Mon. Not. R. Astron. Soc. **398,** 109–118 (2009) doi:10.1111/j.1365-2966.2009.15138.x

The Imperial *IRAS*-FSC Redshift Catalogue

Lingyu Wang★ and Michael Rowan-Robinson

Astrophysics Group, Blackett Laboratory, Imperial College of Science Technology and Medicine, London SW7 2BZ

Accepted 2009 May 22. Received 2009 April 23; in original form 2008 September 9

ABSTRACT

We present a new catalogue, the Imperial *IRAS*-FSC Redshift Catalogue (IIFSCz), of 60 303 galaxies selected at 60 μm from the *IRAS Faint Source Catalogue* (FSC). The IIFSCz consists of accurate position, optical, near-infrared and/or radio identifications, spectroscopic redshift (if available) or photometric redshift (if possible), predicted far-infrared (FIR) and submillimetre (submm) fluxes ranging from 12 to 1380 μm based upon the best-fitting infrared template. About 55 per cent of the galaxies in the IIFSCz have spectroscopic redshifts, and a further 20 per cent have photometric redshifts obtained through either the training set or the template-fitting method. For $S(60) > 0.36$ Jy, the 90 per cent completeness limit of the FSC, 90 per cent of the sources have either spectroscopic or photometric redshifts. Scientific applications of the IIFSCz include validation of current and forthcoming infrared and submm/mm surveys such as *AKARI*, *Planck* and *Herschel*, follow-up studies of rare source populations, large-scale structure and galaxy bias, local multiwavelength luminosity functions and source counts. The catalogue is publicly available at http://astro.imperial.ac.uk/∼mrr/fss/.

Key words: catalogues – surveys – galaxies: distances and redshifts – quasars: general – large-scale structure of Universe – infrared: galaxies.

1 INTRODUCTION

The *IRAS Faint Source Catalog* (FSC; Moshir et al. 1992) contains 173 044 sources extracted from image plates of co-added data. It is 2–2.5 times deeper than the *IRAS Point Source Catalogue* (PSC), reaching a depth of ∼0.2 Jy at 12, 25 and 60 μm. The sky coverage of the FSC is limited to $|b| > 20^\circ$ in unconfused regions at 60 μm. For sources with high-quality flux density,[1] the minimum reliability of the whole catalogue is ∼99 per cent at 12 and 25 μm and ∼94 per cent at 60 μm. Around 41 per cent of the sources are detected at 60 μm (FQUAL = 2 or 3 at 60 μm).

The construction of a redshift catalogue of the FSC 60 μm sources is made possible by overlaps (in terms of depth and area) with various surveys either spectroscopic or photometric, such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Two Micron All Sky Survey (2MASS; Skrutskie et al. 1997) and the 6dF Galaxy Survey (Jones et al. 2004, 2005). We use two photometric redshift techniques, the empirical training set method and the spectral energy distributions (SED) fitting procedure, to provide estimates of redshifts from optical and near-infrared (NIR) broad-band photometry.

Other recent or planned all-sky surveys include *AKARI* and *Planck*. The *AKARI* (previously known as ASTRO-F) All-Sky Survey, which ended on 2007 August 26, has observed 94 per cent of the sky from mid- to far-infrared (FIR) (Murakami et al. 2007). The 5σ point source detection limit of the Far-Infrared Surveyor with a single scan coverage is estimated to be 2.4, 0.55, 1.4 and 6.3 Jy at 65, 90, 140 and 160 μm, respectively (Kawada et al. 2007; Wang et al. 2008). The European Space Agency (ESA) mission *Planck* is going to map the cosmic microwave background (CMB) with unprecedented angular resolution and sensitivity and produce all-sky catalogues of infrared and radio galaxies in the frequency bands ranging from 30 to 850 GHz. The study of these *Planck* extragalactic point sources can not only help clean up the foreground contaminants in the CMB images but also constrain galaxy formation and evolution models and allow searches for high-redshift dusty galaxies. For the High Frequency Instrument (HFI) channels, the estimated *Planck* All Sky Survey sensitivity at the 3σ level is 26, 37, 75, 180, 300 mJy at 2100, 1380, 850, 550, 350 μm, respectively [*Planck* Bluebook, ESA-SCI(2005)-1, Version 2] and thus many of the sources detected by *Planck* will also be present in the *IRAS* catalogues. Another ESA mission, *Herschel* (Pilbratt 2004), which is to be launched together with *Planck*, will observe the Universe in the FIR and submillimetre (submm) range (approximately 57–670 μm) with the formation and evolution of galaxies and stars and stellar systems as its key science goals. The SPIRE instrument (Griffin et al. 2007) will perform imaging in the broad-band photometry mode centred at 250, 350 and 500 μm with the predicted point source sensitivity in the range 8–11 mJy (5σ, 1 h). To aid these missions, we have provided predicted fluxes at the relevant mission wavelengths. The Imperial *IRAS*-FSC Redshift Catalogue (IIFSCz) will be an important

★E-mail: lingyu.wang05@imperial.ac.uk

[1] In the *IRAS* FSC, the flux density quality (FQUAL) is classified as high (=3), moderate (=2) or upper limit (=1).

input and validation catalogue for current and forthcoming wide-area infrared/submm surveys described above.

There are also several extragalactic science programmes that can be carried out with the IIFSCz. The convergence of the cosmological dipole is particularly worth further investigation as previous measurements with all-sky surveys such as *IRAS* or 2MASS have not yet shown a consensus on the convergence depth (Rowan-Robinson et al. 2000; Maller et al. 2003; Erdoğdu et al. 2006). While 2MASS samples the local Universe (within 200 h^{-1} Mpc) very well, it is not quite deep enough to detect potential contributions to the dipole signal from large distance. The IIFSCz provides a huge sample for large-scale structure and velocity studies, e.g. the baryon acoustic oscillation (Eisenstein et al. 2005), the two-point correlation function and its dependence on the FIR luminosity and star formation rate (Mann, Saunders & Taylor 1996; Szapudi et al. 2000; Hawkins et al. 2001), the local multiwavelength luminosity functions (Serjeant & Harrison 2005), follow-up studies of rare source populations such as ultraluminous infrared galaxies (ULIRGs) and hyperluminous infrared galaxies (HLIRGs) and number counts.

The layout of this paper is as follows. The criteria used to select galaxies from the *IRAS* FSC are described in Section 2.1. The issue of extended sources is discussed in Section 2.2. In Section 3, we first obtain spectroscopic redshifts from a number of data bases and then cross-identify (using the likelihood ratio technique) FSC sources with their optical, near-infrared and/or radio counterparts. In Section 4, we carry out photometric redshift estimation (using both the training set and the template-fitting method) and then make flux predictions at FIR and submm wavelengths. The overall properties of the IIFSCz are described in Section 5. Finally, discussions and conclusions are given in Sections 6 and 7, respectively. We adopt a flat cosmological model with $\Lambda = 0.7$ and $h_0 = 0.72$.

2 CATALOGUE CONSTRUCTION

2.1 Sample selection

To obtain a complete sample of galaxies from the *IRAS* FSC, our selection criteria are as follows.

(1) To ensure reliability, we select sources with FQUAL $\geq$3 and signal-to-noise ratio (SNR) >5 at 60 μm. There are two types of SNR in the *IRAS* FSC, the SNR at a given pixel (LOCSNR) and the SNR which uses a noise value derived from a local region. The latter is adopted here. This condition leaves us with 63 842 sources.

(2) To exclude cirrus, we require log(S100/S60) < 0.8 if FQUAL $\geq$ 2 at 100 μm. The sample size is reduced to 63 117. The upper panel in Fig. 1 shows the predicted 100–60 μm colour as a function of redshift based upon four infrared templates, cirrus, M82 starburst, Arp 220 starburst and active galactic nuclei (AGN) dust torus (Rowan-Robinson et al. 2004, 2008). It is clear that our constraint on the 60–100 μm flux ratio should not exclude any infrared galaxy type for $z < 4$.

(3) To discriminate against stars, we first require log(S60/S25) > −0.3 if FQUAL $\geq$ 2 at 25 μm and then log(S60/S12) > 0 if FQUAL $\geq$2 at 12 μm. A total of 60 381 sources have met the above criteria, the faintest of which has a flux density of 0.12 Jy at 60 μm. Fig. 2 is the colour–colour diagram of FSC sources with detections at 12 and 25 μm after applying the 60–100 μm colour cut. The Rayleigh–Jeans predictions for the 12–60, 25–60, 25–12 μm flux ratio are the dotted lines, while our stellar rejection criteria are indicated by the solid lines. The concentrations of objects to the upper right are stars. Admittedly, our colour cuts might be a little harsh and therefore some low-redshift AGNs might be missing from our sample (see the middle and lower panels in Fig. 1). By examining Digitized Sky Survey (DSS) images and cross-identification in NASA/IPAC Extragalactic Data base (NED) or SIMBAD (if available) of sources

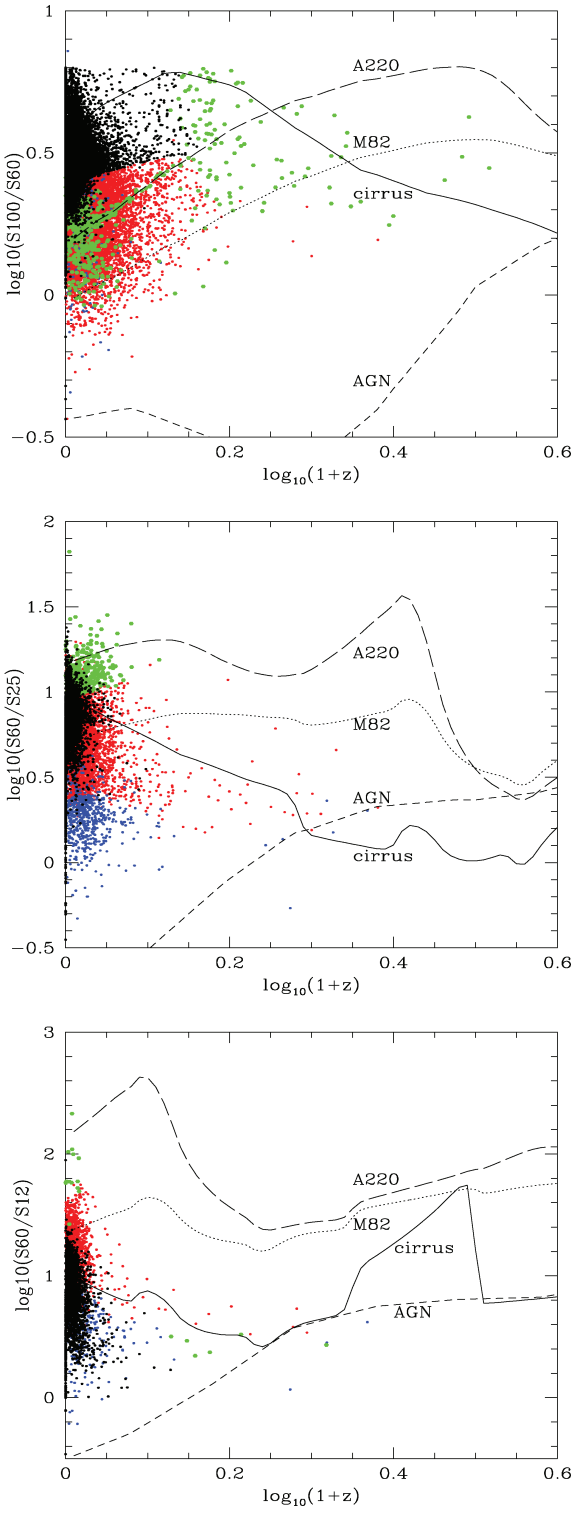

Figure 1. Upper panel: colour at 100–60 μm versus spectroscopic redshift colour-coded by four infrared templates (black: cirrus; red: M82; green: A220; blue: AGN dust torus). Middle panel: colour at 25–60 μm versus spectroscopic redshift. Lower panel: colour at 12–60 μm versus spectroscopic redshift.

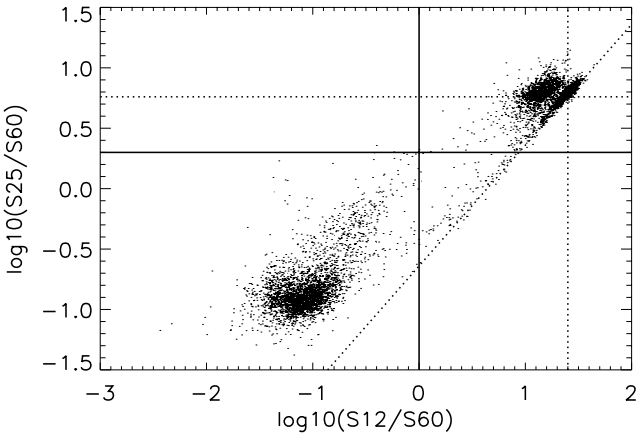

Figure 2. Colour–colour distribution of sources detected at 12 and 25 μm after applying the cirrus rejection criterion. The dotted lines are the Rayleigh–Jeans predictions, while the solid lines indicate our stellar rejection criteria. Sources with log 10(S25/S60) < 0.3 and log 10(S12/S60) < 0 are selected as galaxies.

in regions where $0 < \log(S12/S60) < 0.7$ and $\log(S25/S60) > 0$ or $0 < \log(S12/S60) < 0.3$ and $\log(S25/S60) < 0.3$, we managed to retrieve three Seyfert galaxies, eight unidentified or confused galaxies, while the rest are mostly carbon stars, post-AGB stars, etc. At this stage, our sample contains 60 392 galaxies which form the base catalogue of the IIFSCz. However, we point out that the sample size will undergo one more change in Section 3.1.

2.2 Extended sources

In the *IRAS* Faint Source Survey, the median filtering method was used to maximize the detection of faint point sources. However, the fluxes of extended sources are attenuated. This may cause nearby galaxies to be incorrectly excluded from our catalogue. In the construction of the PSCz catalogue, Saunders et al. (2000) used the ADDSCAN/SCANPI (Scan Processing and Integration) to derive fluxes for sources with blue-light isophotal major diameters $D_{25} >$ 2.25 arcmin. For sources with $D_{25} > 8$ arcmin, they used fluxes from the catalogue of *IRAS* observations of 85 large optical galaxies (Rice et al. 1988). There are a total of 1402 galaxies with blue isophotal diameters >2.25 arcmin in the PSCz, 1290 of which are associated with PSC sources. Of the 1290 PSC sources, 938 were found in the FSC after applying our first selection criterion, and so we have adopted their PSCz fluxes. Of the remaining 112 sources not found in the PSC, we identified 76 in the FSC and their fluxes were also replaced by PSCz fluxes. Some of the extended PSCz sources were not found in the FSC because they lie at $|b| < 20^\circ$.

ADDSCAN was designed for accurate flux measurement of bright sources and extended sources which need higher in-scan resolution. From a sample of 62 unambiguously extended galaxies in the Virgo Cluster, the ratio of the ADDSCAN flux to the FSC flux stays close to unity for sources fainter than ∼0.4 Jy at 60 μm (Moshir et al. 1992). Indeed, almost all FSC sources cross-matched with extended PSCz sources ($D_{25} > 2.25$ arcmin) have $S60 >$ 0.4 Jy. In order to estimate the likely number of FSC sources to have their fluxes seriously underestimated, we show the K_s band total magnitude diameter ($D_{\rm Ktot}$) for around 28 000 FSC sources with 2MASS identifications (see Section 3.3) in Fig. 3. The mean K_s total magnitude diameter for FSC sources associated with extended PSCz sources (red symbols in Fig. 3) is $D_{\rm Ktot} = 2.57 \pm 0.90$ arcmin. Therefore, the K_s total magnitude diameter $D_{\rm Ktot}$ is approximately

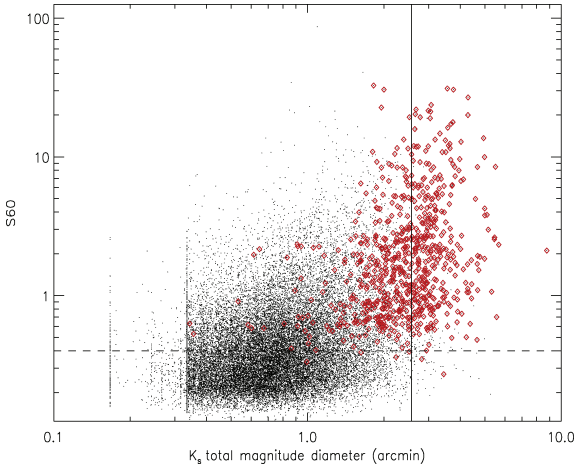

Figure 3. Flux densities at 60 μm versus K_s-band total magnitude diameter for around 28 000 FSC sources with 2MASS identifications. The red symbols are the FSC sources associated with extended PSCz sources (i.e. $D_{25} >$ 2.25 arcmin). The solid line shows the mean K_s total magnitude diameter $D_{\rm Ktot} = 2.57$ arcmin for the red symbols and the dashed line is where $S60 = 0.4$ Jy.

the same as the blue isophotal diameter D_{25} (Jarrett et al. 2003). In total, there are 1677, 430, 97 galaxies with $S60 > 0.4$ Jy and $D_{\rm Ktot} > 1.7$, 2.6, 3.5, arcmin, respectively. Excluding extended sources whose fluxes have already been replaced by PSCz fluxes, we conclude there are still ∼<1 per cent of the galaxies in our catalogue that suffer from flux underestimation.

3 SOURCE IDENTIFICATION

3.1 Redshift compilation from NED, FSSz, PSCz and 6dF

Having constructed the base catalogue, the next step is to obtain spectroscopic redshifts from past redshift surveys and the literature. The NED currently contains 10.4 million objects, 1.4 million redshifts and 16.0 million cross-identifications based upon astrometry, photometry and avoidance of confusion.

An NED all-sky query[2] of FSC sources at $|b| > 20^\circ$ with $S60 >$ 0.1 Jy returned 64 219 objects (we will refer to this sample as NED-FSC). In the NED, two constraints were used to select FSC sources, $\log(S60/S25) > -0.3$ and FQUAL ≥3 at 60 μm. Therefore, our sample is expected to be smaller than the NED-FSC. The NED position of a given FSC source is the best possible position, that is to say optical or near-infrared position if available and *IRAS* position if not. The accuracy of *IRAS* position depends on the size, brightness and SED of the source but is usually <20 arcsec (1σ). Within a radius of 300 arcsec, 60 320 out of 60 392 sources in our base catalogue are matched with sources in the NED-FSC, more than 23 000 of which have spectroscopic redshifts in the NED. A manual checking of the matched sources with separations ≥80 arcsec (∼170 objects) proved that these cross-identifications are correct. Fig. 4 shows the number of cross-identifications as a function of the angular separation between the *IRAS* position and the NED position. Three FSC sources are matched manually to their cross-identifications due to their large positional separations.

[2] An all-sky query in NED usually returns a list of sources with information such as source name, type, position and redshift. However, photometric data for each source are not included and have to be retrieved individually.

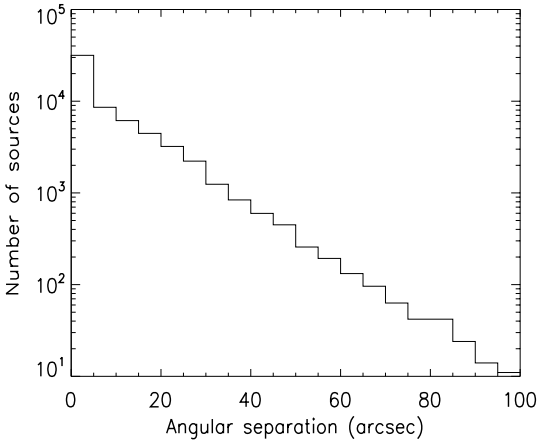

Figure 4. Distribution of the angular separations between the *IRAS* positions and the NED positions of the FSC sources.

44 FSC sources are not found in NED. In addition, we have removed a few FSC sources which are identified in NED as galactic objects from our base catalogue. This is the final change to the size of our base catalogue which now contains 60 303 galaxies.

The FSS redshift survey (FSSz; Oliver 1993) covering 700 deg^2 measured redshifts for 1546 FSC sources. It provides an additional 568 redshifts for our catalogue.

So far, around 2000 bright FSC sources (S60 $\geq$ 0.6 Jy) in our base catalogue do not have spectroscopic redshifts. The PSC Redshift Survey (PSCz; Saunders et al. 2000) is complete to S60 = 0.6 Jy. NED has not folded in the PSCz due to issues such as significant positional difference and incorrect NGC/IC/MCG cross-identifications. We managed to get 1972 spectroscopic redshifts from the PSCz via exact source name matching.

The 6dF Galaxy Survey covers roughly two-third of the southern sky, measuring redshifts for over 80 000 galaxies. The primary redshift sample is selected from the 2MASS Extended Source Catalog (XSC), together with 13 other samples such as the SuperCOSMOS catalogue, the *ROSAT* All-Sky Survey and the *IRAS* FSC. NED used a 3 arcsec radius in matching the 6dF with the FSC; however, matches were not made if there were more than one 6dF object within the *IRAS* beam. In total, NED has matched ∼6000 FSC sources with the 6dF observations.

In the 6dF data base, 10 111 objects appear in both the *IRAS* FSC and the SPECTRA table where all the observational and redshift-related information is stored. Again, by matching FSC source names, the 6dF provides an extra 2794 redshifts with acceptable quality (on quality scale $Q = 3$ or 4).

To summarize, we have obtained 29 037 spectroscopic redshifts from NED, FSSz, PSCz and 6dF, which comprises 48 per cent of our base catalogue. However, there are a few thousand more redshifts to be gained from the SDSS DR6 survey (see below).

3.2 CROSS-MATCH WITH SDSS

The SDSS photometric DR6 survey has covered 9583 deg^2 at u, g, r, i, z with magnitude limits 22.0, 22.2, 22.2, 21.3, 20.5, respectively. The spectroscopic DR6 survey has covered 7425 deg^2, the main samples of which are magnitude-limited at Petrosian $r < 17.77$ for galaxies.

As mentioned in Section 3.1, one does not get photometric data using the NED all-sky search. Therefore, in order to get optical magnitudes, we have to cross-identify FSC sources with their optical counterparts in the SDSS DR6 catalogues. In addition, because NED has only entered galaxies from the spectroscopic DR5 survey (covering 5740 deg^2) so far, we can also get new redshifts from the spectroscopic DR6 survey.

3.2.1 *The likelihood ratio technique*

We use the likelihood ratio (LR) technique (Wolstencroft et al. 1986; Sutherland & Saunders 1992; Ciliegi et al. 2003; Brusa et al. 2007) to cross-identify FSC sources with their optical counterparts. The LR technique essentially compares the probability of a candidate being the true counterpart as a function of magnitude m and separation d with that of a chance association, i.e.

$$L = \frac{p(m, d)\ \mathrm{d}m\ \mathrm{d}x\ \mathrm{d}y}{n(m)\ \mathrm{d}m\ \mathrm{d}x\ \mathrm{d}y} = \frac{q(m) f(d)}{n(m)}. \quad (1)$$

The probability distribution function $f(d)$ is usually assumed to be a two-dimensional Gaussian,

$$f(d) = \frac{1}{\sigma_1 \sigma_2 2\pi} \exp\left[-\frac{1}{2}\left(\frac{d_1^2}{\sigma_1^2} + \frac{d_2^2}{\sigma_2^2}\right)\right] = \frac{1}{\sigma_1 \sigma_2 2\pi} \exp\left(-\frac{1}{2} d^2\right), \quad (2)$$

where σ_1 and σ_2 are the axes of the positional uncertainty ellipse of a given FSC source, d_1 and d_2 are the positional separation along the axes and d is the normalized angular distance. For FSC sources, the mean 1σ positional uncertainty along the minor and major axes are 5 and 18 arcsec, respectively. The magnitude distribution function of background objects $n(m)$ is taken from objects around some random positions in the sky. The magnitude distribution function of the true optical counterparts $q(m)$ is obtained by subtracting $n(m)$ from the magnitude distribution of objects' each FSC sources. In Bayesian inference, it can be shown that L contains all the information about the probability of a candidate being the true counterpart. In cases where there are multiple candidates, the one with the highest likelihood is selected as the true counterpart.

3.2.2 *Obtaining redshifts from SDSS DR6*

First, we need to test the reliability of the LR method. This is easily achievable as we have already obtained 29 022 spectroscopic redshifts in Section 3.1 which can be used to compare with the redshifts of the optical counterparts found in the spectroscopic DR6 survey.

In the spectroscopic DR6 catalogue, we search for all primary objects[3] within 1 arcmin from the *IRAS* position of each FSC source with known spectroscopic redshift gained in Section 3.1. Although, in principle, we can use accurate positions for these FSC sources, the *IRAS* positions are used to investigate the impact of large positional uncertainty.

Of the 29 022 FSC sources with spectroscopic redshifts, 8979 were found to have optical counterparts, and, of these, an average of 1.2 optical counterparts were found for each. We exclude candidates with $r > 17.6$ (above the magnitude limit) and $z < 0.0005$ in order to exclude stars. Fig. 5 shows the r-band magnitude distribution of

[3] Whenever the SDSS makes multiple observations of the same object, the one with the best photometry will be assigned as the 'primary' observation.

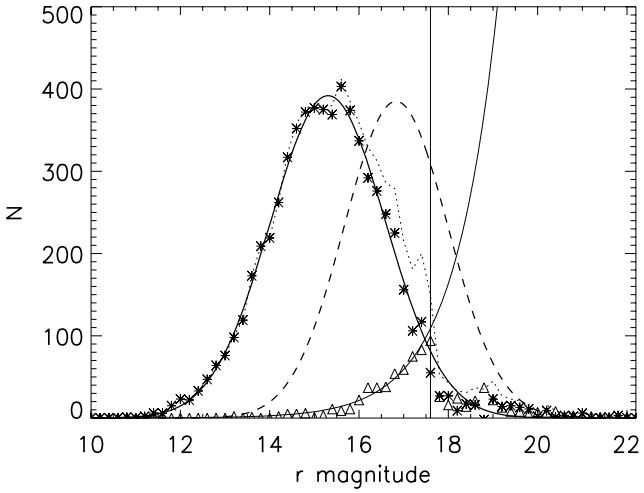

Figure 5. The *r* magnitude distribution of objects in the SDSS spectroscopic DR6 around FSC sources with redshifts obtained in Section 3.1 (black dotted line). The triangles show the magnitude distribution of random background objects, i.e. $n(m)$, and the thin solid curve is an exponential fit to these points. The asterisks depict the magnitude distribution of the true optical counterparts, i.e. $q(m)$, and the thick solid line is a Gaussian fit to these points. The dashed curve is a Gaussian fit to $q(m)$ for optical objects around FSC sources which do not gain redshifts in Section 3.1. The vertical line is where $r = 17.6$.

the true counterparts and that of the random background objects. The thick solid line is a Gaussian fit to $q(m)$,

$$q(m) \propto \exp\left[-\frac{1}{2}\left(\frac{m-15.31}{1.29}\right)^2\right] \quad (3)$$

and the thin solid curve is an exponential model of $n(m)$. Using the LR method, we managed to cross-match 8166 FSC sources with optical galaxies in the spectroscopic DR6 survey. For an FSC source, if its redshift obtained in Section 3.1 agrees with the redshift of its optical counterpart, then we say that the cross-identification is correct. Thus, the reliability, defined as the ratio of the number of correction identifications to the total number of sources in the cross-matched sample,

$$\text{Reliability} = \frac{N_{\text{correct}}}{N_{\text{total}}}, \quad (4)$$

of the cross-matched sample is estimated to be ∼97 per cent.

Now, we can apply the same procedure to the 31 260 FSC sources which did not receive spectroscopic redshifts in Section 3.1. Using a search radius of 1 arcmin, 4654 FSC sources are matched with 5462 optical candidates. The dashed curve in Fig. 5 is a Gaussian fit to the new $q(m)$, where the peak is now shifted to ∼16.8. The LR technique gives rise to new optical identifications and spectroscopic redshifts for 3844 FSC sources.

3.2.3 Obtaining photometry from SDSS DR6

Source identification becomes much more complicated in matching FSC sources with the photometric DR6 catalogue as the probability of chance association is $\propto n\pi d^2$, where n is the number density of background optical objects. Using a search radius of 1 arcmin, an average of 28 optical counterpart candidates were found for each FSC source. The LR approach produced 9425 cross-identifications between the sample of 31 260 FSC sources and the photometric DR6 catalogue. The adopted $q(m)$ is the blue dotted line in Fig. 5. SDSS objects classified as stellar are rejected not only because the contamination caused by stars is significant but also because the vast majority of the quasi-stellar objects (QSOs) already have redshifts.

To estimate the reliability of our cross-matched sample, first let $\mathbb{F}$ denote the set of 9425 FSC sources and $\mathbb{O}$ the set of optical counterparts. The reliability of $\mathbb{O}$ is simply the ratio of the number of true optical IDs in $\mathbb{O}$ to the size of $\mathbb{O}$. Consider a subset $\mathcal{F} \subset \mathbb{F}$ for which the true optical counterparts $\mathcal{T}$ are known. If the set of optical counterparts of $\mathcal{F}$ found using the LR method is denoted as $\mathcal{O} \subset \mathbb{O}$, the reliability of $\mathcal{O}$ is given by the ratio

$$\text{Reliability} = \frac{N_{\mathcal{O}\cap\mathcal{T}}}{N_{\mathcal{O}}}, \quad (5)$$

where the numerator is the size of the intersection between $\mathcal{O}$ and $\mathcal{T}$ and the denominator is the size of $\mathcal{O}$. Assuming the subset $\mathcal{F}$ is representative of $\mathbb{F}$, the reliability of the subset $\mathcal{O}$ should reflect the reliability of $\mathbb{O}$.

In order to get reliable optical identifications, accurate positions of the FSC sources are needed so as to reduce the probability of chance association. In the following, three test samples are used.

(i) We select 5000 FSC sources with 2MASS identifications, the positional uncertainty of which is around 1.25 arcsec at the 95 per cent confidence level. Using a search radius of 1 arcsec around the 2MASS positions, 1325 FSC sources (out of 5000) were matched with unique optical counterparts in the photometric DR6 survey (we will refer to this set of optical IDs as '2MASS_1arcsec'). Using '2MASS_1arcsec' as $\mathcal{T}$, the reliability of $\mathbb{O}$ is estimated to be 83 per cent. Excluding 'stellar' optical objects, the reliability is increased to 88 per cent.

(ii) The Faint Images of the Radio Sky at Twenty centimetres survey (FIRST; Becker et al. 1995) with a sensitivity of 1 mJy at 1.4 GHz and an angular resolution of 5 arcsec yielded a catalogue of ∼811 000 sources. The survey area overlaps with that of the SDSS and the radius of the 90 per cent confidence error circle is less than 1 arcsec. An average of 1.1 FIRST sources was found within 1 arcmin from an FSC source. We have selected 4886 unambiguous FSC–FIRST cross-matches, 3273 of which have unique optical counterparts in the photometric DR6 survey ('FSC–FIRST–SDSS'). Using 'FSC–FIRST–SDSS' as $\mathcal{T}$, the reliability of $\mathbb{O}$ is 80 per cent.

(iii) The sample of 3844 new optical identifications obtained in Section 3.2.2 serves as the third test sample and it gives us a reliability of 76 per cent.

Thus, on average the optical identifications obtained from the SDSS photometric DR6 are ∼80 per cent reliable.

3.3 CROSS-MATCH WITH 2MASS

The 2MASS has uniformly surveyed the whole sky in J, H and K_s, reaching a median depth of $z = 0.073$. To date, the 2MASS XSC has been comprehensively assimilated and cross-matched with other surveys in NED. Therefore, it is fairly straightforward to retrieve 28 640 cross-IDs between the IIFSCz and the 2MASS XSC, around 7000 of which do not have redshift information. We have also identified 39 FSC sources in the 2MASS PSC, 25 of which are shown to be QSOs.

3.4 CROSS-MATCH WITH NVSS

The NRAO VLA Sky Survey (NVSS; Condon et al. 1998) is a moderately deep radio survey over 82 per cent of the sky. It has

produced a catalogue of nearly 2 million sources brighter than 2.5 mJy at 1.4 GHz. The observed extragalactic sources include nearby normal galaxies, AGNs, star-forming galaxies, starbursts, etc. The rms positional uncertainties are $\leq$1 arcsec for sources brighter than 15 mJy and 7 arcsec for sources above the detection limit. Given the almost linear FIR to radio luminosity correlation (Helou, Soifer & Rowan-Robinson 1985; de Jong et al. 1985; Yun, Reddy & Condon 2001; Appleton et al. 2004), the NVSS is expected to detect most of the FSC sources and therefore provide much more accurate positions for these sources.

Unique radio counterparts were found for 23 183 FSC sources within a radius of 1 arcmin, and multiple counterparts were found in ~440 cases. Thus, the cross-identification accuracy is $\geq$98 per cent even if we adopt a crude nearest-object cross-matching approach. The NVSS provides cross-IDs for 3123 FSC sources without any observation other than *IRAS*.

4 PHOTOMETRIC REDSHIFT ESTIMATION

4.1 THE TRAINING SET METHOD

We use two different techniques, the empirical training set and the template-fitting technique, to estimate photometric redshifts for FSC sources with optical, NIR and/or radio photometry. In this section, we apply the public Artificial Neural Networks code of Firth, Lahav & Somerville (2003) and Collister & Lahav (2004) (*ANNz*) to the IIFSCz, and results are presented and discussed.

ANNz requires a representative training set[4] to learn the functional relationship between photometry and redshift. Advantages of the training set method include nullifying systematic effects, freedom and flexibility in choosing input parameters, greater accuracy and efficiency, etc. Having said that, the photometric redshift accuracy strongly depends on the quality of the training set. In principle, the training set should occupy the same region in the parameter space (colour, redshift, spectral type, etc.) as the testing set.

The structure of *ANNz* can be roughly divided into three passages, the input layer, the intervening/hidden layers and the output layer. In our case, it can be viewed as: magnitudes at various wavebands $\rightarrow$ functional mapping $\rightarrow$ photometric redshift. Each layer consists of a number of nodes. The network architecture is determined by the number of filters at the input layer, the number and size of each of the hidden layers and the number of outputs. For example, a 3.10.10.10.1 network architecture has photometric measurements in three filters, three hidden layers each of which has 10 nodes, and one output, i.e. the derived photometric redshift. The optimization of the mapping is achieved by tuning the network weights associated with connected nodes. The training process is terminated when the cost function, defined as $(z_{\rm phot} - z_{\rm spec})^2$, is minimal on the validation set. *ANNz* gives two types of errors, photometric noise and network variance. The latter is obtained by using a number of networks known as a committee.

For FSC sources with 2MASS counterparts, we choose the elliptical isophotal aperture based on the $K_{\rm s}$ 20 mag arcsec^{-2} isophote which gives a good estimate of the integrated flux and colour. The 2MASS training set contains 21 050 galaxies, 6664 of which are separated out to form a validation set. The 2MASS testing set contains 6771 galaxies. For FSC sources with SDSS counterparts, the SDSS model magnitudes obtained through the best-fitting model (either a pure deVaucouleurs or an exponential) in the r band is a good choice for accurate colours. The SDSS training set contains 10 096 galaxies, the testing set contains 5586 galaxies and the validation set contains 1856 galaxies. Extinction corrected is applied using the full-sky 100 μm maps (Schlegel, Finkbeiner & Davis 1998). The colour–colour distribution of the training (black plus signs) and testing set (red plus signs for the 2MASS cross-IDs and blue plus signs for the SDSS cross-IDs) is shown in Fig. 6. The SDSS training set is not as representative as the 2MASS training set. The photometric redshift error defined as

$$\sigma = \sqrt{\left\langle \left(\frac{z_{\rm phot} - z_{\rm spec}}{1 + z_{\rm spec}} \right)^2 \right\rangle} \qquad (6)$$

is 1.5 per cent for the 2MASS validation set and 2.3 per cent for the SDSS validation set (see Fig. 7).

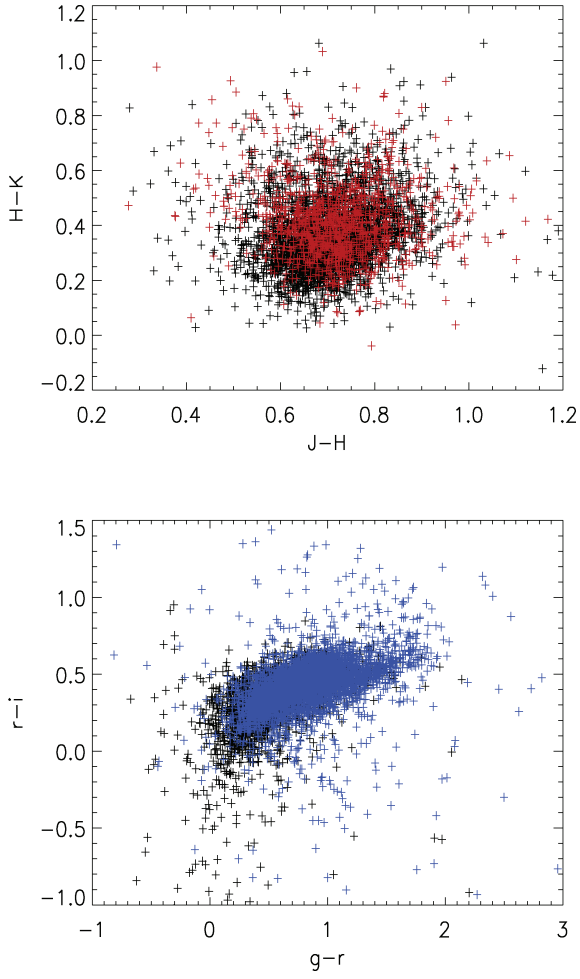

Figure 6. Colour–colour distribution of FSC sources with 2MASS (top) or SDSS cross-IDs (bottom).

4.2 THE TEMPLATE-FITTING TECHNIQUE

Another widely used photometric redshift estimation technique is the template-fitting method where the observed fluxes are compared to those from a library of SED templates (based upon observations or population synthesis models) which represent different galaxy

[4] A training set is a collection of sources with accurate redshifts and the same filter set as the testing set for which we would like to estimate the photometric redshifts. A validation set is a random collection of sources with known redshifts. However, the validation set is not used in the network training process.

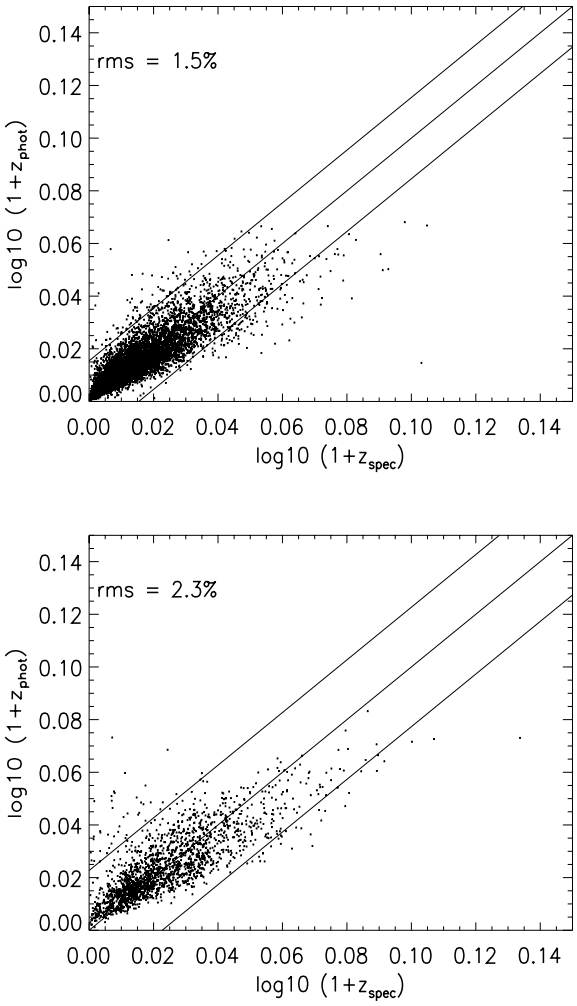

Figure 7. Upper panel: photometric redshift versus spectroscopic redshift, using three near-infrared bands and a 3.10.10.10.1 network architecture. Lower panel: photometric versus spectroscopic redshift using five optical bands and a 5.10.10.10.1 architecture.

populations. Unlike the training set method, a representative set of galaxies with known redshifts is not needed. Moreover, once the spectral type and redshift are determined through χ^2 minimization, we can carry on to predict fluxes at other wavelengths as well as determine useful parameters like the extinction, bolometric luminosity, stellar mass, etc.

Here, we apply the template-fitting method that has been used to construct the SWIRE Photometric Redshift Catalogue (Rowan-Robinson et al. 2008 and references therein) to the IIFSCz galaxies with either optical or near-infrared photometry. We use only a single pass through the data and use a resolution of 0.002 in $\log_{10}(1 + z)$. We use six galaxy templates (E, Sab, Sbc, Scd, Sdm and starburst) and three QSO templates, as in Rowan–Robinson et al. (2008, hereafter RR08). The resulting rms error in $(1 + z)$ for galaxies is 2.7 per cent for 2MASS sources (*JHK*) and 4.4 per cent for SDSS sources (*ugriz*), and the outlier rate is <0.1 per cent. Although the rms values are slightly worse than the neural network method, the template method does have the advantage that it is able to predict redshifts which lie outside the range of values in the training set. This is a definite issue for the SDSS sources and for these we have adopted, in the order of priority, (1) spectroscopic redshift, if available, (2) template method redshift, (3) neural network method redshift. For other sources (2MASS sources or NVSS sources), we have adopted (1) spectroscopic redshift, (2) neural network method redshift, (3) template method redshift.

We are able to determine a redshift for around 74 per cent of the sources in our catalogue. For these sources, we use the infrared template-fitting method of RR08 to fit four infrared templates (cirrus, M82 starburst, Arp 220 starburst, AGN dust torus) to the 12–100 μm *IRAS* data, using spectroscopic redshifts where available. Parameters of these infrared template fits are given in the catalogue, including the infrared luminosities in each component. These template fits are used to predict the fluxes at 12, 25, 60, 90, 100, 110, 140, 160, 250, 350, 500, 850, 1250 and 1380 μm, which cover the survey wavelengths of *AKARI*, *Planck*-Surveyor and *Herschel*. We do not attempt infrared template fits for sources with $z < 0.0003$ (essentially Local Group objects), since the FSS fluxes for these are likely to be serious underestimates. We believe the combination of precise optical, radio or near-infrared positions, redshifts and predicted submm fluxes will make this catalogue invaluable for future large-area FIR and submm surveys.

For the remaining ∼15 000 sources, we do not at the moment have optical identifications or good positions, but we note that most of those that fall outside the areas surveyed by SDSS will be similar to the SDSS sources.

5 CATALOGUE DESCRIPTIONS

The columns in the IIFSCz include source name (as appears in the *IRAS* FSC), position and flag (1 = SDSS, 2 = 2MASS, 3 = NVSS, 4 = NED, 5 = FSC and prioritized in the same order), *IRAS* fluxes and flux-quality flags, *ugrizJHK* magnitudes and errors, photometry flag (1 = 2MASS XSC, 2 = 2MASS PSC), integrated 1.4 GHz flux density and error, spectroscopic redshift and flag (1 = SDSS, 2 = PSCz, 3 = FSSz, 4 = 6dF, 5 = NED and prioritized in the same order), *ANNz* photometric redshift and error, template-fitting photometric redshift, adopted redshift, optical galaxy template type, extinction *Av* from optical galaxy template fit, reduced χ^2 for galaxy template fit, absolute *B* magnitude, optical bolometric luminosity, fraction of contribution at 60 μm of cirrus, M82, AGN and A220 infrared template, bolometric luminosity in cirrus, M82, AGN and A220 component, infrared luminosity, infrared template type (=1 for cirrus galaxies, =2 for M82 starbursts, =3 for A220 starbursts, =4 for AGN dust tori, i.e. $L_{\rm tor} > L_{\rm M82}$), reduced χ^2 for infrared template fit, predicted fluxes at 12, 25, 60, 90, 100, 110, 140, 160, 250, 350, 500, 850, 1250 and 1380 μm, other source names and types. The catalogue and description are available at http://astro.imperial.ac.uk/∼mrr/fss/.

In Fig. 8, the cumulative redshift completeness is plotted against the 60 μm flux. The vertical line shows the intrinsic 90 per cent completeness limit of the FSC, 0.36 Jy at 60 μm. Consequently, for research programmes such as the cosmological dipole where sample completeness is desired, the upper panel in Fig. 7 should be taken into account. The IIFSCz covers about 61 per cent of the whole sky. In Fig. 9, the sky distribution is plotted for all galaxies in the IIFSCz, galaxies with either spectroscopic or photometric redshift and galaxies without any redshift estimate.

To summarize, the IIFSCz contains a total of 60 303 galaxies, 55 per cent of which have spectroscopic redshifts from NED, FSSz, PSCz, 6dF and the SDSS spectroscopic DR6 survey (see the breakdown of spectroscopic redshift sources in Table 1) and 20 per cent of which have photometric redshifts from either the empirical training set or the template-fitting method. At a flux limit of S60 = 0.36 Jy, more than 90 per cent of the galaxies in the IIFSCz have

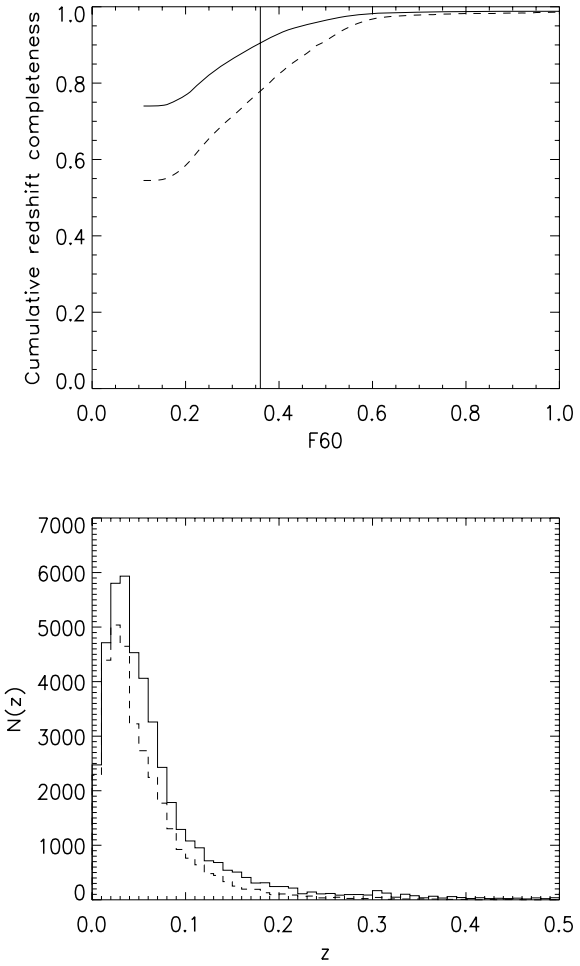

Figure 8. Upper panel: the cumulative redshift completeness as a function of the 60 μm flux density. The dashed line is the completeness curve with only spectroscopic redshifts included and the solid line is with both spectroscopic and photometric redshifts included. The vertical line is S60 = 0.36 Jy. Lower panel: redshift histogram for IIFSCz galaxies.

either spectroscopic or photometric redshifts. There are 344 known QSOs and only six of these do not have redshift information.

6 DISCUSSIONS

Fig. 10 shows the distribution of infrared bolometric luminosity, $L_{\rm IR}$, versus optical bolometric luminosity, $L_{\rm opt}$, from our template fitting. The figure is colour-coded by the infrared template type, and only galaxies with at least two detected FIR fluxes are included. Most cirrus galaxies have $L_{\rm IR} < L_{\rm opt}$, as expected if the emission is from an optically thin interstellar dust distribution. However, there are some cirrus galaxies with $L_{\rm IR} > L_{\rm opt}$, indicating a higher dust optical depth. There is also the interesting population of cool luminous galaxies, with $L_{\rm IR} > 10^{12}\,{\rm L}_{\odot}$, discussed by Rowan-Robinson et al. (2005, 2008). The highest infrared luminosities are dominated by Arp 220 template types, but there are significant numbers of ultraluminous infrared galaxies which are M82 template types. Objects with $L_{\rm opt} > 10^{12}\,{\rm L}_{\odot}$ are type 1 QSOs and most have $L_{\rm IR} < L_{\rm opt}$, as expected if the infrared emission is dominated by a dust torus illuminated by the QSO.

One of the interesting discoveries of the *IRAS* FSC was the existence of hyperluminous infrared galaxies (Rowan-Robinson et al. 1991; Rowan-Robinson 2001), galaxies with infrared luminosities $>10^{13}\,{\rm L}_{\odot}$. In our catalogue, we now find 159 hyperluminous infrared galaxies with S60 > 0.2 Jy, of which 38 have S60 > 0.36 Jy, our completeness limit. Fig. 11 shows the distribution of bolometric infrared luminosity, $L_{\rm IR}$, versus redshift, colour-coded by the infrared template type. Only galaxies with more than one FIR flux detected are included. In our template fitting, we do not permit the cirrus template to be used if $L_{\rm IR} > 10^{13}\,{\rm L}_{\odot}$. We see that the highest luminosities are dominated by the high optical depth starburst, Arp

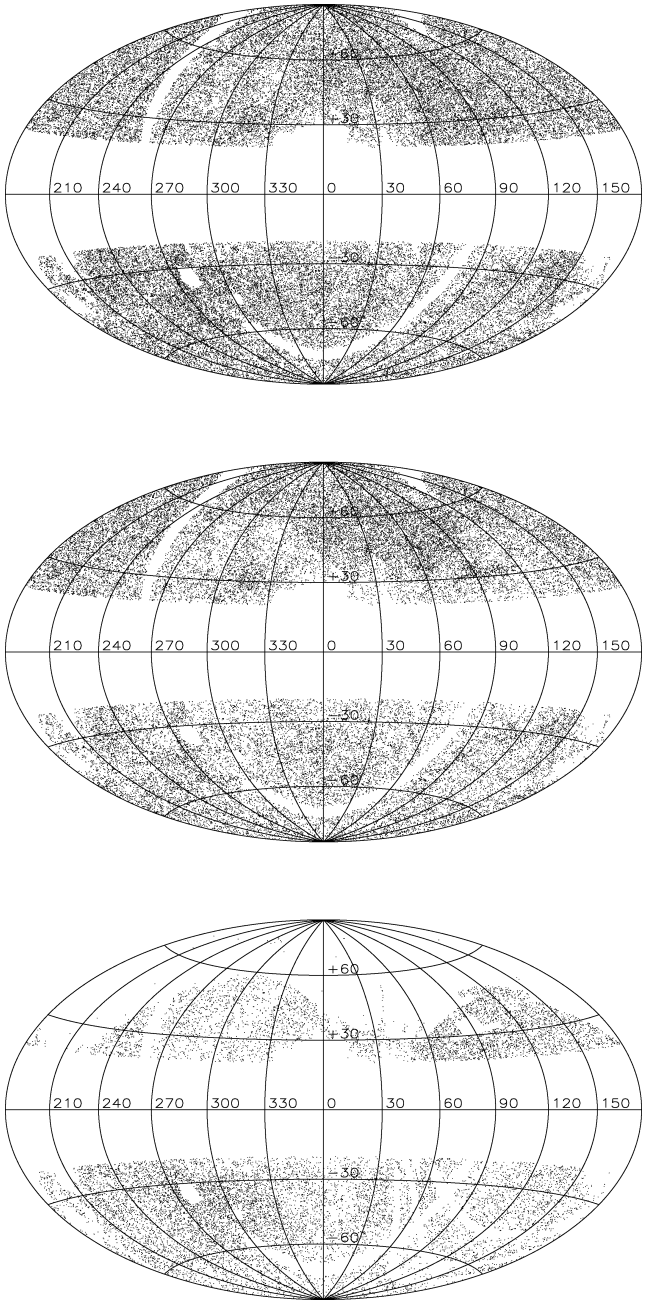

Figure 9. The sky distribution of all galaxies in the IIFSCz (top), galaxies with either spectroscopic or photometric redshift (middle) and galaxies without redshift (bottom), in galactic coordinates. The depth of the IIFSCz is affected by variations in the *IRAS* coverage across the sky. Note that in the bottom panel, the redshift completeness is very high in regions covered by the SDSS DR6 survey.

Table 1. Source of spectroscopic redshift for the IIFSCz.

Source of z_{spec}	Number	Fraction (per cent)
NED	23 703	39.3
IRAS FSSz	568	0.9
IRAS PSCz	1972	3.3
6dF Galaxy Survey	2794	4.6
SDSS spectroscopic DR6	3844	6.4
Total	32 881	55.0

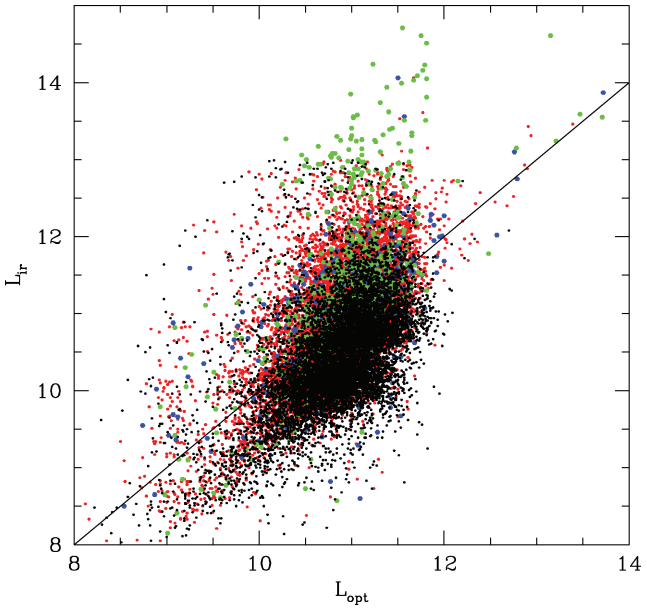

Figure 10. Infrared bolometric luminosity (L_{IR}) versus optical bolometric luminosity (L_{opt}), colour-coded by the infrared template type.

220, templates, but with some sources dominated by AGN dust tori and a few M82 starbursts.

There are 50 galaxies with $L_{IR} > 10^{14}$ L$_\odot$, of which 15 are QSOs, and eight with $L_{IR} > 10^{15}$ L$_\odot$. The latter include the well-known hyperluminous galaxies IRAS F10214+4724 and 08279+5255, both of which are lensed and are discussed in Rowan-Robinson (2001). Five of these eight galaxies have spectroscopic redshifts. The hyperluminous infrared galaxies in our catalogue will be discussed in a subsequent paper (Rowan-Robinson et al., in preparation).

Fig. 12 illustrates our predicted submm fluxes derived from our template fits. It shows the predicted 350 μm flux versus redshift, again colour-coded by infrared template type. We have indicated the predicted flux limit of the *Planck* Surveyor all-sky survey. We are predicting that over 23 000 of the sources in our catalogue will have 350 μm fluxes above the *Planck* all-sky survey limit of 0.3 Jy. *Planck* should detect significantly more sources than this because (i) not all the sources in our catalogue have redshift estimates and template fits and (ii) some high-redshift infrared galaxies may fall below our 60 μm detection limit but still be detectable at 350 μm. Our catalogue should, however, include all sources detected by *Planck* with $z < 0.5$.

7 CONCLUSIONS

We have presented the IIFSCz. It contains 60 303 galaxies selected at 60 μm from the *IRAS* FSC, covering around 61 per cent of the whole sky. The process of retrieving spectroscopic redshifts, multiwavelength cross-matching and photometric redshift estimation is described in some detail. In the catalogue, we give the best possible position, *IRAS* fluxes, optical, near-infrared and/or radio identifications, spectroscopic redshift (if available) or photometric redshift (if possible), predicted fluxes at wavelengths ranging from 12 to 1380 μm. Overall, 32 881 galaxies in the IIFSCz (55 per cent) have received spectroscopic redshifts from past redshift surveys such as the *IRAS,* PSCz, FSSz and 6dF and around 12 000 galaxies (∼20 per cent) obtained photometric redshifts through either the training set (for sources with 2MASS or NVSS photometry) or the template-fitting method (for sources with SDSS photometry). At a flux limit of S(60) = 0.36 Jy, the redshift completeness of the

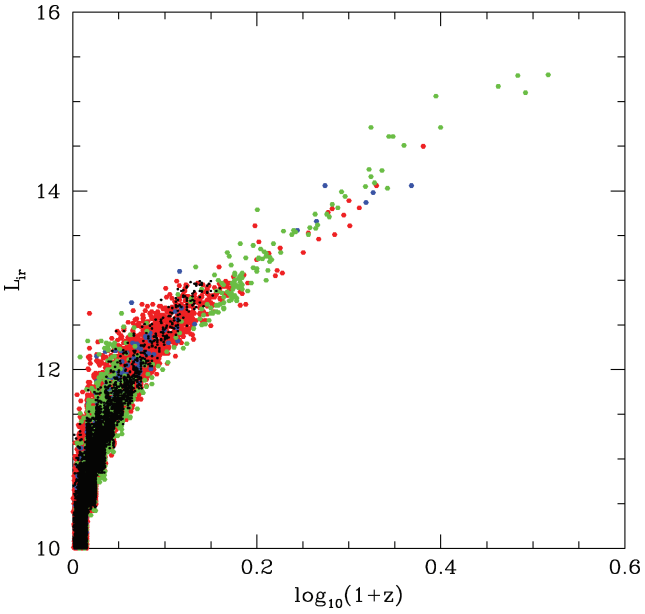

Figure 11. Infrared bolometric luminosity versus redshift, colour-coded by the infrared template type.

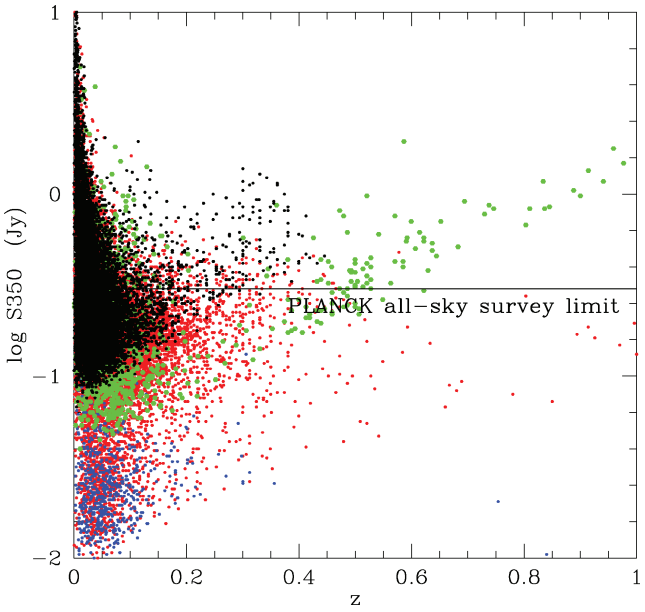

Figure 12. The predicted 350 μm flux in Jy versus redshift, colour-coded by the infrared template type.

catalogue, including both spectroscopic and photometric redshifts, is increased to 90 per cent.

The IIFSCz provides a huge data set for large-scale structure studies and validations of recent and future infrared and submm surveys (e.g. *AKARI*, *Planck* and *Herschel*). Potential users should be aware of issues such as the intrinsic *IRAS* FSC completeness limit, the redshift completeness variations across the sky and the varying quality of photometric redshift derived for different subsets of the catalogue.

ACKNOWLEDGMENTS

We thank Seb Oliver for providing the FSSz catalogue and helpful discussions on obtaining optical identifications in the SDSS DR6 survey. This research has made extensive use of the NED and the SDSS. LW thanks Marion Schmitz and Joseph Mazzarella for their help on various catalogues in the NED. LW is supported by a Dorothy Hodgkin Postgraduate Award (DHPA).

The NED is operated by the Jet Propulsion Laboratory, CALTECH, under contract with the National Aeronautics and Space Administration.

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 and the Max Planck Society and the Higher Education Funding Council for England. The SDSS web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, The University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

REFERENCES

Appleton P. N. et al., 2004, ApJS, 154, 147
Becker R. H., White R. L., Helfand D. J., 1995, AJ, 450, 559
Brusa M. et al., 2007, ApJS, 172, 353
Ciliegi P., Zamorani G., Hasinger G., Lehmann I., Szokoly G., Wilson G., 2003, A&A, 398, 901
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
de Jong T., Klein U., Wielebinski R., Wunderlich E., 1985, A&A, 147, L6
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Erdoğdu P. et al., 2006, MNRAS, 368, 1515
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195
Griffin M. et al., 2007, Adv. Space Res., 40, 612
Hawkins E., Maddox S., Branchini E., Saunder W., 2001, MNRAS, 325, 589
Helou G., Soifer B. T., Rowan-Robinson M., 1985, ApJ, 298, L7
Jarrett T. H., Chester T., Cutri R., Schneider S. E., Huchra J. P., 2003, AJ, 125, 525
Jones D. H. et al., 2004, MNRAS, 355, 747
Jones D. H., Saunders W., Read M., Colless M., 2005, PASA, 22, 277
Kawada M. et al., 2007, PASJ, 59, S389
Maller A. H., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJ, 598, L1
Mann R. G., Saunders W., Taylor A. N., 1996, MNRAS, 279, 636
Moshir M. et al., 1992, Explanatory Supplement to the IRAS Faint Source Survey, Version 2, JPL D-10015 8/92 JPL, Pasadena
Murakami H. et al., 2007, PASJ, 59, S369
Oliver S. J., 1993, PhD Thesis, Univ. London
Pilbratt G., 2004, in Mather J. C., ed., Proc. SPIE Vol. 5487, The Herschel Mission: Status and Observing Opportunities. SPIE, Bellingham, p. 401
Rice W., Lonsdale C. J., Soifer B. T., Neugebauer G., Kopan E. L., Lloyd L. A., de Jong T., Habing H. J., 1988, ApJS, 68, 91
Rowan-Robinson M., 2001, IAUS, 204, 265
Rowan-Robinson M. et al., 1991, Nat., 351, 719
Rowan-Robinson M. et al., 2000, MNRAS, 314, 375
Rowan-Robinson M. et al., 2004, MNRAS, 351, 1290
Rowan-Robinson M. et al., 2005, AJ, 129, 1183
Rowan-Robinson M. et al., 2008, MNRAS, 386, 697 (RR08)
Saunders W. et al., 2000, MNRAS, 317, 55
Serjeant S., Harrison D., 2005, MNRAS, 356, 192
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Skrutskie M. F. et al., 1997, in Garzón F., Epchtein N., Omont A., Burton B., Persi P., eds, ASSL Vol. 210, The Impact of Large Scale Near-IR Sky Surveys. Dordrecht, Kluwer, p. 25
Sutherland W., Saunders W., 1992, MNRAS, 259, 413
Szapudi I., Branchini E., Frenk C. S., Maddox S., Saunders W., 2000, MNRAS, 318, L45
Wang L. et al., 2008, MNRAS, 387, 601
Wolstencroft R. D., Savage A., Clowes R. G., MacGillivray H. T., Leggett S. K., Kalafi M., 1986, MNRAS, 223, 279
York D. G. et al., 2000, AJ, 120, 1579
Yun M. S., Reddy N. A., Condon J. J., 2001, ApJ, 554, 803

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