\mu\mathrm{m}} \geqslant 32$ mJy are detected in our $r \leqslant 22.4$ mag SDSS catalog, and of those we identify 2423 counterparts with reliability $R \geqslant 0.8$, which we consider robust. Of these reliable counterparts, 1252 also have GALEX detections in at least one ultraviolet band, and each source has either a reliable spectroscopic redshift (1099 galaxies) or photometric redshift.

Figure 8 shows the fractional completeness in our identification catalogue as a function of the 250 μm flux, and of the SDSS DR7 r–band magnitude of the counterparts. The shaded areas indicate the 1σ uncertainty on the completeness derived from the Poisson errors on the number of sources brighter than a given magnitude/flux.

4.1 The sub-millimetre colours of SPIRE sources

In Figure 7 we display the S_{250}/S_{350} colours of the 250 μm sources from the *Herschel*–ATLAS SDP catalogue with detections at $\geqslant 5\sigma$ in the 250 μm band, and $\geqslant 3\sigma$ in the 350 μm band. The sources have been divided into three sub-sets; those sources with reliable r–band counterparts classified as galaxies ($R \geqslant 0.8$, black shaded histogram), those for which the most reliable candidate has $R < 0.8$ (red shaded histogram) and all of the remaining sources in the MAD-X catalogue without any r–band counterparts (grey shaded histogram). The median SPIRE colours for the three samples are quite different, with median colours of $S_{250}/S_{350} = 1.51, 1.23, \& \ 1.16$ for the three respective samples. It is clear that the $R \geqslant 0.8$ sources are considerably bluer than the SPIRE sources without (reliable) counterparts, and a series of Kolmogorov–Smirnov tests confirms that no two of the three sets of histograms are drawn from the same parent distribution at $\gg 99.9999\%$ confidence.

The differences between the colour populations may be due to those sources without reliable counterparts residing at higher redshift than those for which we can identify reliable counterparts, causing the peak of each such source's far–infrared spectral energy distribution to move to longer wavelengths.

It is also clear that those sources for which the most reliable candidate counterpart has $R < 0.8$ are a different population from those for which we identify no r–band counterparts. Half of these sources can be explained by the expected number which are above the SDSS limit but for which we cannot determine a reliable counterpart, and the other half simply have an unrelated SDSS r–band source within the search radius. This is reflected in the histogram for these sources having colours intermediate between the reliable counterparts and the "no potential counterparts" samples - it contains roughly equal fractions of both types of object (presumably high and low redshift).

We also compare the far–infrared colours with the results of Amblard et al. (2010 – their Figure 1 and our Figure 9). In this figure (which uses the same colour scheme as Figure 7, with the $R \geqslant 0.8$ counterparts in black) we consider only those sources at $\geqslant 5\sigma$ in the 250 μm and 350 μm bands and $\geqslant 3\sigma$ in the 500 μm bands to ensure a fair comparison. We identify 133 such sources with $R \geqslant 0.8$ counterparts in our SDSS r–band catalogue.

4.2 Lensed sources in *H*-ATLAS

Wide–field sub–millimetre wavelength surveys such as the *Herschel*–ATLAS are particularly well-suited to detecting large numbers of strongly–lensed sources (e.g. Blain, 1996, Negrello et al., 2007), in which intrinsically faint distant galaxies may be magnified by an otherwise unrelated foreground massive object along the line of sight (e.g. galaxy, galaxy cluster), and observed at more readily–detectable flux densities. Strong lensing can not only amplify the brightness of these distant soures, but also increase their angular size, allowing galaxies to be studied on scales smaller than would otherwise be possible, making samples of strongly–lensed galaxies an important cosmological probe (e.g. Swinbank et al. 2010). At bright 500 μm flux densities (> 100 mJy), after removing very local galaxies and blazars from the source counts, the surface density of sources on the sky is dominated by strongly-lensed galaxies, with large–area surveys such as *Herschel*–ATLAS required to detect them due to their paucity on the sky ($\sim 0.5\,\mathrm{deg}^{-2}$). This method of selecting lensed sources has one huge advantage over other methods, in that the selection efficiency is almost 100 percent (Negrello et al., 2010).

Of particular interest in Figure 9 is the positioning of the 58 $R \geqslant 0.8$ galaxies with $S_{250}/S_{350} \leqslant 1.5$. The redshift loci of Arp220 and M82 templates from Silva et al. (1998), shown in green

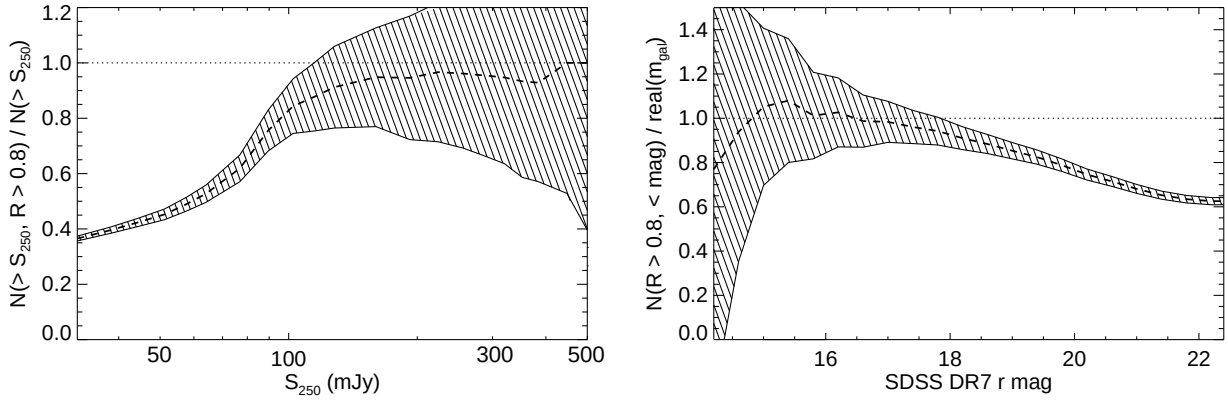


Figure 8. Left: Completeness of our identification catalogue as a function of 250 μm flux. The ordinate indicates the fraction of 250 μm sources brighter than the flux given by the abscissa, for which we have reliably identified counterparts in the SDSS DR7 r–band data. We can reliably identify the counterparts to 36.6 percent of the 6621 sources in our 5σ 250 μm–selected catalogue down to the limit of our r–band SDSS DR7 catalogue. Right: The fraction of sources with statistical identifications in our SDSS DR7 r–band catalogue (i.e. real(m)) which can be reliably identified with $R > 0.80$, not accounting for the value of Q_0. We reliably identify $\sim$63% of those counterparts that are detected in our $r < 22.4$ mag survey data. The shaded regions indicate the 1σ uncertainties in these completeness fractions, determined based on the Poisson errors on the number counts.

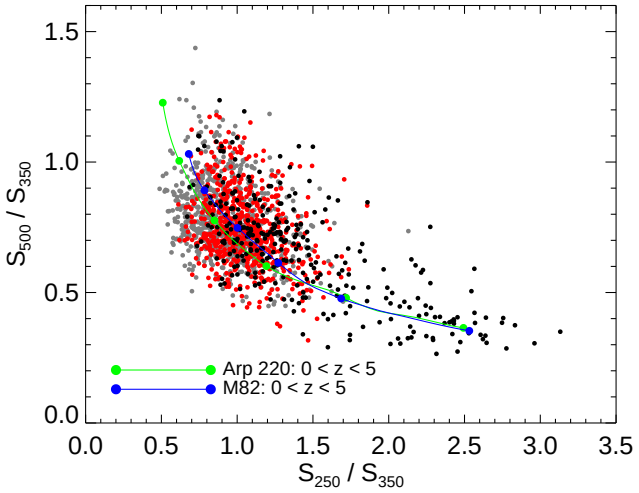


Figure 9. SPIRE colour–colour diagram, showing the colours of 5σ 250 μm and 350 μm sources with $\geqslant 3\sigma$ detections in the 500 μm band. The colour scheme is as in Figure 7. The green and blue lines represent redshift colour tracks between $0.0 < z < 5.0$ based on Arp220 and M82 template SEDs from Silva et al., 1998, with solid circles along the tracks indicating the locations in colour–colour space of integer redshifts between these two values for that template.

and blue respectively, suggest that sources with such colours reside at $z > 1.0$, despite the photometric and spectroscopic redshifts of their reliable optical counterparts residing at $z < 1.0$ (Figure 6). This disparity is, for some of these sources, caused by the blending of galaxies in the 350/500 μm bands (the results of Rigby et al., 2010, suggest that the extracted 500 μm flux densities of more than a quarter of $> 5\sigma$ sources are enhanced by factors of up to $\sim$2 due to multiple sources residing within a beam, for example). However, it is also possible that some of these are intrinsically high–redshift far–infrared sources which are strongly lensed by low–redshift foreground galaxies. The models of Negrello et al. (2007) predict that the fraction of lensed sources to these sensitivity limits is $\sim$4%. In our catalogue, $\sim$14% of *Herschel*–ATLAS sources with low–redshift $R \geqslant 0.8$ counterparts have $S_{250}/S_{350} \leqslant 1.5$, consistent with high-$z$ galaxies (209 out of 1480 sources that are detected at $\geqslant 5\sigma$ at 250 and 350 μm and $\geqslant 3\sigma$ at 500 μm). These numbers suggest that approximately one third of these sources may be strongly-lensed galaxies, although more realistic simulations will be required to thoroughly test this interesting hypothesis.

4.3 Multiple sources

We may also use our catalogue to identify 250 μm sources with multiple counterpart galaxies, by considering those which have at least one $L > 5.0$, $R < 0.80$ optical source within 10 arcsec (of course, this will also select low-probability superpositions of sources on the sky, e.g. Arp, 1967). There are a total of 118 such 250 μm sources which have at least one counterpart with $L > 5.0$ and $R < 0.8$ in our catalogue. It is possible that these sources contain multiple interacting counterparts, and indeed four of these sources have at least two counterparts with spectroscopic redshifts with $\Delta z \lesssim 0.001$ (including one of the radio sources mentioned in appendix A2, H–ATLAS J090631.3+004605). SDSS three-colour images of each spectroscopically-confirmed galaxy interaction are shown in Figure 10. There may be further examples for which we do not have spectroscopic redshifts.

For a more "complete" sample of cross–identifications, sources above some threshold in L could be considered, however in this case there is no immediate information to decide which of the multiple counterparts contributes most to the SPIRE flux, without resorting to priors on e.g. the colours of sources (e.g. Roseboom et al., 2009). This is work which we are pursuing and will look to implement in our next data release.

Finally, there is one additional source (H–ATLAS 090130.2-00215) with two high LR counterparts that have differing spectroscopic redshifts. This 250 μm source has counterparts with $L = 15.8$ & 35.0, residing at $z_{\rm spec} = 0.196$ and $z_{\rm spec} = 0.255$, respectively. The latter counterpart is also a $P < 0.20$ radio source, mentioned in appendix A2, and presumably constitutes one of the low-probability superpositions mentioned above. We also note that merging sources may have real positional offsets between the dust emission in the far–infrared and the starlight which dominates the

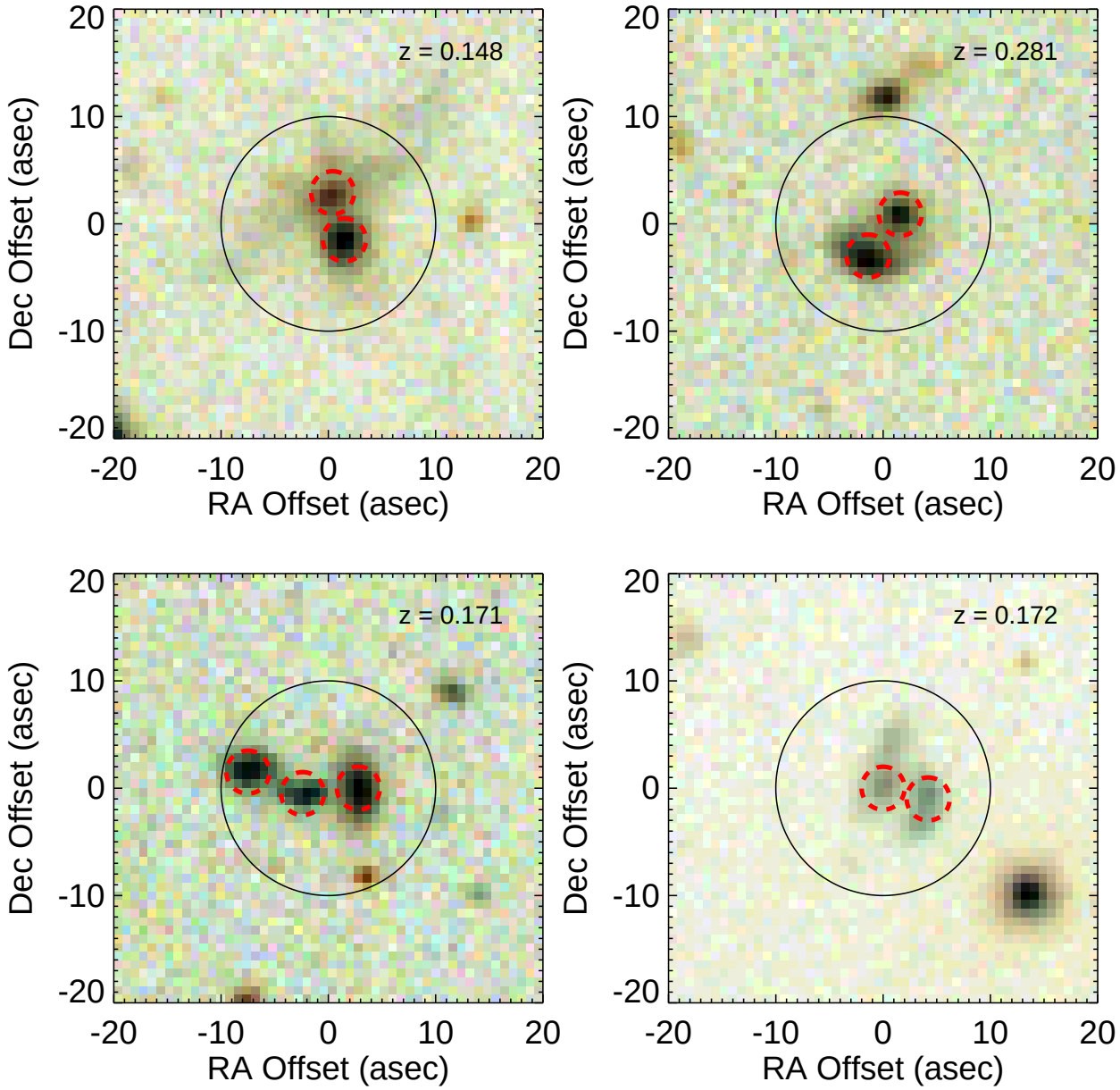


Figure 10. SDSS gri colour images centred on the positions of four SPIRE 250 μm 5σ sources with at least one counterpart with $L > 5.0$ and $R < 0.8$, and with at least two spectroscopic redshifts within $\Delta z = 0.001$ of each other (indicated by the red dashed circles). Each image is 40 arcsec on a side orientated such that North is up and East is to the left.

optical (see e.g. Zhu et al., 2007, Ivison et al., 2008, or Smith et al., 2010a).

4.4 Unresolved sources

Although the main focus of this paper is the reliable identification of galaxies selected at 250 μm, we have also applied the LR method separately to identify any reliable unresolved sources. We have used our spectroscopic data set to further split the population of unresolved sources in to groups of candidate stars and QSOs. There are a total of five $R \geqslant 0.80$ unresolved sources in the 250 μm selected sample (green asterisks in Figure 2), of which three occupy the stellar colour-colour locus, while two have colours or spectroscopic redshifts consistent with being QSOs.

Studying these objects in detail is beyond the scope of this paper, but see e.g. Thompson et al., (2010) for a discussion of stellar sources in the *Herschel*–ATLAS and the identification of two candidate debris disks.

5 CONCLUSIONS

We have demonstrated that the likelihood ratio method of Sutherland & Saunders (1992) is an appropriate way to determine reliable counterparts for 250 μm–selected galaxies from the *Herschel*–ATLAS science demonstration phase observations in the SDSS

DR7 r–band observations of the GAMA 9 hour field. We have determined reliable ($R \geqslant 0.8$) counterparts to 2423 out of 6621 sources detected at a SNR $\geqslant 5$, and found that $\sim$59.3% have counterparts brighter than r = 22.4 (the limit of our catalogue). We identify reliable counterparts to 36.6% of the 250μm sources (2423 out of 6621), and our calculations in section 2.2 suggest that our sample is 61.8% complete down to the SDSS r–band limit of our catalogue, in the sense that we have reliably identified 2423 counterparts out of the $\tilde{Q}_0 \times 6621 \approx 3925$ counterparts that are actually detected in the SDSS DR7 data.

We show from a consideration of their sub-mm colours that those sources without optical counterparts appear to reside at higher redshifts than those with optical counterparts in our available ancillary data. We compute the completeness of our reliable catalogue as a function of redshift, and find that *Herschel*–ATLAS sources with SDSS counterparts have a lower median redshift than the general SDSS population, suggesting a bimodal $n(z)$ for *Herschel*–ATLAS sources. For this bimodal $n(z)$, we find that the lower redshift population has a median redshift of $z = 0.40^{+0.25}_{-0.19}$ (with the errors calculated according to the 16th and 86th percentiles of the redshift cumulative frequency distribution), and that the high redshift population peaks at $z > 1$. We also find evidence for a population of sub–millimetre–selected interacting galaxies, and suggest a possible method for selecting samples of strongly–lensed galaxies. Finally, we find five new positions for *IRAS*–FSC/IIFSCz sources based on our LR analysis and higher-resolution PACS and SPIRE data.

The UV/optical/near–infrared identifications to the 250 μm–selected sample, as well as their photometric and spectroscopic redshifts, are available for download from the *Herschel*–ATLAS webpage; `http://www.h-atlas.org`.

ACKNOWLEDGMENTS

Herschel is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia with significant participation from NASA. U.S. participants in *Herschel*–ATLAS acknowledge support provided by NASA through a contract issued from JPL. GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programs including *GALEX* MIS, VST KIDS, VISTA VIKING, *WISE*, *Herschel*–ATLAS, GMRT and ASKAP providing UV to radio coverage. The GAMA website is: `http://www.gama-survey.org/`. This work used data from the UKIDSS DR5 and the SDSS DR7. The UKIDSS project is defined in Lawrence et al. (2007) and uses the UKIRT Wide Field Camera (WFCAM; Casali et al. 2007). Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, The National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The Italian group acknowledges partial financial support from ASI contract I/009/10/0 'COFIS'.

REFERENCES

Abazajian K.N., et al., 2009, ApJS, 182, 543
Amblard et al., 2010, A&A, 518, 9
Arp H., 1967, ApJ, 148, 321
Baldry I.K., 2010, MNRAS, 404, 86
Becker R.H., White R.L. & Helfand D.J., 1995, ApJ, 459, 559
Biggs A.D., et al., 2010, arXiv:1012.0305
Blain A.W., 1996, MNRAS, 283, 1340
Cannon R., et al., 2006, MNRAS, 372, 425
Chapin E.L. et al., 2010, arXiv:1003.2647
Chapman, S.C., Blain A.W., Ivison R.J., Smail I., 2003, Nature, 422, 695
Ciliegi P., et al., 2003, A&A, 398, 901
Ciliegi P., et al., 2005, A&A, 441, 879
Clements D.L., et al., 1996, MNRAS, 279, 459
Clements D.L., et al., 2004, MNRAS, 351, 447
Clements D.L., et al., 2010, A&A, 518, 8
Collister A.A., Lahav O., 2004, PASP, 116, 345
Condon J.J., Cotton W.D., Greisen E.W., Yin Q.F., Perley R.A., Taylor G.B., & Broderick J.J. 1998, AJ, 115, 1693
Connolly A.J., et al., 2002, ApJ, 579, 42
Croom S.M., et al., 2009, MNRAS, 392, 19
Davis M., et al., 2003, SPIE, 4834, 161
Davis M., et al., 2007, ApJL, 660, 1
Downes A.J.B., Peacock J.A., Savage A., & Carrie D.R., 1986, MNRAS, 218, 31
Driver S., et al., 2010, arXiv:1009.0614
Dunlop, J.S., 2010, MNRAS, 408, 2022
Duval V.G., Irace W.R., Mainzer A.K., Wright E.L., 2004, SPIE, 5487, 101
Dye S., et al., 2009, ApJ, 703, 285
Dye S., et al., 2010, A&A, 518, 10
Eales S., et al., 2009, ApJ, 707, 1779
Eales S., et al., 2010, PASP, 122, 499
Fanaroff B.L. & Riley J.M., 1974, MNRAS, 167, 31
Griffin, M., Abergel A., Abreu A., et al. 2010, A&A, 518, L3
Hill D.T., et al., 2011, MNRAS, arXiv:1009.0615
Ibar E., et al., 2010, MNRAS, 409, 38
Irwin M.J., 1985, MNRAS, 214, 575
Ivison R., et al., 2007, MNRAS, 380, 199
Ivison R., et al., 2008, MNRAS, 390, 1117
Jones D.H., et al., 2009, MNRAS, 399, 683
Lagache, G., et al., 2004, ApJS, 154, 112
Lawrence A., et al., 2007, MNRAS, 379, 1599
Lilly S.J., Eales S.A., Gear W.K.P., Hammer F., Le Fèvre O., Crampton D., Bond J.R., Dunne L., 1999, ApJ, 518, 641
Lilly S., et al., 2007, ApJS, 172, 70
Maddox S.J., et al., 2010, A&A, 518, 11
Martin D.C., Fanson J., Schiminovich D., et al., 2005, ApJ, 619, 1
Moshir M., et al., 1992, Explanatory Supplement to the IRAS Faint Source Survey, Version 2., JPL D-10015 8/92 (Pasadena JPL)
Munn J.A. et al., 2004, AJ, 127, 3034
Negrello M., Perrotta F., González-Nuevo J., Silva L., de Zotti G., Granato G.L., Baccigalupi C. & Danese L., 2007, MNRAS, 377, 1557
Negrello M., et al., 2010, Sci, 330, 800
Oyaizu, H., Lima M., Cunha C.E., Lin H., Frieman J., Sheldon E.S., 2008, ApJ, 674, 768
Pascale et al., 2010, arXiv:1010.5782
Pilbratt, G.L., Riedinger J.R., Passvogel T. et al., 2010, A&A, 518, L1
Poglitsch, A., Waelkens C., Gels N. et al. 2010, A&A, 518, L2
Richter, 1975, Astron. Nachrichten, 296, 65
Rigby E.E., et al., 2010, arXiv:1010.5787
Robotham A., 2010, PASA, 27, 76
Roseboom I.G., Oliver S., Parkinson D., Vaccari M., 2009, MNRAS, 400, 1062
Serjeant S., et al., 2003, MNRAS, 334, 887
Smith D.J.B., Simpson C., Swinbank A.M., Rawlings S.G. & Jarvis M.J., 2010a, MNRAS, 404, 1089

Sutherland W.J., Maddox S.J., Saunders W., McMahon R.G., Loveday J., 1991, MNRAS, 248, 483
Sutherland R., & Saunders R., 1992, MNRAS, 259, 413
Sutherland, W., 2009, "Science with the VLT in the ELT Era", ed A. Moorwood, Astrophysics & Space Science Procs, Springer, 171
Swinbank A.M., et al., 2010, Nature, 464, 733
Thompson M. et al., 2010, A&A, 518, 134
Wang L. & Rowan-Robinson M., 2009, MNRAS, 398, 109
Wilman R.J., 2008, MNRAS, 388, 1335
Wilman R.J., Jarvis M.J., Mauch T., Rawlings S., Hickey S., 2010, MNRAS, 405, 447
York et al., 2000, AJ, 120, 1579
Zhu M., Gao Y., Seaquist E.R., Dunne L., 2007, AJ, 134

This paper has been typeset from a TEX/ LATEX file prepared by the author.

APPENDIX A: CHECKING THE IDENTIFICATION PROCESS

A1 *IRAS* sources

In the 9hr field SDP region, there are a total of 35 detections from the Imperial *IRAS*–FSC Redshift Catalogue (IIFSCz, Wang & Rowan-Robinson, 2009, building on the *IRAS* Faint Source Catalogue of Moshir et al. 1992), the majority with associated optical/near–infrared positions of high reliability. By matching our 250 μm selected catalogue with the IIFSCz, and comparing the results to our LR analyses, we can provide a first check on the accuracy of our associations. There are 30 sources in the IIFSCz that have catalogue positions within 10.0 arcsec of the 250 μm source positions. Each of the positions in the IIFSCz catalogue for these sources matches an SDSS DR7 position within 2 arcsec, and has reliability $R > 0.80$.

There remain five IIFSCz sources for which we do not recover SDSS/SPIRE matches within 10.0 arcsec. In Figure A1, we show greyscale images of the *Herschel*–ATLAS PACS 100 μm observations of the regions surrounding these five IIFSCz sources, centred on the quoted IIFSCz catalogue positions. It is clear that each IIFSCz source has a bright PACS detection less than one arcminute away, with the PACS 100 μm observations being of considerably higher sensitivity and resolution than that of *IRAS* at 60 μm (this is the band on which the IIFSCz is selected).

Here we discuss each of these sources individually.

- F08555+0145: This source has an IIFSCz position derived from the SDSS DR6, residing approximately 40 arcsec away from the *IRAS* centroid. The bright PACS/SPIRE source within the 1σ positional errors of the IRAS centre is associated with an r = 17.9 mag galaxy approximately 1 arcsec away, which is not in the SDSS primary photometry catalogue. It is clear that this is the correct association, with a reliability based on its newly–measured magnitude and separation of $R = 0.999$, and that this source was mis-identified in the IIFSCz.
- F08598-0103: IIFSCz contains only an IRAS–derived position for this source in the absence of any counterparts detected in the ancillary data available at the time. Using our higher-resolution PACS/SPIRE observations, we are able to identify the optical counterpart, approximately 30 arcsec away from the IIFSCz position, with $R = 0.999$.
- F08599+0139: The IIFSCz position for this source is also derived from the SDSS DR6, suggesting a source approximately 10 arcsec to the North of the IRAS position. *Herschel*–ATLAS data, however, reveal a bright sub-millimetre source $\sim$30 arcsec to the East, associated with an r–band counterpart at $R = 0.994$. This source was therefore also mis-identified in the IIFSCz catalogue.
- F09009-0054: The IIFSCz catalogue position for this source was derived using data from the NRAO VLA Sky Survey (NVSS, Condon et al., 1998), and resides within the extended stellar halo of the $z = 0.04$ galaxy 2MASX J09033081-0106127 in the SDSS/UKIDSS–LAS data (which is also detected in each of the PACS and SPIRE bands). The higher resolution of the SDSS DR7 catalogue compared with the NVSS data and *IRAS* positions enables us to derive a more accurate position for the counterpart to this source, with $R = 0.999$.
- F09047-0040: The PACS/SPIRE detection of this source is located approximately 40 arcsec away from the IRAS position quoted in the catalogue. We identify an SDSS DR7 optical counterpart with a more accurate position, and reliability $R = 0.999$.

Table A1 contains new positions for our reliable counterparts to these *IRAS* sources. Assuming that our new identifications to the *IRAS* sources are correct, we recover reliable counterparts (with accurate positions) to all of the IIFSCz sources at $R \geqslant 0.8$, as compared with $\sim$89 percent (31/35) of sources for the IIFSCz itself (we exclude the two sources with clearly mis–identified counterparts, and also the two sources with no identified counterparts).

A2 A comparison with radio observations

We also compared the results of our likelihood ratio analysis to data from the Faint Images of the Radio Sky at Twenty centimetres (FIRST) Survey (Becker, White & Helfand, 1995). The FIRST survey covers 9,000 square degrees of sky with a resolution of 5 arcsec, with a source density of approximately 90 per square degree brighter than the detection threshold of $1\,mJy$. At these relatively bright flux limits, the source population is dominated by Active Galactic Nuclei (AGN) rather than star–forming sources (e.g. Wilman et al., 2008); as a result the overlapping population of sources between the *Herschel*–ATLAS and FIRST catalogues is not expected to dominate the number counts.

To make the comparison between our LR analysis and FIRST sources, we used the frequentist identification procedure of Downes et al. (1986), commonly used to quantify the formal significance of possible counterparts to sub-millimetre galaxies in radio survey data (e.g. Lilly et al., 1999, Ivison et al. 2007). In this procedure, the statistic used to assess the probability that a nearby radio source is *not* associated with the SPIRE source is $S = \pi r^2 \times n(>F)$, where r is the angular distance between the SPIRE source and the radio source, F is the flux density of the radio source, and $n(>F)$ is the surface density of radio sources with flux densities greater than this. For each SPIRE source, we looked for radio sources in the FIRST catalogue within 10.0 arcsec, and treated the radio source with the lowest value of S ($S_{\rm min}$) as the one most likely to be associated with the SPIRE source. We used a Monte-Carlo simulation (e.g. Eales et al. 2009) to determine the probability distribution of $S_{\rm min}$ on the null hypothesis that there are no genuine associations between radio sources and SPIRE sources.

We then used this probability distribution to determine the probability that each measured value of $S_{\rm min}$ would have occured by chance. We call this probability P'. Of the 6621 5σ 250 μm SPIRE sources, 105 have radio counterparts within 10.0 arcsec, all with values of $P' < 0.002$. However, this does not take account of the fact that with such a large sample of SPIRE sources one

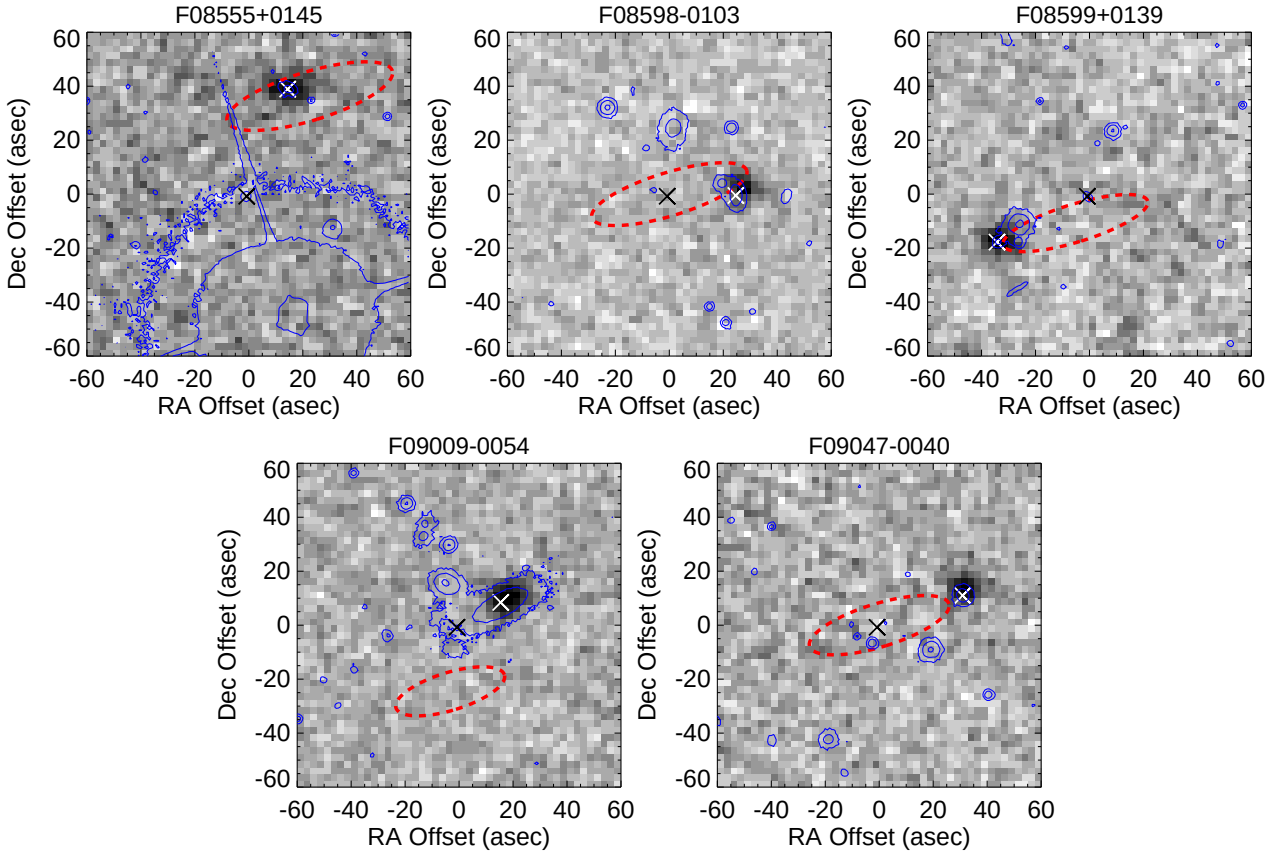


Figure A1. PACS 100 μm greyscale cutout images showing the regions surrounding the five IIFSCz catalogue positions for which we do not find a match within 10.0 arcsec in our SDSS/SPIRE catalogue. The IIFSCz catalogue positions are denoted by a black cross (2 derived from SDSS positions, 1 from NVSS and 2 from the original FSC), with the SDSS DR7 r–band contours overlaid in blue and the *IRAS*–FSC 1σ error ellipse overlaid in dashed red. The white crosses denote the positions of the $R \geqslant 0.8$ SDSS DR7 counterparts from our likelihood ratio analysis. These sources are discussed in more detail in section A1. Using our SDSS DR7 likelihood ratio analysis and the higher-resolution SPIRE 250 μm positions as our starting point, we are able to derive $R \geqslant 0.8$ counterparts for four of the five IIFSCz sources, positions of which are given in Table A1. The exception is F08555+0145, for which the bright galaxy approximately centred on the PACS 100 μm source is not present in the SDSS DR7 primary photometry catalogue (however we include a manually-measured position in Table A1).

Table A1. Updated positions of the IIFSCz sources, which were previously mis-identified, or identified with only NVSS/*IRAS* positions in Wang & Rowan-Robinson (2009). The position angles of the *IRAS* positional error ellipses were orientated 107° East of North.

	New positions				IIFSCz catalogue positions			*IRAS*–FSC catalogue positions			
IIFSCz ID	RA	Dec	source	R	RA	Dec	source	RA	Dec	$\sigma_{\rm maj}$	$\sigma_{\rm min}$
F08555+0145	134.535	1.5649	This paper	0.999	134.539	1.5538	SDSS DR6	134.533	1.5639	32″	9″
F08598-0103	135.602	−1.2622	SDSS DR7	0.999	135.609	−1.2623	*IRAS*–FSC	135.609	−1.2623	30″	8″
F08599+0139	135.636	1.4582	SDSS DR7	0.994	135.627	1.4629	SDSS DR6	135.628	1.4599	28″	7″
F09009-0054	135.879	−1.1033	SDSS DR7	0.994	135.883	−1.1059	NVSS	135.884	−1.1127	21″	7″
F09047-0040	136.829	−0.8693	SDSS DR7	0.999	136.838	−0.8726	*IRAS*–FSC	136.838	−0.8726	27″	8″

expects to find some low values of P' even if there were no genuine associations between the SPIRE sources and FIRST objects. We used a Monte-Carlo simulation to determine that 15 of the 105 associations are likely to be spurious. To correct for this, we calculated a new probability for each association, $P = \alpha P'$, where α is a constant that we calculated using $\sum_i \alpha P'_i = 15$. We took the conservative decision to treat associations with $P < 0.2$ as counterparts which are likely to be genuine, which rejected 29 of the original 105 associations.

There were a total of 76 SPIRE sources with $P < 0.20$ counterparts, and each of these was scrutinised using the FIRST and SDSS images displayed side–by–side with the Downes et al. and LR analysis overlaid. In this manner, we compared the results of the two independant identification methods. In forty-two cases, the $P < 0.20$ radio counterpart is also identified as having $R \geqslant 0.80$ in the r–band data, and the two methods choose the same counterpart.

There are thirty SPIRE sources with high quality ($P < 0.20$) FIRST counterparts which we do not recover in our LR analysis, including twenty–three SPIRE sources which do not have any r–band counterparts in our SDSS DR7 data (presumably distant, optically–

faint radio sources). Of the remaining seven sources with $P < 0.20$ FIRST counterparts:

- Four counterparts are detected in the optical data but have low reliabilities due to their faint magnitudes, or large separations in comparison to the value of $\sigma_{\rm pos}$ derived based on the 250 μm source SNR.
- Two sources have multiple, possibly interacting components with $L > 10.0$ but $R < 0.8$, only one of which is a radio source (these sources are discussed in section 4.3).
- In one further instance, the radio source has a double-lobed structure (a so-called FR-II, following Fanaroff & Riley, 1974), not coincident with either the dust emission or the starlight in the plane of the sky. The lobes of this FRII are extremely bright; as a result, the P statistic suggests that there is a low probability of a chance association, even though the separation between the SPIRE position and the FIRST centroid is large. The LR technique identifies the apparent host galaxy – aligned at the centre, between the two luminous radio jets – as having $L = 0.0$ due to its large separation (~ 9 arcsec) from the SPIRE centroid; this is an example of the limitations of the Downes et al. method.

However, these possibilities do not contaminate the 250 μm selected sample with incorrect associations. There are however, four instances where distinct counterparts have $P > 0.20$ and $R \geqslant 0.80$; here, the opposite is potentially true and the two methods conflict. These sources have derived reliabilities of 0.87, 0.98, 0.81 and 0.93 as compared with distinct Downes et al. counterparts with P statistics of 0.08, 0.07, 0.02 and 0.19, respectively. These sources are shown in Figure A2, in which the 10.0 arcsec search radius centred on the 250 μm position is shown in red, any unreliable optical counterparts in black, the reliable optical ID in light blue, and the radio contours overlaid in royal blue. In two of the four cases, the additional sources implied by the radio data are visible in K_S–band observations from VIKING (Sutherland, 2009), indicating that these sources are not merely effects of the larger positional uncertainty in FIRST as compared with SDSS. Futhermore, three of the four sources have SPIRE colours $S_{250}/S_{350} \leqslant 1.5$, suggesting high redshifts ($z > 1$) or cold dust temperatures, with the former being at odds with the photometric redshifts of their most reliable counterparts ($z < 0.55$). Sources with similar SPIRE colours and low-redshift counterparts are discussed in more detail in section 4.1.

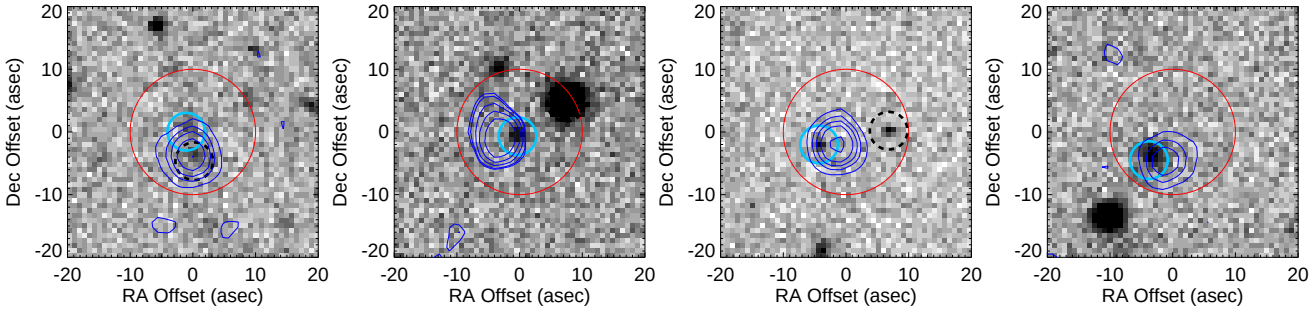
Finally, we note that probabilistic arguments such as those discussed here will inevitably present apparent disagreements for a small number of sources within large samples. In the remaining 101 out of 105 cases however, the results of our LR analysis are consistent with those using the FIRST catalogue and the P statistic, and crucially we recover an additonal 2,348 counterparts, compared with 31 extra counterparts to the 250 μm sources obtained by using only the radio data.

A3 *Spitzer* observations

An additional check on the identification process was conducted by searching for mid–infrared data from the *Spitzer Space Telescope* heritage archive, in order to compare the reliabilities from our r–band catalogue with near- and mid–infrared images between 3.6 and 160 μm. Four sets of observations were found which overlapped with the *Herschel*–ATLAS SDP observations These data can be used to examine the regions surrounding the SPIRE IDs for additional sources which may not be present in the r–band catalogue used for the identification process, as a visual check on the effectiveness of the LR technique.

There are a total of 49 sources that have *Spitzer* data, and although these data vary in sensitivity, there is no evidence that would suggest a mis-identification from the r–band catalogue. Such indications of wrong IDs would include reliable ($R \geqslant 0.8$) r–band counterparts indicated for SPIRE sources which have previously unrevealed bright *Spitzer* sources nearer to the centre of the SPIRE centroid. Indeed, in one case in particular (H–ATLAS J090913.2+012111), the sensitive IRAC data reveal the power of the LR technique. Although there are three potential counterparts in the SDSS DR7 r–band catalogue all within 6 arcsec of the SPIRE centroid, they have all been given low reliability ($R \leqslant 0.30$, and also $L \leqslant 0.20$). The IRAC 3.6 μm data reveal a fourth candidate counterpart within 1 arcsec of the SPIRE position, which is presumably the true counterpart. The r–band and IRAC 3.6 μm data are presented in Figure A3, with the various source positions overlaid to demonstrate the robustness of the LR method for this particular source, but also the need for longer-wavelength observations in order to be able to reliably identify the counterparts to higher redshift sources. The forthcoming data from the VISTA Kilo-degree INfrared Galaxy (VIKING) survey and from the Wide-field Infrared Survey Explorer (*WISE* – Duval et al., 2004) satellite will enable this.

This example also highlights one crucial advantage of using the LR technique for *Herschel* surveys rather than opting simply for the Downes method; the LR method takes into account the fact that not every source has a counterpart that is brighter than the detection limit in ancillary survey data.



(a) $P = 0.08, R = 0.87$ (b) $P = 0.07, R = 0.98$ (c) $P = 0.02, R = 0.81$ (d) $P = 0.19, R = 0.93$

Figure A2. Cases in which the LR method applied to the SDSS DR7 data and the P–statistic (Downes et al. 1986) method applied to the FIRST data produce different robust counterparts to 250 μm sources. In each panel, the 10.0 arcsec search radius around the 250 μm position is shown in red, with any unreliable counterparts circled in black. Reliable counterparts from the LR analysis are circled in light blue, while the royal blue contours reveal the FIRST counterpart. The P statistic for the FIRST source and the value of the reliability, R, of the most reliable SDSS DR7 counterpart are given in the subfigure captions for each 250 μm object. Thumbnail images are orientated such that North is up and East is to the left.

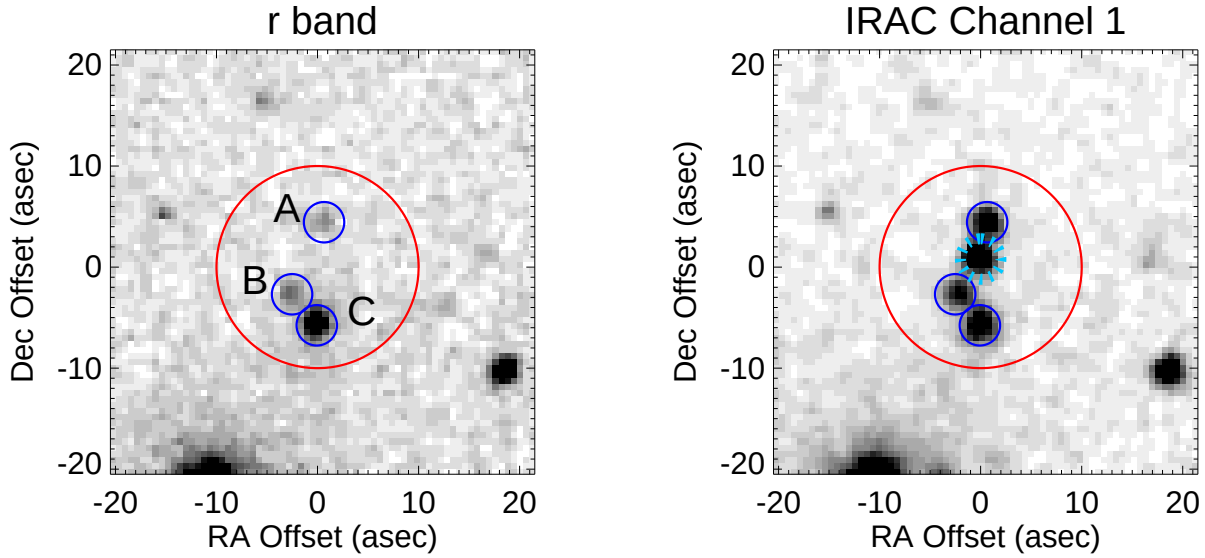


Figure A3. SDSS r–band (left) and *Spitzer Space Telescope* IRAC 3.6 μm image of the region surrounding source H–ATLAS J090913.2+012111, at α =137.305, δ = 1.3532 (position shown by the red circle, which has a radius of 10.0 arcsec). The r–band image contains three sources (blue 2 arcsec circles) that are identified as potential counterparts to the SPIRE source, with reliabilities of 0.00, 0.29, and 0.00 (and $L = 0.00, 0.17$ and 0.00) for the sources labelled A, B and C, respectively. The IRAC 3.6 μm channel image (right) reveals an additional source within 1 arcsec of the SPIRE centroid (dotted light blue 2 arcsec radius circle). The low reliabilities associated with the r–band detections indicates the power of the LR technique in this context. Images are orientated such that North is up and East is to the left.