Automated cross-identifying radio to infrared surveys using the LRPY algorithm: a case study

S. D. Weston,[1,2]★ N. Seymour,[3] S. Gulyaev,[1] R. P. Norris,[2,4] J. Banfield,[5]
M. Vaccari,[6,7] A. M. Hopkins[8] and T. M. O. Franzen[3]

[1]*Institute for Radio Astronomy & Space Research, AUT University, Auckland, New Zealand*
[2]*CSIRO, PO Box 76, Epping, NSW-1710, Australia*
[3]*International Centre for Radio Astronomy Research, Curtin University, Perth, Australia*
[4]*Western Sydney University, Penrith, NSW, Australia*
[5]*Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT-2611, Australia*
[6]*Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, 7535 Bellville, Cape Town, South Africa*
[7]*INAF - Istituto di Radioastronomia, via Gobetti 101, Bologna, I-40129, Italy*
[8]*Australian Astronomical Observatory, PO Box 915, North Ryde, NSW-1670, Australia*

ABSTRACT

Cross-identifying complex radio sources with optical or infra red (IR) counterparts in surveys such as the Australia Telescope Large Area Survey (ATLAS) has traditionally been performed manually. However, with new surveys from the Australian Square Kilometre Array Pathfinder detecting many tens of millions of radio sources, such an approach is no longer feasible. This paper presents new software (LRPY – Likelihood Ratio in PYTHON) to automate the process of cross-identifying radio sources with catalogues at other wavelengths. LRPY implements the likelihood ratio (LR) technique with a modification to account for two galaxies contributing to a sole measured radio component. We demonstrate LRPY by applying it to ATLAS DR3 and a *Spitzer*-based multiwavelength fusion catalogue, identifying 3848 matched sources via our LR-based selection criteria. A subset of 1987 sources have flux density values for all IRAC bands which allow us to use criteria to distinguish between active galactic nuclei (AGNs) and star-forming galaxies (SFG). We find that 936 radio sources ($\approx$47 per cent) meet both of the Lacy and Stern AGN selection criteria. Of the matched sources, 295 have spectroscopic redshifts and we examine the radio to IR flux ratio versus redshift, proposing an AGN selection criterion below the Elvis radio-loud AGN limit for this dataset. Taking the union of all three AGNs selection criteria we identify 956 as AGNs ($\approx$48 per cent). From this dataset, we find a decreasing fraction of AGNs with lower radio flux densities consistent with other results in the literature.

Key words: methods: data analysis – catalogues – surveys – radio continuum: galaxies.

1 INTRODUCTION

By obtaining large datasets at different wavelengths, which are sensitive to different galaxy properties, one can separate the different influences on the formation and evolution of galaxies. One key problem in combining these multiwavelength surveys is determining which sources, observed at different wavelengths, are truly associated with one another and which are unrelated. We are now entering a new epoch of radio astronomy with even larger surveys of galaxies such as the Evolutionary Map of the Universe (EMU, Norris 2012)

and the MeerKAT International GigaHertz Tiered Extragalactic Exploration survey (MIGHTEE, Van der Heyden & Jarvis 2010) using the Square Kilometre Array (SKA) precursors ASKAP and MeerKAT, which will increase the number of faint (sub-mJy) radio sources by orders of magnitude. All of these faint radio sources will need to be cross-matched with other wavelength datasets in order to tackle numerous science questions, in particular the star-formation history of the Universe, a key goal for the SKA (Prandoni & Seymour 2015). Where surveys have similar wavelengths, resolutions and sensitivities, this matching of sources is generally straightforward and unambiguous (e.g. by simple nearest-neighbour matching). However, where the surveys are dissimilar, with very different resolutions and/or sensitivities, such as between radio and infrared,

★ E-mail: stuart.weston@aut.ac.nz

a nearestneighbour approach becomes unreliable. Furthermore, radio sources may have complex structure which is very different to the appearance at optical and near-infrared (NIR) wavelengths.

There are methods which overcome the problem of cross-matching different wavelengths, such as the Likelihood Ratio (LR) method first proposed by Richter (1975), the Poisson Probability (PP) method Downes et al. (1986) and a Bayesian method Fan et al. (2015). The LR method not only uses positional information but also flux and source density. In the rare complex cases, beyond the capabilities of computational algorithms, human pattern recognition can be used via citizen science projects such as Radio Galaxy Zoo (Banfield et al. 2016). Despite having many participants, such methods are time consuming with each cross-match requiring 10 or more classifications. Hence, with potentially 70 million sources to be detected by EMU, it is desirable to maximize the fraction which can be cross-matched in an automated way.

In this paper, we present an implementation of the LR method within an algorithm for cross-identifying sources between a radio catalogue and an infrared (IR) catalogue. The Likelihood Ratio in PYTHON code (LRPY) developed for this paper is also released to the public.

The paper is structured as follows. In Section 2, we present the ATLAS and Fusion surveys. In Section 3, we present our implementation of the LR technique. In Section 4, we review the selection criteria for possible true matches and to identify multiple components in the infrared domain for one radio source. In Section 5, we take the cross-matches between radio and infrared and examine some of their colour–colour properties and radio to infrared flux as a function of redshift. We present an additional catalogue to ATLAS DR3 in Section 6 by including our matches with the *Spitzer* Data Fusion catalogue with the corresponding LR and reliability values.

2 DATA

This work and analysis concentrates on the ATLAS DR3 catalogue (Franzen et al. 2015) and the *Spitzer* Data Fusion catalogue (Vaccari et al. 2010; Vaccari 2015).[1]

2.1 ATLAS DR3 radio catalogue

The ATLAS survey was conducted with the Australia Telescope Compact Array (ATCA) between 2002 April and 2010 June and covers 1.3–1.8 GHz, over an area coinciding with the *Chandra* Deep Field South (CDFS) and the European Large Area ISO Survey – South 1 (ELAIS-S1 Rowan-Robinson et al. 2010). These areas also coincide with the Spitzer Wide-area Infrared Extragalactic(SWIRE) survey (Lonsdale et al. 2003), thus providing the multiwavelength data required for identification of the radio sources and further analysis (e.g. Mao et al. 2012). ATLAS Data Release 1 (DR1) was presented in Norris et al. (2007) and Middelberg et al. (2008) and DR2 in Hales et al. (2014).

The ATLAS DR3 component source catalogue (hereafter referred to as the 'ATLAS' catalogue) presented in Franzen et al. (2015) found a total of 5191 radio source components over 5σ for both fields. There are 3078 source components in CDFS above 70 μJy beam^{-1} and 2113 source components above 85 μJy beam^{-1} in ELAIS-S1. The restoring beam parameters used in our subsequent analysis for the two ATLAS fields are given in Table 1. Middelberg et al. (2008) identified a positional offset between the ATLAS DR1

Table 1. Restoring beam for each ATLAS radio image.

Field	Φ_{Maj} (arcsec)	Φ_{Min} (arcsec)	Position angle (degrees)
CDFS	16.8	6.9	1.0
ELAIS-S1	12.2	7.6	−11.0

and SWIRE catalogues in ELAIS-S1. A systematic offset of mean value 0.08 ± 0.03 arcsec in right ascension and 0.06 ± 0.03 arcsec in declination was found. We find an identical offset between the ATLAS DR3 and Fusion catalogues in our analysis and apply this to the ATLAS catalogue.

2.2 UV to Mid-IR Fusion Catalogue

The *Spitzer* Data Fusion catalogue is a multiwavelength far-UV to mid/far-IR catalogue of *Spitzer* selected sources, hereafter referred to as the 'Fusion' catalogue (Vaccari, Marchetti & Gonzalez-Solares 2010; Vaccari 2015), which covers the CDFS and ELAIS-S1 fields. This catalogue is based on detections at 3.6 μm with the IRAC instrument (Fazio et al. 2004) on board the *Spitzer Space Telescope* (Werner et al. 2004), down to a flux density of 4.6 μJy in the CDFS field and 4.8 μJy in the ELAIS-S1 field. There are 391 518 *Spitzer*/IRAC sources in ELAIS-S1 and 462 638 in CDFS. The cross-identification performed in this analysis makes use of the IRAC 3.6 μm flux density. We note that this catalogue contains very few photometric and spectroscopic redshifts pertaining to radio sources, but the Herschel Extragalactic Legacy Project (HELP) Vaccari (2016) is in the process of putting together multiwavelength data, compute photometric redshifts and physical parameters for sources in ATLAS (and ASKAP/EMU Early Science) fields.

2.3 OzDES data

OzDES (Yuan et al. 2015) a multiyear, 100 night spectrscopic survey of the Dark Energy Survey (DES Dark Energy Survey Collaboration et al. 2016) deep supernova fields with the 4 m Anglo-Australian Telescope. While the primary goal is to obtain spectra of supernovae detected by DES and their host galaxies, the design of the survey allows other projects to target sources in these fields. Five of the DES deep fields overlap with CDFS and ELAIS-S1, hence OzDES is able to provide spectra of many radio sources. In this work, we use the 2016-02-25 version of the Global Redshift Catalogue which includes OzDES and literature spectra in the DES deep fields.

2.4 Catalogue coverage

We overlay all the sources in the ATLAS and Fusion catalogues in Fig. 1. While the Fusion catalogue completely covers the ATLAS observations in ELAIS-S1, part of the ATLAS CDFS data is not covered by the Fusion catalogue. Hence, we restrict all our analysis to the following sub-region for CDFS ($51.7° \leq$ RA $\leq 54.2°$ and $-29.0° \leq$ Dec $\leq -27.2°$). This region has been placed so that it is inside Fusion and 100 arcsec from the edge. We also restrict our analysis in ELAIS-S1 to ($7.3° \leq$ RA $\leq 9.7°$ and $-44.6° \leq$ Dec $\leq -42.9°$).

3 SOURCE CROSS-MATCHING TECHNIQUES

In this work, we further adapt the LR for the ATLAS and Fusion catalogues but also extend this technique to account for multiple
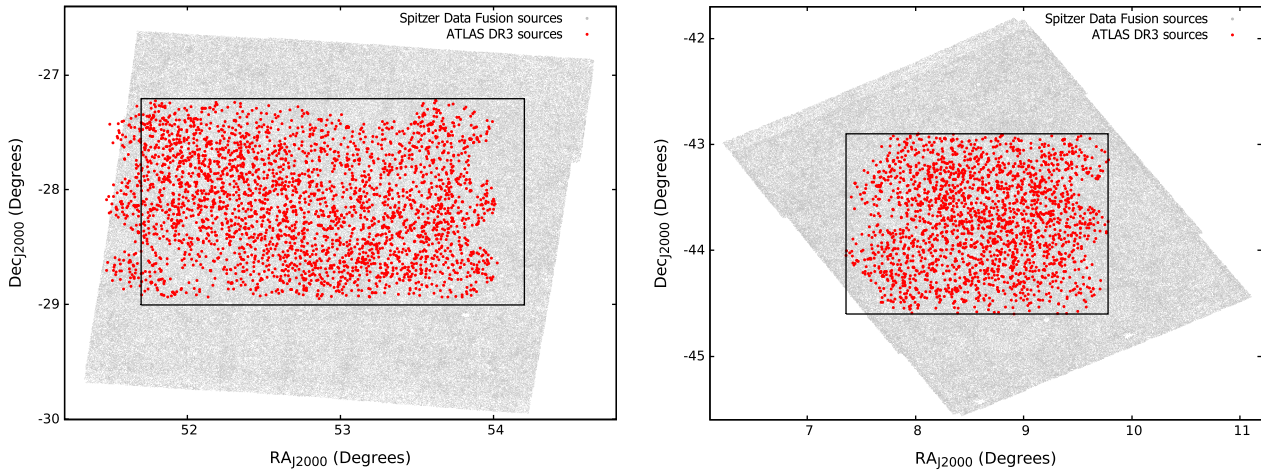
Figure 1. Spatial distribution of ATLAS and Fusion sources in CDFS field (left) and in ELAIS-S1 (right). We also mark the rectangular regions used in this work.

infrared candidates. In future work, we shall show how we adapt it for complex radio sources.

3.1 Likelihood ratio technique

Richter ([1975](#)) presented an early statistical treatment to cross-match optical sources to the low resolution 5C3 radio survey by applying the statistical separation of real and chance identifications. This technique was then developed by de Ruiter, Willis & Arp ([1977](#)) to match optical sources to radio sources detected with the Westerbork Synthesis Radio Telescope using a probability ratio, referred to as the LR. It uses the ratio of the a priori probability, $\mathrm{d}p(r|id)$, that radio source and optical counterpart are intrinsically located at the same position, and the probability that the optical object is an unrelated background or foreground source. This method was further refined by Sutherland & Saunders ([1992](#)) who defined the LR as the ratio between the probability that a candidate source is the correct identification and the probability that it is an unrelated background or foreground source as a function of magnitude.

The LR technique is commonly used to cross-match low-resolution long wavelength surveys with optical data of higher resolution. For example Ciliegi et al. ([2003](#)) used this method to find optical counterparts for the VLA 6 cm Lockman Hole survey. Ciliegi et al. ([2005](#)) used the same technique to look for optical and NIR counterparts for the VLA 1.4 GHz survey in the VIMOS VLT deep survey. More recently Smith et al. ([2011](#)) used the technique with some further refinements to Sutherland & Saunders ([1992](#)) to identify optical counterparts to 250 μm sources from the *Herschel*–ATLAS survey. The refinements from the Smith et al. ([2011](#)) technique has been followed by Fleuren et al. ([2012](#)) with some modifications when matching sources between the NIR VISTA VIKING and *Herschel*–ATLAS SPIRE catalogues.

In recent applications (e.g. by Smith et al. [2011](#)), this is presented in the form:

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

where $q(m)$ is the probability distribution of the true counterparts as a function of magnitude m, $f(r)$ is the distribution of probability density per unit solid angle and $n(m)$ is the surface density of background and foreground objects.

In the following sub-sections the terms in equation (1) are discussed with reference to the Fusion catalogue and the ATLAS catalogue. As the Fusion catalogue provides 3.6 μm flux densities we simply use flux densities, S_ν, rather than magnitudes from this point.

3.1.1 The Probability Distribution Function

Here, we follow the standard approach to the definition of the LR (Sutherland & Saunders [1992](#)). Therefore, we use $f(r)$ in equation (1) as the probability distribution function (PDF) of the positional errors, see also the definition of $f(r)$ given by Fleuren et al. ([2012](#)). We note a confusion in definition of $f(r)$ in Smith et al. ([2011](#)), where they first define $f(r)$ as 'the radial probability distribution of *offsets* between the 250-μm positions and the SDSS *r*-band centroid', that is as the PDF of the offsets between objects of two catalogues, then (in the next paragraph) as the 'probability distribution function of the positional error'. The difference between two definitions is significant, because the 'probability distribution function of the positional errors' is determined by the Gaussian function, whence the PDF of the offsets between objects of two catalogues is described by the Rayleigh distribution function.[2] In our case, $f(r)$ is a two-dimensional Gaussian distribution of the form:

$$f(r) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right). \tag{2}$$

Here, r is the angular distance (in arcseconds) from the radio source position and σ is the combined positional error given by:

$$\sigma = \sqrt{\sigma_{\mathrm{Posn}}{}^2 + \sigma_{\mathrm{Atlas}}{}^2 + \sigma_{\mathrm{Fusion}}{}^2}. \tag{3}$$

The Fusion absolute position uncertainty, σ_{Fusion}, is taken as 0.1 arcsec (Vaccari et al. [2010](#)) and the ATLAS absolute position uncertainty, σ_{Atlas}, is taken from Huynh et al. ([2005](#)), who argued that the positional accuracy of 1.4 GHz ATCA observations for 10σ detections is 0.6 arcsec.

The positional uncertainty term, σ_{Posn}, of the individual lower resolution ATLAS sources depends on the signal-to-noise ratio (SNR) and the full-width at half-maximum (FWHM) of the radio restoring beam (point spread function in other words). We use the value for

---

[2] We thank the reviewer who attracted our attention to this confusion in Smith et al. ([2011](#)).
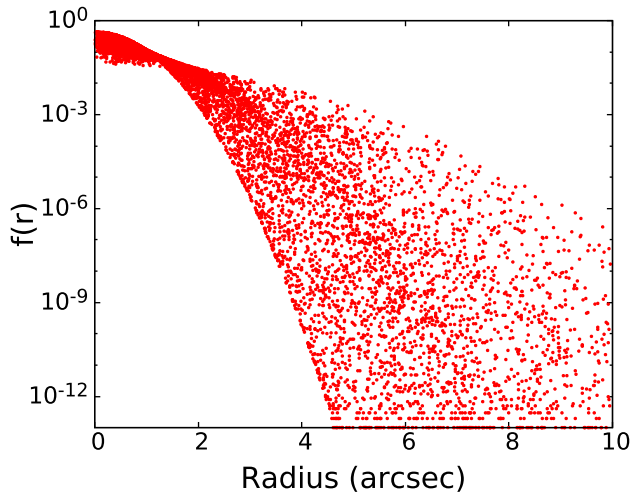
Figure 2. $f(r)$ versus r for all the potential candidates for both fields CDFS and ELAIS-S1 within the initial 10 arcsec search radius.

σ_{Posn} as provided in Ivison et al. (2007) and used in Huynh et al. (2005):

$$\sigma_{Posn} \simeq 0.6 \left(\frac{\text{FWHM}}{\text{SNR}} \right) \qquad (4)$$

As the position angle of the restoring beam is small for both fields (see Table 1) we can assume it is zero, hence:
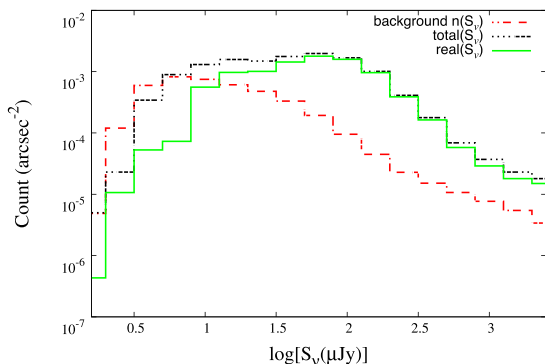
$$\sigma_{Posn} = \frac{0.6}{\text{SNR}} \times \left(\left(\frac{\sin\theta}{\Phi_{Min}} \right)^2 + \left(\frac{\cos\theta}{\Phi_{Maj}} \right)^2 \right)^{-1/2}, \qquad (5)$$

where θ is the position angle of the candidate Fusion counterpart relative to the radio source defined clockwise from North. The SNR values are taken from the ATLAS radio catalogue for each source. The terms Φ_{Min} and Φ_{Maj} are the values of minor and major axes of the beam given in Table 1.

The distribution of the values of $f(r)$ with radius from equation (2) for the individual candidate Fusion counterparts found within an initial search radius of 10 arcsec is shown in Fig. 2. We can see that $f(r)$ is $<10^{-3}$ for $r > 6$ arcsec. We further discuss the rationale for choosing a final search radius of 6 arcsec in Section 3.1.4.

3.1.2 The background flux density probability function

The quantity $n(S_\nu)$ is the surface density of background and foreground Fusion sources with flux density, S_ν. The surface density of

Fusion sources not related to ATLAS radio sources can be obtained from the Fusion catalogue by one of two methods, both of which have been implemented within the LRPY algorithm:

(i) Use all Fusion sources within an annulus of 6 arcsec $< r < 100$ arcsec around each radio candidate – this is referred to as the *local* method.

(ii) Use all Fusion sources from the area of overlap between the two catalogues (defined in Section 2.4) – this is referred to as the *global* method and is the default in LRPY.

With the *local* method, care must be taken that the annuli are not too close to the edge of the field as this can result in a lower count for the background sources as they encompass regions beyond the survey with no sources. To mitigate this edge effect, only annuli 100 arcsec from the inside edge of the area are used. The flux densities are binned and the resultant $n(S_\nu)$ values are then divided by the total area covered to produce a density function, for the CDFS and ELAIS-S1 fields (see Fig. 3). These values are stored in a database lookup table for use later in the final LR calculations.

There are advantages and disadvantages to both methods: the *local* method can account for variations in depth and density of a catalogue, and for very large surveys the entire catalogue is not required, but the *global* method can provide better statistics, if area is limited and depth is uniform, which can be important for both bright and faint flux densities where numbers are small. We use the global method as default as it best suits our situation with the Fusion catalogue being uniform in depth.

3.1.3 The true counterpart probability distribution

The true counterpart probability distribution, $q(S_\nu)$, is the probability that a true Fusion counterpart to a radio source has a flux of S_ν at 3.6 μm:

$$q(S_\nu) = \frac{\text{real}(S_\nu)}{\sum_{S_\nu} \text{real}(S_\nu)} \times Q_0 \qquad (6)$$

Here, $real(S_\nu)$ is the background subtracted distribution of flux densities of Fusion sources around an ATLAS source. The coefficient Q_0 represents the probability that a real counterpart is above the detection limit in the matching catalogue; it does not depend on the search radius. To determine $real(S_\nu)$ we take:

$$\text{real}(S_\nu) = \text{total}(S_\nu) - n(S_\nu), \qquad (7)$$

where $n(S_\nu)$ is the surface density of unrelated background/foreground sources introduced in the previous subsection
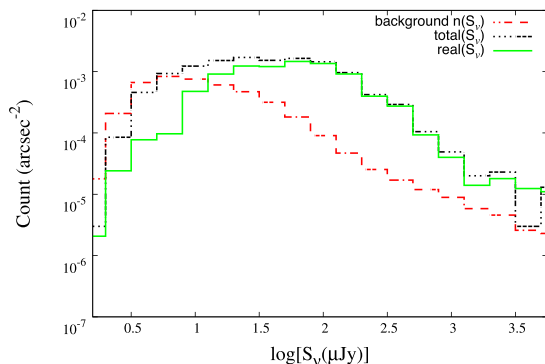

Figure 3. Histograms of the Fusion values for $n(S_\nu)$ background (red dashed line), total(S_ν) (black dotted line) and real(S_ν) (green solid line) for CDFS (left) and ELAIS-S1 (right). Note that S_ν is the 3.6 μm flux.
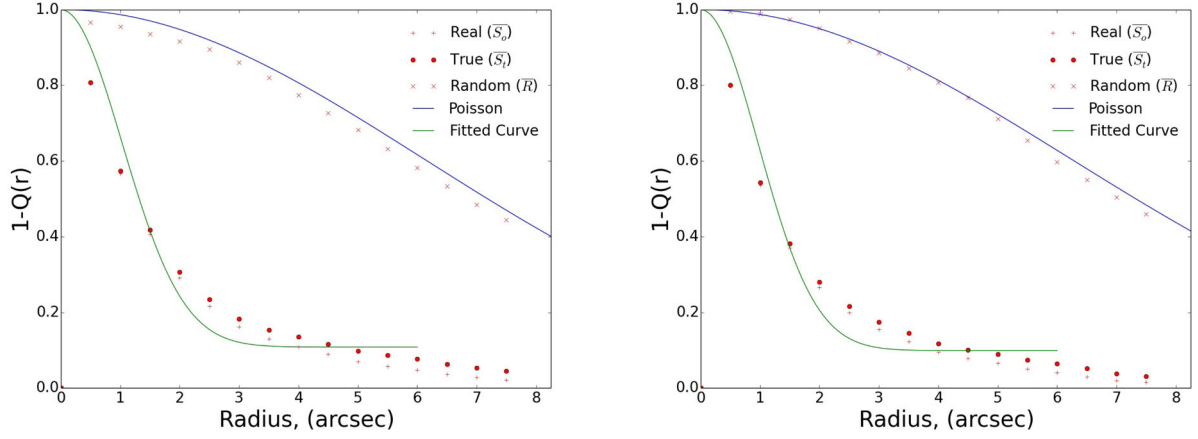
Figure 4. Estimation of Q_0 for CDFS (left) and ELAIS-S1 (right) determined from fitting the ratio, $\overline{S_t}$ (red filled circles), of the fraction of observed blanks, $\overline{S_o}$ (crosses), and the fraction of random blanks, $\overline{R}$ (plusses). The green line represents the functional fit to the ratio (equation 11), and the blue line is an estimate of the fraction of random blanks from Poisson statistics using equation (14). Taking (Fleuren et al. 2012), the dependence of Q on the search radius can be presented in the form $Q(r) = Q_0 \exp\left(-r^2/2\sigma^2\right)$.

and total(S_ν) is the surface density of all Fusion sources to be matched within the search radius, r, including the true counterpart (if above the detection limit) plus unrelated background and foreground sources. These values are kept in the same LRPY database table as $n(S_\nu)$ for use later by the algorithm.

The distributions of real(S_ν) and total(S_ν), as well as distribution of $n(S_\nu)$ are shown in Fig. 3. It should be noted that for the unphysical condition where $n(S_\nu) >$ total(S_ν) (i.e. when the background exceeds the measured distribution), a method is adopted to set *real*(S_ν) to be positive. This occurs at faint and bright flux densities when there is a small number of Fusion sources in a given flux density bin. To keep our estimate of real(S_ν) positive and physical, we replace negative values of real(S_ν)/total(S_ν) with a value determined from the last positive value at faint and bright flux densities. This adaption ensures we account for potential counterparts at the extreme flux density values.

A reasonablyx accurate determination of Q_0 in equation (6) is naturally required. If we were simply to estimate Q_0 by summing real(S_ν) and dividing by the total number of ATLAS sources, we would likely overestimate Q_0 due to source clustering and genuine multiple matches (which we deal with in Section 4.2). While this simple method finds values of $Q_0 = 0.845$ for CDFS and 0.822 for ELAIS-S1, we undertake the following process to estimate its value more accurately. We follow Fleuren et al. (2012), who, to avoid these issues, estimate the value $1 - Q_0$, which in this case will be the fraction of ATLAS sources *without* a Fusion counterpart, which we refer to here as 'blanks'. These will principally be ATLAS sources with true counterparts below the Fusion detection limit, or ATLAS sources with true Fusion counterparts outside the search radius. The latter case is possible when ATLAS sources are complex and the Fusion counterpart may well not correspond to any radio component, but lie between components (lobes) which can be separated by tens of arcsec.

The true fraction of blanks, $1 - Q_0 = \overline{S_t}$, will be greater than the observed fraction of blanks, $\overline{S_o}$, because a fraction of true blanks will have random (i.e. physically unrelated) Fusion sources within the search radius. Hence, we do not wish to falsely count such sources as matches. Therefore, $\overline{S_t}$ equals $\overline{S_o}$ plus some fraction of true blanks 'contaminated' by random Fusion sources. Hence,

$$\overline{S_t} = \overline{S_o} + \overline{S_t} \times \frac{R}{N}, \tag{8}$$

where R is the number of sources, out of N with randomly generated positions, containing one or more Fusion sources within the search radius and N is the total number of radio sources. If we define $\overline{R}$ as the number of N randomly generated sources which do not have a Fusion counterpart within the search radius, such that $N = R + \overline{R}$, it is straightforward to show:

$$\overline{S_t} = \overline{S_o} + \overline{S_t} \times \left(\frac{N - \overline{R}}{N}\right) \tag{9}$$

$$\frac{\overline{S_t}}{N} = \frac{\overline{S_o}}{\overline{R}} \tag{10}$$

Hence, one can determine the fraction of true blanks, $\overline{S_t}/N$, as a function of search radius, r, by determining the ratio of the number of observed blanks, $\overline{S_o}$, to the number of blanks from a randomly generated catalogue, $\overline{R}$, as a function of r. We calculate this result for our case by counting the number of observed blanks with increasing search radii across 0 arcsec $< r <$ 20 arcsec and repeat for a catalogue of N randomly generated positions of Fusion sources. We present these results in Fig. 4 showing, as a function of radius, the fraction of observed blanks, $\overline{S_o}/N$, and the fraction of random blanks, $\overline{R}/N$ and their ratio which equals $\overline{S_t}/N$. As the radius, r, increases to encompass all true counterparts, this result tends towards $1 - Q_0$. We can fit the distribution in Fig. 4 with the following expression:

$$\frac{\overline{S_t}(r)}{N} = 1 - Q_0 \times (1 - e^{-r^2/2\sigma^2}) \tag{11}$$

from Fleuren et al. (2012), where σ is positional uncertainty. This function returns unity at $r = 0$ and $1 - Q_0$ for large r. By fitting for Q_0 and taking σ as the maximum value for the field($\sigma_{\text{CDFS}} = 1.08$ arcsec, $\sigma_{\text{ELAIS-S1}} = 0.868$ arcsec) to the function, using a non-linear least-squares fit, we obtain for both these fields the values and uncertainties for Q_0 presented in Table 2. These values are fairly similar to the ones from our earlier crude estimate, but with the CDFS being a little higher and ELAIS-S1 being slightly lower. We note that this function must pass through (0,1) by definition, but may deviate within the best match search radius due to physical clustering of sources or from the existence of multiple true components.

We also note that we can model the distribution of random blanks, $\overline{R}$, in Fig. 4. The probability that an observed area of sky, $a = \pi r^2$,

Table 2. Estimated fraction, of the non-blanks, Q_0 (ATLAS sources with a true counterpart), and its error δQ_0.

Field	Q_0	δQ_0
CDFS	0.831	0.018
ELAIS-S1	0.825	0.017

has one or more random Fusion source is given by the Poisson distribution $P(a) = 1 - e^{-a\lambda}$, where λ is the density of Fusion sources. Hence, from equation (8), we can write

$$\overline{S_t} = \overline{S_o} + \overline{S_t}(1 - e^{-a\lambda}) \tag{12}$$

which we can rearrange to

$$\overline{S_t} = \overline{S_o}e^{a\lambda} \tag{13}$$

and therefore from equation (10), we get:

$$\frac{\overline{R}}{N} = e^{-a\lambda} \tag{14}$$

In Fig. 4, we overlay this function on the random blanks distribution with radius using a density of Fusion sources of $\lambda = 0.004\,\mathrm{arcsec}^{-2}$, for both fields. We note this theoretical determination matches our empirical determination well.

3.1.4 The search radius

Fleuren et al. (2012) deal with 1 376 606 near-IR sources in the area of 56 deg², which results in density of near-IR sources of $\lambda = 6.8\,\mathrm{arcmin}^{-2}$ and mean intersource distance of $r_0 = (\pi\lambda)^{-1/2} \sim 13$ arcsec. They chose the search radius $r = 10$ arcsec, which is 77 per cent of the mean intersource distance. In our case, the Fusion source density is higher ($\sim 15.1\,\mathrm{arcmin}^{-2}$) and therefore the mean intersource distance is smaller: $r_0 = 8.7$ arcsec. To be consistent with Fleuren et al. (2012), we chose the search radius at 77 per cent of our mean intersource distance: $r = 8.7$ arcsec $\times\, 0.77 \sim 6$ arcsec. Also as shown in Section 3.1.1, the function $f(r)$ exponentially decreases making the LR vanishingly small, $<10^{-3}$, outside $r = 6$ arcsec.

4 ANALYSIS

In this section, we analyse different aspects of the resultant cross-matches and present how we determine criteria for selecting true matches from the LR and Reliability values. We then present a method to identify potential pairs of Fusion sources where both are likely contributing to the radio emission of an ATLAS source.

4.1 Proposed selection criteria

Due to the high density of background sources in the Fusion catalogue there can be $0 \le M \le 5$ possible candidate Fusion counterparts for a given radio source within the search radius of 6 arcsec (see Table 3). Included in Table 3 is the expected number of radio sources in each field with N fusion potential counterparts from a random distribution, i.e. via Poisson statistics. The numbers we find are higher than those from Poisson statistics suggesting (a) potentially more than one galaxy is contributing to the radio emission and (b) there may be clustering around the host galaxies of radio sources. The former option is discussed in Section 4.2 and we noted the latter point in Section 3.1.3.

Table 3. Statistics of Fusion counterparts inside the 6 arcsec search radius around ATLAS sources. The first column is the number of Fusion matches; the second column is the number of ATLAS sources with the corresponding number of Fusions matches (M) for CDFS; the third column is the percentage of the total. Columns four and five are the same, but for the ELAIS-S1 field.

M (matches)	CDFS Poisson	CDFS Count	CDFS Percent	ELAIS S1 Poisson	ELAIS S1 Count	ELAIS S1 Percent
0	1905	378	12.2	1307	177	8.3
1	914	1657	53.8	628	1157	54.7
2	219	832	27.0	151	615	29.1
3	35	185	6.0	24	138	0.6
4	4	23	1.0	3	24	1.1
5	0.4	3	0.1	0.3	2	0.1
Totals		3078			2113	

To select from these M possible candidates a reliability value for each can be determined, thus:

$$R_j = \frac{L_j}{\sum_{i=1}^{M} L_i + (1 - Q_0)}, \tag{15}$$

where R_j is the reliability that the candidate Fusion counterpart j of M possible counterparts is associated with the radio source. The sum is taken over all M potential candidates within the 6 arcsec search radius and Q_0 is the probability that the true Fusion counterpart is above the detection limit (determined in Section 3.1.2 and presented in Table 2). Plots of Reliability versus LR for each candidate counterpart for both fields are presented in Fig. 5.

There is always a trade-off between maximizing the number of radio sources with 'reliable' counterparts and minimizing the contamination of false associations. Equation (15) permits us to compare the relative likelihood of an association between an ATLAS and a Fusion source in the situation where we have two or more potential counterparts. Determining the appropriate cut-off values in LR and Reliability is therefore crucial for any scientific analysis.

Reliability can also be calculated for the case of a single Fusion source, $M = 1$ (one Fusion source in the search radius):

$$R_j = \frac{L}{L + (1 - Q_0)}. \tag{16}$$

Hence, once an LR cut-off, L_c, is determined, the corresponding cut-off value of reliability, R_c, can easily be calculated for single sources as we know Q_0 (here we take $Q_0 = 0.85$).

Fig. 6 shows the families of theoretical curves L versus S_ν and R versus S_ν for the range of r (distance between the radio source and Fusion candidate) from 0 to 5 arcsec (all inside the search radius of 6 arcsec) and $Q_0 = 0.85$. They are calculated for the set of real(S_ν) and $n(S_\nu)$ we observe in the CDFS field for radio sources detected. The upper plots are computed for $\sigma = 1.2$ arcsec, which is close to maximum value of σ we deal with in CDFS field (Section 3.1.1), the bottom plots correspond to $\sigma = 0.6$ arcsec (close to minimum value of σ in CDFS field).

We can choose the L_c for single Fusion sources in such a way that for $\sigma = 1.2$ arcsec almost all single Fusion sources within $r = 5$ arcsec are considered as true counterparts. This condition is fulfilled when $L_c = 0.01$ (horizontal dashed line in Fig. 6) and corresponds to a reliability cut-off of $R_c = 0.055$ for CDFS and 0.054 for ELAIS-S1. These cut-off values are shown in graphs with horizontal solid lines. The horizontal dashed lines show a much stronger criterion for cut-off values of $L_c = 0.1$ and the corresponding $R_c = 0.37$ for CDFS
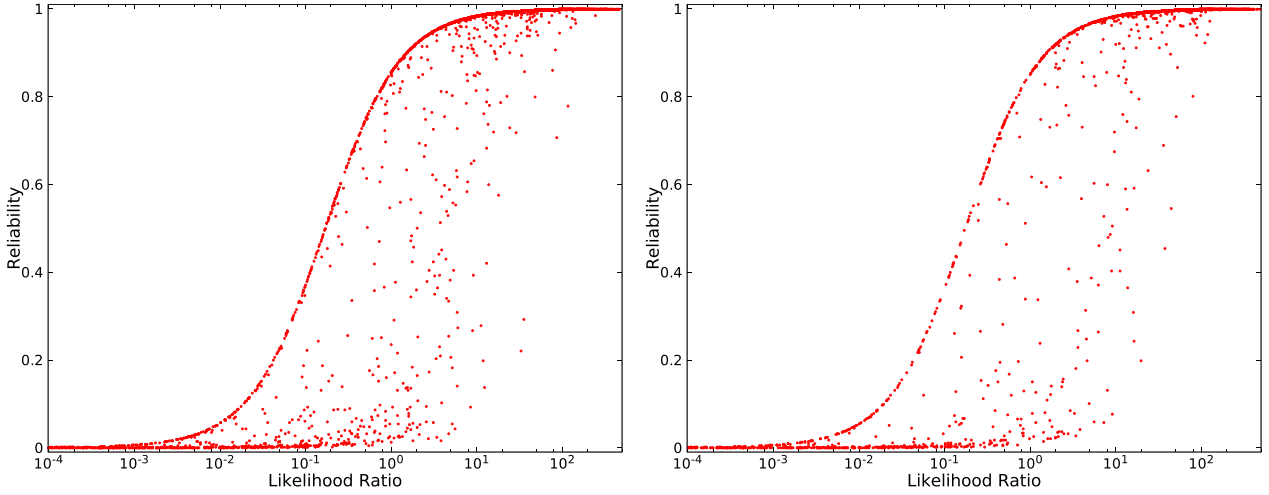
Figure 5. Plots showing the variation of the Reliability, R, as a function of the LR for CDFS (left) and ELAIS S1 (right). For both plots, we note some symmetry of points around $R \approx 0.5$ and as discussed in Section 4.2 could be used to identify potential Fusion pairs being related to one radio source.
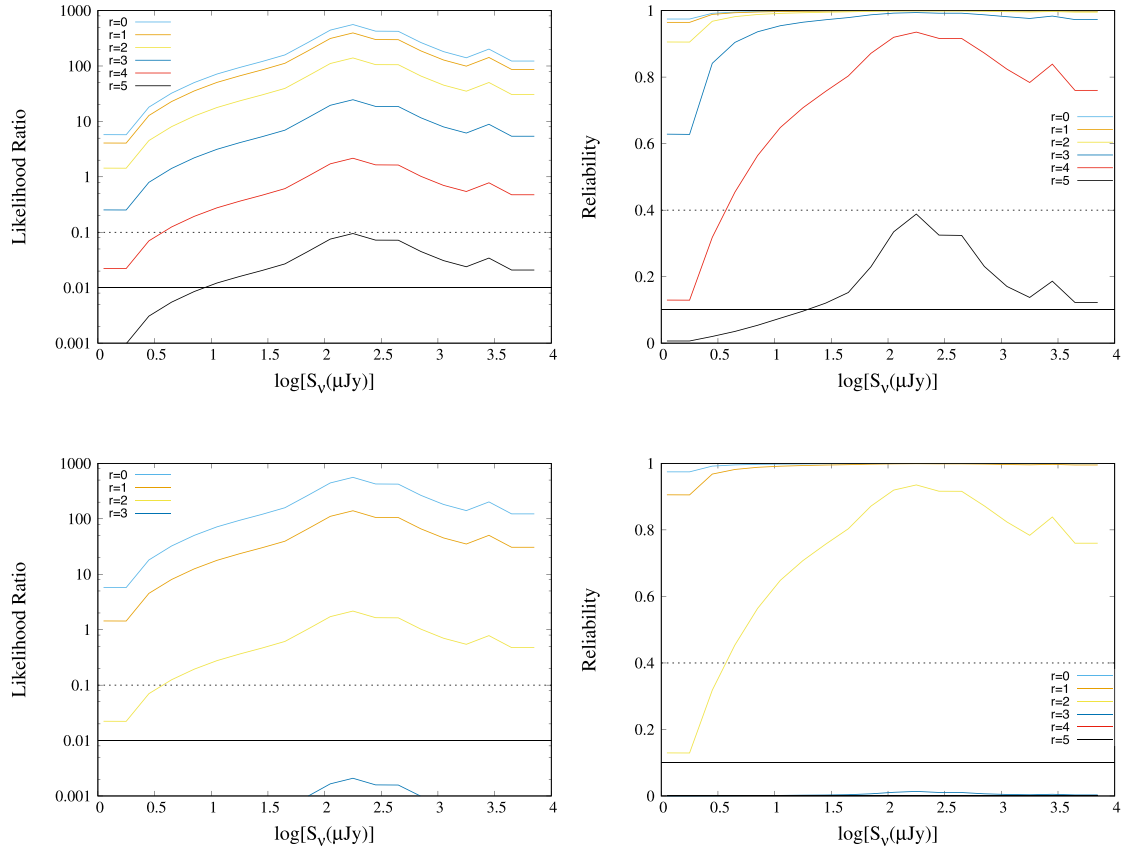

Figure 6. LR against $\log(S_v)$ (left column) and reliability against $\log(S_v)$ (right column) for extreme values of the positional uncertainty $\sigma = 0.6$ arcsec (bottom row) and 1.2 arcsec (top row) in CDFS. Families of curves are computed for distances r between a candidate Fusion counterpart and the ATLAS source in the range $r = 0$–5 arcsec. Distributions of real(S_v) and $n(S_v)$ used to determine the LR are taken from the CDFS field statistics. Horizontal solid lines corresponding to suggested cut-off values presented in Section 4.1, for $L = 0.01$ and $R = 0.1$ are drawn on the figures. The horizontal dashed lines show a much stronger selection criteria $L = 0.1$ and $R = 0.4$. Note that S_v is the 3.6 µm flux.

and 0.36 for ELAIS-S1. In this case, all Fusion sources with $r > 4$ arcsec are excluded from consideration as possible counterparts.

Another approach to determining a value for the reliability cut-off, where those candidates with a reliability greater than R_c can be treated as true counterparts, was used by Smith et al. (2011) who estimated the number of false cross-matches using:

$$N_{\text{false}}(R_c) = \sum_{R_i^{\text{Max}} \geqslant R_c} (1 - R_i). \qquad (17)$$
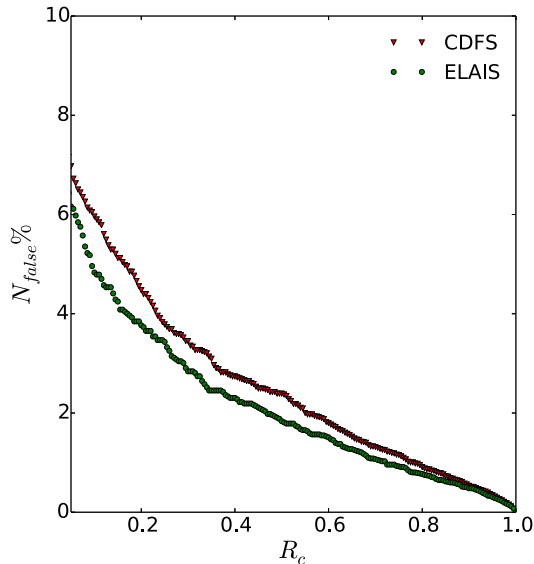
Figure 7. Estimated percentage of the false cross-matches, N_{false}, as a function of the Reliability cut-off, R_c, for CDFS (red) and ELAIS-S1 (green) determined with equation (17).

Fig. 7 shows N_{false} as a function of R_c for our two fields. Smith et al. (2011) used a reliability limit of 0.8 which gave them a contamination rate of 4.2 per cent. Bonzini et al. (2012) selected only those candidates with a reliability greater than 0.6 as the threshold to ensure the expected number of spurious associations was below 5 per cent of the auxiliary catalogue; and at the same time maximizing the number of identified sources. Using a similar acceptable contamination threshold at 5 per cent for our datasets, results in $R_c = 0.1$ for both CDFS and ELAIS-S1 fields (Fig. 7). As we discussed above, this value of R_c corresponds to $L_c = 0.01$ for single Fusion sources.

Using the 5 per cent contamination threshold, we can accept only Fusion counterparts with $L \geq L_c$. Here, we use $L_c = 0.01$ and we reject all Fusion counterparts below this value. We apply this to all Fusion sources, whether they are single or multiple.

For the ATLAS fields using a LR cutoff of 0.01 (reliability cutoff of 0.1) and using equation (17), we have for CDFS $N_{\text{false}} = 159$ which is 5.2 per cent; and for ELAIS-S1 we have $N_{\text{false}} = 99$ which is 4.8 per cent. Using this cut-off there are 2135 ATLAS sources with at least one match in the CDFS field and 1580 in the ELAIS-S1 field. We give an example of this in the Fig. A4.

4.2 Double and multiple fusion counterparts

One ATLAS radio source due to its unresolved peak in a low-resolution radio image could potentially be produced by two or more radio sources blended into one apparent 'source' by the large radio beam. In this section, we modify the LRPY algorithm to identify possible double blended radio sources using the background sources from Fusion.

When reviewing the Reliability versus LR plots in Fig. 5, we note a symmetry of some data points at high values of the likelihood around $R \approx 0.5$. This symmetry has also been noted by Smith et al. (2011) in a similar area in their analysis and they surmise that these could be interacting galaxy counterparts; four of their sources had multiple counterparts with spectroscopic redshift differences of $\Delta z \lesssim 0.001$. Also Fleuren et al. (2012) highlight these possible multiple counterparts, and propose that these could be either merg-
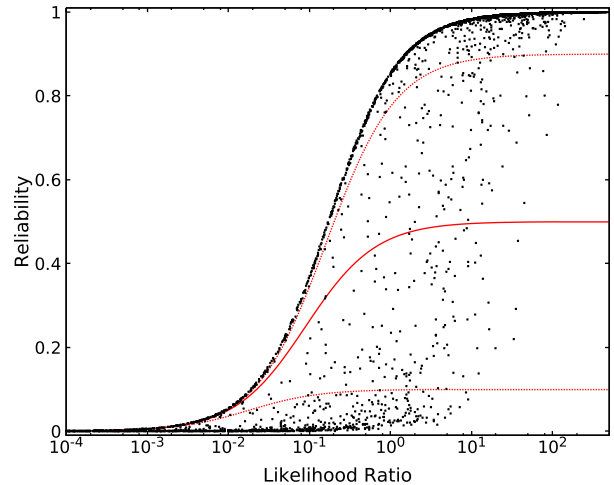
Figure 8. LR versus Reliability for all possible matches within the 6 arcsec search radius for both fields. Also included are the selection limits using equation (19), the upper and lower selection limits ($\beta \pm 0.4$) are marked with red dotted lines and the axis of symmetry of points ($\beta = 0$) is marked by a solid red line.

ing galaxies or members of the same cluster. Fleuren et al. (2012) found matches to 37 sources (out of 1444) with a mean redshift difference of 0.0011 with a maximum difference of $\Delta z = 0.0187$. Within our catalogue, we found these ATLAS sources with two candidate Fusion counterparts to have similar flux density and similar angular separation. Thus, we consider the possibility that these pairs of Fusion counterparts could be close or interacting galaxies and may both be contributing to the radio emission from one source. We investigate this further below and group them together introducing the term InfraRed Double (IRD).

When two Fusion sources with similar LRs are found in the search field around a radio source, the reliability of both sources is determined by

$$R = \frac{L}{(1 - Q_0) + L/0.5} \, , \tag{18}$$

which follows from equation (16) when $L = L_1 = L_2$. Equation (18) results in $R = 0.5$ when $L \gg 1 - Q_0$.

In Fig. 8, the axis of symmetry for pairs is shown with the red solid curve which follows equation (18). The red dashed curves above and below the axis of symmetry are given by:

$$R = \frac{L}{(1 - Q_0) + L/(0.5 \pm \beta)} \, , \tag{19}$$

where in this case $\beta = 0.4$, so that when $L \gg 1 - Q_0$ these tend towards $R = 0.1$ and 0.9 for the dashed lines. We may then make the hypothesis that if both counterparts have $0.1 \leq R \leq 0.9$ then they might both be counterparts Fusion sources and both be contributing to the radio emission. However, if one counterpart has $R < 0.1$ then we consider the other counterpart to be the sole true match.

For example, if we have two sources inside the search radius, one with $R_1 = 0.05$ and the other with $R_2 = 0.95$ (so $R_1/R_2 < 1/19$), we reject the first source, even if $L_1 > L_c$, and consider the second Fusion source as a single source and sole counterpart. Hence, all components of pairs below the lower dashed line are rejected, and all Fusion sources above the upper dashed line are now considered as singles. For this work, we take a value of $\beta = 0.4$ based on the LR and Reliability cut-off values in Fig. 6. This acceptance zone

Table 4. Results of cross-identification of ATLAS sources with FUSION sources using the LRPY code. We present the total number of radio sources, the number having Fusion coverage, the number with any Fusion counterpart within 6 arcsec and the number with high reliability Fusion counteparts per Sections 4.1 and 4.2.

Field	CDFS	ELAIS-S1	Both
Total	3078	2113	5191
With Fusion coverage	2922	2113	5035
With any Fusion candidates	2700	1936	4636
With high reliability XID	2222	1626	3848

can be narrowed or widened by decreasing or increasing β in the algorithm.

We have a relatively small subset of multiple Fusion counterparts between the dashed lines in Fig. 5 and with $L > 0.01$, but clearly if this selection is applied to much larger catalogues then a significant number of sources would be selected as such. In our case, we have 38 pairs of fusion counterparts in the CDFS field and 26 in ELAIS-S1 which makes ≈2 per cent of all radio sources with cross-identification in Table 4. Hence, from the 7×10^7 radio sources expected in EMU we might expect more than a million with multiple matches.

Using the selection rules as outlined earlier in this section, we find 64 pairs for the two fields. To explore the possibility that some of these 64 pairs of galaxies could be members of the same group of galaxies or even physically interacting, we perform a nearest-neighbour match of the Fusion sources with objects from the OzDES survey (presented in Section 2.3)

If the Fusion source is within 1 arcsec of an OzDES object, we consider it to be the same object. We found 22 out of 64 doubles to have spectroscopic redshifts of *both* galaxies which we present in Table 5. In 20 cases, both objects have similar redshifts, $\Delta z/z < 0.01$, and in two cases Fusion sources have significantly different redshift. Postage stamp images of some IRDs are given in Fig. A2 with the ATLAS radio contours overlaid on the grey-scale IR images to demonstrate these objects.

In addition, using the Great Observatories Origins Deep Survey (GOODS; Dickinson et al. 2003; Renzini et al. 2003), we find two of our CDFS IRDs have *Hubble Space Telescope* (*HST*) archive images.[3] In Figs A3 and A4, we present these images with the IR source positions marked and the radio contours overlaid. As well as the ATLAS radio contours we present contours from the deep JVLA 1.4 GHz survey of this sub-region of the CDFS (Miller et al. 2013). The *HST* images clearly indicated that these two pairs of galaxies are interacting via their disturbed morphologies and tidal tails.

It is possible that more than two Fusion sources high enough LR and reliabilities above the cut-off. In this work, we do not have a situation of this sort. However, when dealing with large datasets, we can expect cases of multiple counterparts, and an automated approach has to be elaborated for this case.

4.3 Results of cross-identification

Table 4 presents the results of our cross-identification of the ATLAS catalogue with the Fusion catalogue. As we described in Section 2.3,

---

[3] Based on observations made with the NASA/ESA *HST*, obtained from the data archive at the Space Telescope Science Institute. STScI is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555.

Table 5. Redshifts for possible IR doubles taken from OzDES, by a nearest-neighbour match between Fusion *Spitzer* and OzDES (Yuan et al. 2015) within 1 arcsec. Of these 22, 20 have pairs of galaxies with $\Delta z/z_{\rm spec} < 0.01$.

ATLAS ID	Fusion ID	Ang sep (arcsec)	OzDES ID	OzDES z	OzDES Δz
CI0069	309081	0.528	281939.8	0.6789	0.0019
	309075	0.045	00036776	0.68084	
CI0099C2	295215	0.106	0076-01223	0.3339	0.0004
	295098	0.285	2940685175	0.3344	
CI0175	333146	0.448	91-274837.9	0.1816	0.0001
	333165	0.223	32564	0.1817	
CI0191	467746	0.321	S117	0.0909	0.0001
	467716	0.121	NAO_0552_119829	0.09078	
CI0548	322386	0.188	57-280213.3	0.5368	0.0004
	322361	0.236	2940877666	0.53631	
CI0561	151844	0.013	20-283323.1	0.3402	0.006
	151810	0.106	0082-01440	0.3462	
CI0632	328609	0.231	NOAO_0334_R126091	0.32776	0.0006
	328657	0.076	NOAO_0552_126052	0.32708	
CI0633	197618	0.277	S477	0.2511	0.0005
	197687	0.258	63053	0.2516	
CI0757	171555	0.1063	0084-00302	0.6633	0.0067
	171511	0.1062	0085-00883	0.6701	
CI1000	178274	0.149	2939983811	0.3380	0.0002
	178269	1.070	49-275932.3	0.3378	
CI1036	183594	0.072	92922	1.0967	0.0007
	183614	0.119	49-275217.7	1.096	
CI1042	162527	0.119	2940728894	0.406	0.006
	162601	0.106	0084-01738	0.400	
CI0418	184058	0.115	0139-01350	0.2864	0.0105
	184031	0.642	26362	0.2759	
CI0961	147928	0.094	2940682513	0.5963	0.0004
	147978	0.106	0082-00421	0.5959	
CI1905	315366	0.106	0079-00072	0.5815	0.2892
	315305	0.106	0078-01319	0.2923	
CI1906	187271	0.189	34546	0.4357	0.2216
	187229	0.055	J033244.87	0.2141	
EI0151	215007	0.224	2970674536	0.12434	0.0004
	215061	0.191	2970674654	0.12478	
EI0455	221400	0.254	2971105989	0.198	0.0036
	221459	0.136	0092-01998	0.195	
EI0487	101702	0.144	J003459.03	0.330	0.0011
	101761	0.111	J003458.95	0.329	
EI0863	247565	0.100	0094-01686	0.217	0.0064
	247574	0.333	2971175179	0.224	
EI1034	196663	0.20	0096-00993	0.34711	0.0022
	196655	0.11	2971105849	0.3493	
EI1219	73440	0.21	2970777434	0.4001	0.0008
	73483	0.14	2970777513	0.3993	

and illustrated in Fig. 1, ∼96 per cent of total number of ATLAS sources are covered by the Fusion catalogue, which makes 2922 radio sources in the CDFS field and 2113 in the ELAIS-S1 field. So there are in total 5035 radio sources for XID with the Fusion catalogue. Not all radio sources we deal with have Fusion candidates inside of the search radius used in this work (6 arcsec). This number of 'blanks' is small consisting of 222 for CDFS and 177 for ELAIS-S1. So the number of 'candidates' [radio sources with one or more Fusion source(s) in the search area] drops to 2700 for the CDFS field and 1936 for the ELAIS-S1 field. We found that a large percentage of these candidate radio sources have just one Fusion source in the search radius ∼60 per cent (see Table 3). About 40 per cent of all (non-blank) radio sources have two or more (up to 5) Fusion sources within the search radius.
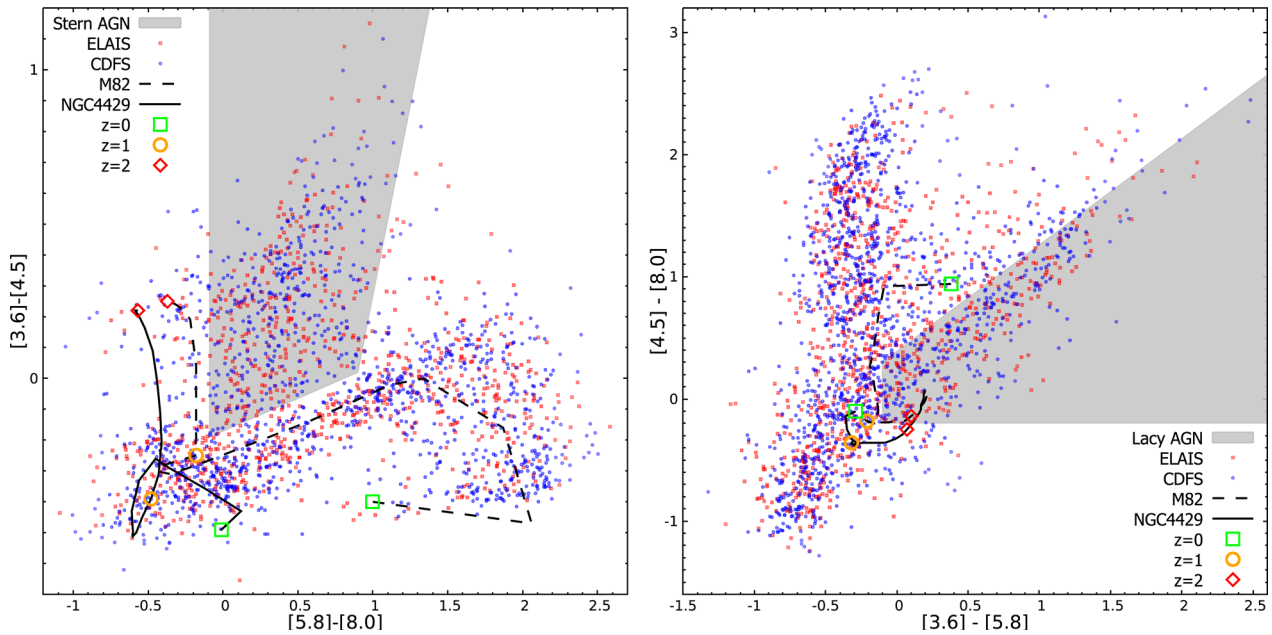
Figure 9. We present the colour–colour diagrams of the Fusion counterparts to the ATLAS sources (as determined in Section 4.1). Left: is the $[5.8] - [8.0]$ versus $[3.6] - [4.6]$ colour–colour plot. The grey shaded area shows the location of the Stern et al. (2005) selection for AGNs. Also included are the evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 2$ taken from Seymour et al. (2007). Right: using the same Fusion counterparts is the $[3.6] - [5.8]$ versus $[4.5] - [8.0]$ colour–colour with the grey shaded area showing the location of the Lacy et al. (2004) selection for AGNs. Again the evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 2$ taken from Seymour et al. (2007) are included.

Applying the LR criteria for 'single' sources and both LR and reliability criteria for the situation when two or more Fusions sources are in the search radius, we find that about ∼84 per cent of all candidates correspond to the criteria we use in this work for cross-identification in Sections 4.1 and 4.2. The ATLAS sources without secure Fusion counterparts likely have counterparts below the Fusion detection limit.

5 HOST GALAXY PROPERTIES

In this section, we use the cross-identified sources in order to examine the nature of the faint ATLAS radio sources. The primary question we wish to address is whether our radio sources are AGNs or SFGs. The advent of surveys with *Spitzer* has presented us with new methods of distinguishing between AGNs and SFGs. These methods work by being sensitive to the hot (∼1000 K) dust around the AGN causing excess emission in the mid-IR compared to regular SFGs.

5.1 IRAC colour–colour plots

Earlier work by Eisenhardt et al. (2004) presented a vertical spur in the $[3.6] - [4.6]$ versus $[5.8] - [8.0]$ colour–colour diagram which may be associated with AGNs (where the magnitude difference $[i] - [j] = -2.5 \log \left(\frac{S_i}{S_j} \right)$, where i and j are the wavelengths of the *Spitzer* IRAC bands in μm). This is also supported by Stern et al. (2005) who proposed a region in this parameter space which separates AGNs from Galactic stars and SFG. Lacy et al. (2004) presented a $[8.0] - [4.5]$ versus $[5.8] - [3.6]$ colour–colour diagram and also identified an area to select AGNs.

As we now have cross-matches (based on the selection criteria in Section 4.1) of the ATLAS catalogue against the Fusion catalogue, we will use these to determine if a source is an AGN or SFG. We

Table 6. Results of the classification of cross-identification of ATLAS sources. The first row presents the total number of XIDs with a complete set of IRAC bands. The following rows show the AGN identified by the three methods (Stern, Lacy and Flux Density ratio) followed by the total number of AGN identified, i.e. the union of the preceding three sets, and the percentage.

Field	CDFS	ELAIS-S1	Both
# With Fusion XID and all IRAC bands	1153	834	1987
# AGNs			
Stern	298	235	533
Lacy	490	358	848
Ratio	29	19	48
Total ∪	550 (48 per cent)	406 (49 per cent)	956 (48 per cent)

do so by plotting these sources in the same colour–colour diagrams as Stern et al. (2005) and Lacy et al. (2004). Not all of the cross-identifications have infrared detections in the four *Spitzer* bands and, as such, comparisons can only be made where data is available at all wavelengths. We present the numbers of radio sources with cross-matches and their break down in Table 6.

Fig. 9 contains two colour–colour plots for the Fusion counterparts to the ATLAS sources. Following Stern et al. (2005), the left plot presents the $[3.6] - [4.5]$ versus $[5.8] - [8.0]$. The plot on the right is $[8.0] - [4.5]$ versus $[5.8] - [3.6]$ as per Lacy et al. (2004). Evolutionary tracks from redshift 0 to 2 for a late type starburst galaxy (M82) and an early type galaxy (NGC 4429) are included in both plots. Markers have been placed to indicate $z = 0$, $z = 1$ and $z = 2$. This data has been taken from Seymour et al. (2007) based on the work from Devriendt, Guiderdoni & Sadat (1999). We see how these evolutionary tracks generally remain outside the Stern et al. (2005) and Lacy et al. (2004) AGN selection 'wedge' (grey

shaded areas) and neither would be selected as an AGN candidate if located below $z = 2$.

Many sources in the left-hand plot of Fig. 9 are spread along the evolutionary track of M82 to a redshift of $z = 1$ as there are very few sources in the $z = 1$ to $z = 2$ region of this track. Of note is the vertical spur in the Stern AGN zone grey shaded consistent with Eisenhardt et al. (2004) and Stern et al. (2005). In the right-hand plot there is a clear fork with the right hand arm entering the Lacy AGN zone grey shaded. This is also consistent with Lacy et al. (2004). We note that Mao et al. (2012), using the ATLAS DR1 data release and associated spectroscopic classifications, showed that many spectroscopic AGNs lay outside the Stern and Lacy wedges.

By using the selection criteria in Section 4.1 for Fusion cross-identifications in the Lacy AGN selection zone there are a total of 848 XIDs and for the Stern AGN selection zone there are a total of 533 XIDs. A total of 956 XIDs satisfy the union of the Stern and Lacy selection criteria for AGNs.

5.2 Radio to 3.6 μm flux density ratio

In order to account for the relative radio emission from RL AGN, we examine the radio to 3.6 μm flux density ratio of all the cross-matched sources with known redshifts in our sample in Fig. 10. We compare these to tracks of known sources shifted to higher redshifts (i.e. we compare the ratio of the observed frame 1.4 GHz and 3.6 μm flux densities shifted with redshift). For comparison, we include the tracks for the radio-loud and radio-quiet AGNs from Elvis et al. (1994) and the two galaxies used in the previous section (the starburst M82 and the quiescent galaxy NGC4429). The various galaxy template tracks are relatively flat although some variation is seen with redshift. We note that the AGN templates are for unobscured AGNs which do not include any potential obscuration of the AGN by dust from a torus or the host galaxy. Any obscuration would increase these flux ratios by suppressing the observed 3.6 μm flux density as it gets shifted to the NIR and optical rest-frame at higher redshift.

The redshifts for 295 sources come from the OzDES global redshift catalogue (Yuan et al. 2015, Lidman et al. in prep). We observe that most of our redshifts are at $z < 0.3$, which is due to the targeting of the brightest optical counterparts by OzDES and earlier surveys (Mao et al. 2012), although there is a tail to $z \sim 2.8$. This $z < 0.3$ grouping typically have a flux ratio from ∼0.2 to 2, below that for the starburst and quiescent galaxy tracks. Why does our group of SFGs have lower flux ratios than these two sources? To first approximation we can say that, if these are star forming galaxies, the radio emission traces the star-formation rate (SFR) and the 3.6 μm emission traces the stellar mass. Hence, the galaxies in this grouping are likely similar to M82, but with a lower specific SFR (SFR per unit stellar mass). As the ATLAS sources are selected on an SFR proxy, radio flux, and lie at higher redshift, it is likely the higher stellar masses are pulling this observed ratio down despite the higher SFRs compared to M82.

In terms of identifying which sources in this plot have radio emission powered by AGNs, we can use the Elvis RL AGN track as a guide. Allowing for uncertainty in the models and the fluxes we suggest that any radio source with a ratio greater than one third of the track from the RL AGNs, marked by the grey shaded area in Fig. 10, is likely powered by an AGN. We find only nine sources above this line which also have a redshift value. We also have 57 matches with no redshift and the 3.6 μm flux is below the detection limit. Taking their radio flux and dividing by the 3.6 μm flux limit would place these matches in the grey AGN region.
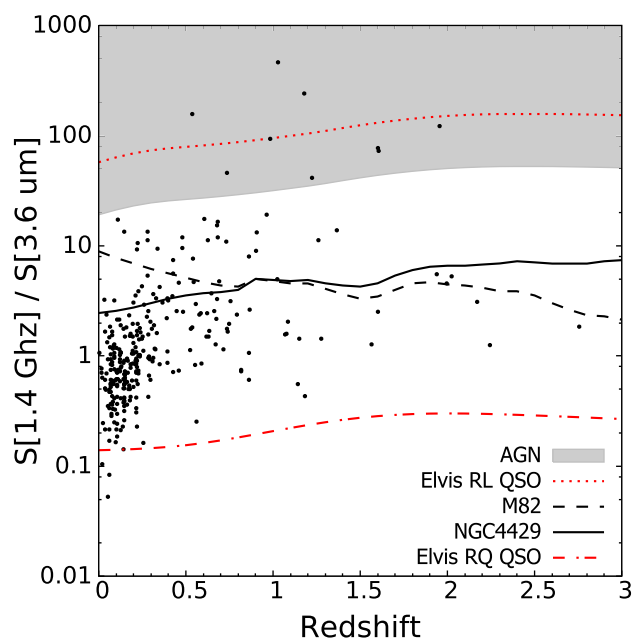

Figure 10. The ratio between the radio 1.4 GHz and Fusion 3.6 μm flux density plotted as a function of redshift for all XIDs (determined in Section 4.1). The red dotted line near the top of the figure indicates the loci of a classical radio-loud (radio-quiet) QSO Elvis et al. (1994), and the red dot–dashed line in the lower part of the figure indicates the loci for radio-quiet QSO Elvis et al. (1994). The grey shaded area denotes the population that we identify as the radio loud AGNs. Also included are the evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 3$ taken from Seymour et al. (2007).

5.3 Results of AGN identification

We have used the identification of Fusion counterparts to determine if our sample of ATLAS sources are AGNs or SFGs. We find 1987 cross-matched radio sources with flux values for all four IRAC bands. Of these (≈27 per cent) meet the Stern AGN selection criteria, and (≈43 per cent) Lacy, in addition 48 (≈2 per cent) of these sources lie above a line one-third of the RL AGN track. Taking the union of all of these across our three selection criteria, we find 956 (≈48 per cent) are possible AGNs.

6 CATALOGUE AND CODE

6.1 Catalogue description

In this section, we describe the catalogue containing the results of the ATLAS cross-identification with Fusion using the LRPY algorithm discussed in this paper. The information is divided into two tables, one for each field CDFS and ELAIS-S1. Example subsets are given in Table 7 for CDFS and Table 8 for ELAIS-S1. The cross-identification catalogue columns are organized as follows:

Column (1) - ATLAS DR3 Identification number of the radio source 'cid'

Column (2) - Right Ascension (J2000) of the radio source, decimal degrees

Column (3) - Declination (J2000) of the radio source, decimal degrees

Column (4) - Integrated Radio Flux Densities (μJy) at 1.4 GHz

Column (5) - Fusion Identification number 'swire_index_spitzer'

Column (6) - Right Ascension (J2000) of the IR candidate, decimal degrees

Table 7. ATLAS/FUSION SWIRE Cross-Identification Catalogue for the CDFS field. A description of the table is given in Section 6. (This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content, a full copy of the catalogue is available online.)

(1) ATLAS ID	(2) RA (J2000)	(3) Dec (J2000)	(4) S_{Int} (mJy)	(5) SWIRE (ID)	(6) RA_{IR} (deg)	(7) Dec_{IR} (deg)	(8) $S_{3.6\mu m}$ (μJy)	(9) $\sigma_{3.6\mu m}$ (μJy)	(10) $\log_{10}(LR)$	(11) Reliability
CI0001C1	52.1516	−28.6982	132.5	432065	52.1520	−28.6988	2565	7.07	−1.154	0.290
CI0002	51.8917	−28.7726	158.151	428929	51.8917	−28.7725	101.84	0.8	2.556	0.999
CI0003	53.5387	−28.4055	74.8	158805	53.5388	−28.4053	67.29	1.09	2.011	0.998
CI0005C1	51.9088	−28.0239	19.8	456752	51.9079	−28.0232	5.16	0.43	−6.966	0.000
CI0005C2	51.9127	−28.0357	0.692	456300	51.9127	−28.0358	24.75	0.81	1.906	0.997
CI0005C3	51.9067	−28.0251	69.6	456683	51.9071	−28.0250	200.42	1.63	1.424	0.993
CI0007	53.9722	−27.4613	118.207	63449	53.9722	−27.4612	398.33	2.01	2.458	0.999
CI0008	52.1943	−28.4379	56.6902	303864	52.1940	−28.4383	68.82	0.88	0.693	0.966
CI0009	53.8646	−27.3308	95.6314	209688	53.8646	−27.3307	157.62	1.77	2.610	0.999
CI0010	53.6121	−27.7338	55.7524	190007	53.6121	−27.7338	201.51	1.55	2.546	0.999

Table 8. ATLAS/FUSION SWIRE Cross-Identification Catalogue for the ELAIS-S1 field. A description of the table is given in Section 6. (This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content, a full copy of the catalogue is available online.)

(1) ATLAS ID	(2) RA (J2000)	(3) Dec (J2000)	(4) S_{Int} (mJy)	(5) SWIRE ID	(6) RA_{IR} (deg)	(7) Dec_{IR} (deg)	(8) $S_{3.6\mu m}$ (μJy)	(9) $\sigma_{3.6\mu m}$ (μJy)	(10) $\log_{10}(LR)$	(11) Reliability
EI0001	9.06966	−43.15964	160.032	87322	9.06957	−43.15968	103.24	1.53	2.565	0.999
EI0002C2	7.87001	−43.68909	193.85	373161	7.87013	−43.68917	46.44	0.96	2.038	0.998
EI0003	8.28863	−43.99076	69.6057	220919	8.28844	−43.99079	141.9	1.8	2.554	0.999
EI0004C1	8.67872	−43.50959	49.97	237221	8.67851	−43.50948	89.18	1.45	2.352	0.999
EI0004C3	8.67261	−43.51230	0.538	237403	8.67237	−43.51141	228.49	2.27	−3.161	0.003
EI0005	8.01844	−44.19189	35.5887	350583	8.01828	−44.19192	48.61	0.84	2.042	0.998
EI0006	8.20550	−44.36404	39.7013	341374	8.20543	−44.36403	373.25	2.28	2.420	0.999
EI0007	9.34721	−44.37919	44.6797	159474	9.34722	−44.37909	51.05	0.81	2.308	0.999
EI0008	9.19112	−43.09654	30.4775	87851	9.19100	−43.09654	10.76	0.68	1.705	0.996
EI0009C1	9.32550	−44.50327	49.35	162876	9.32449	−44.50391	133.17	1.3	−7.356	0.000

Column (7) - Declination (J2000) of the IR candidate, decimal degrees

Column (8) - IR Flux Density at 3.6 μm (μJy)

Column (9) - IR Flux Density Uncertainty at 3.6 μm (μJy)

Column (10) - Log_{10} of LR of the IR candidate

Column (11) - Reliability of the IR candidate

These two tables are available in their entirety including all Fusion sources within 6 arcsec of an ATLAS source in a machine-readable format in the supplementary material. No filtering on reliability or LR has been undertaken. Column 10 has been presented with its $\log_{10}$ value to make the column width manegeable, also Column 11 is presented to three decimal places. Both columns in the supplementary material will be presented to a higher precision. From our findings in Section 4.1, we recommend the following selection criteria for accepting identifications: $L_c \geq 0.01$ and the reliability is greater than R limit given by using equation (19) with $\beta = 0.4$. The data is also available as a series of normalized relational database tables where, by using the index columns 'cid' and 'swire_index_spitzer' as the relationship to join the tables, it is possible for the reader to construct their own version of the catalogue or work with the data in other ways.

6.2 The LRPY algorithm

The method described in Section 3 along with the selection rules presented in Section 4 have been coded in Python and the full set

of files is available from github[4] under a GNU General Public License (GPL) V3.0. We intend to use this code on the ASKAP/EMU survey as one of the several complimentary methods to determine counterparts to the estimated 7×10^7 faint radio sources. Anyone is permitted to use this code for research purposes and we ask that they cite this paper.

7 CONCLUSIONS

We have developed a PYTHON based code to allow cross-matching of lower resolution survey data to higher source density, higher resolution catalogues implementing the LR technique. Our motivation is to use this code as one of several techniques to cross-match the ASKAP/EMU survey to higher resolution optical/near-IR data. This code, LRPY, accurately accounts for sources below the detection limit in an automated fashion. It is suitable for any cross-matching of catalogues with significantly mismatched resolutions. We have added an extension to the LR algorithm to identify potential multiple matches to a lower resolution source, i.e. doubles. We make this code publicly available to the community. Future updates to this code will include the (optional) ability to handle more complex radio sources (e.g. radio doubles and triples).

[4] https://github.com/sdweston/LikelihoodRatio

We have used LRPY to cross-match the ATLAS DR3 radio survey to the *Spitzer* Data Fusion catalogues with a search radius of 6 arcsec. Setting the possible false detection rate of 5 per cent, and using the new resultant LR and reliability cut-off selection criteria described in Section 4.1, rather than simply a reliability threshold, we obtain 2222 (82 per cent) matches in the CDFS field and 1626 (83 per cent) in the ELAIS-S1 field. Of these matches, we obtain a subset with detections in all four IRAC bands consisting of 2133 sources (1243 for CDFS and 890 for ELAIS-S1). A much smaller subset has redshifts consisting of 295 sources (186 CDFS and 109 for ELAIS-S1). Hence, from this work we present a new catalogue listing ATLAS DR3 radio sources with their *Spitzer* Data Fusion counterparts including the likelihood and reliability figure to allow the reader to use their own selection criteria as required.

We have identified a subset of 64 *Spitzer* Data Fusion doubles (38 in CDFS and 26 in ELAIS-S1), i.e. radio sources with two *Spitzer* Data Fusion candidates meeting our selection criteria in Section 4.2. From these pairs we find 22 with a redshift for each member; we find 20 of these have a $\Delta z/z < 0.01$ and we identify them as potentially interacting galaxies contributing to the one radio source. Two pairs are confirmed as interacting galaxies from deep *HST* imaging.

Taking the available *Spitzer* Data Fusion colour–colour information for the possible matches we present their characteristics with respect to the Stern and Lacy AGN selection criteria. For the two fields, we identify 848 AGN radio sources using the Lacy selection criteria which is ≈ 42 per cent of the candidates, and 533 if using the Stern criteria which is ≈ 27 per cent. Also, we examine the radio to 3.6 μm flux density ratio as a function of redshift to search for RL AGNs. We find a cluster of objects at $z < 0.3$ and flux ratio 0.2–2 which we surmise are SFG, but with a lower relative SFR to stellar mass than the modelled M82 track. We propose a cut-off where the flux ratio is greater than one third of the value of the RL AGN track. Taking the union of all three AGNs selection criteria we identify 956 as ≈ 48 per cent possible AGNs.

ACKNOWLEDGEMENTS

REFERENCES

Banfield J. K. et al., 2016, MNRAS, 460, 2376
Bonzini M. et al., 2012, ApJS, 203, 15
Ciliegi P., Zamorani G., Hasinger G., Lehmann I., Szokoly G., Wilson G., 2003, A&A, 398, 901
Ciliegi P., Zamorani G., Bondi M., Pozzetti L., Bolzonella M., Rizzo L. G. D., 2005, A&A, 441, 879

Dark Energy Survey Collaboration et al., 2016, MNRAS, 460, 1270
de Ruiter H. R., Willis A. G., Arp H. C., 1977, A&AS, 28, 211
Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, A&A, 350, 381
Dickinson M., Giavalisco M., GOODS Team, 2003, in Bender R., Renzini A., eds, The Mass of Galaxies at Low and High Redshift: The Great Observatories Origins Deep Survey. Springer-Verlag, Berlin, p. 324
Downes A. J. B., Peacock J. A., Savage A., Carrie D. R., 1986, MNRAS, 218, 31
Eisenhardt P. R. et al., 2004, ApJS, 154, 48
Elvis M. et al., 1994, ApJS, 95, 1
Fan D., Budavári T., Norris R. P., Hopkins A. M., 2015, MNRAS, 451, 1299
Fazio G. G. et al., 2004, ApJS, 154, 10
Fleuren S. et al., 2012, MNRAS, 423, 2407
Franzen T. M. O. et al., 2015, MNRAS, 453, 4020
Hales C. A. et al., 2014, MNRAS, 441, 2555
Huynh M. T., Jackson C. A., Norris R. P., Prandoni I., 2005, AJ, 130, 1373
Ivison R. J. et al., 2007, MNRAS, 380, 199
Lacy M. et al., 2004, ApJS, 154, 166
Lonsdale C. J. et al., 2003, PASP, 115, 897
Mao M. Y. et al., 2012, MNRAS, 426, 3334
Middelberg E. et al., 2008, AJ, 135, 1276
Miller K. I. et al., 2013, ApJS, 205, 13
Norris R. P., 2012, in Tuffs R. J., Popescu C. C., eds, IAU Symp. 284, The Spectral Energy Distribution of Galaxies. Cambridge University Press, Cambridge, p. 489
Norris R. P., Tingay S., Phillips C., Middelberg E., Deller A., Appleton P. N., 2007, MNRAS, 378, 1434
Prandoni I., Seymour N., 2015, Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14), Giardini Naxos, Italy, p. 67
Renzini A. et al., 2003, in Bender R., Renzini A., eds, The Mass of Galaxies at Low and High Redshift ESO for GOODS' Sake. Springer-Verlag, Berlin, p. 332
Richter G. A., 1975, Astron. Nachr., 296, 65
Rowan-Robinson M. et al., 2010, MNRAS, 409, 2
Seymour N. et al., 2007, , ApJS, 171, 353
Smith D. J. B. et al., 2011, MNRAS, 416, 857
Stern D. et al., 2005, ApJ, 631, 163
Sutherland W., Saunders W., 1992, MNRAS, 259, 413
Vaccari M., 2015, in The Many Facets of Extragalactic Radio Surveys: Towards New Scientific Challenges The *Spitzer* Data Fusion: Contents, Construction and Applications to Galaxy Evolution Studies. p. 27
Vaccari M., 2016, in Napolitano N., Longo G., Marconi M., Paolillo M., Iodice E., eds, Astrophys. Space Sci. Proc. Vol. 42, The Universe of Digital Sky Surveys. Springer International Publishing, Switzerland, p. 71
Vaccari M. et al., 2010, A&A, 518, L20
Vaccari M., Marchetti L., Gonzalez-Solares E., 2010, *Spitzer* Data Fusion: A Spitzer-Selected Multi-Wavelength Catalog, available at http://mattiavaccari.net/df/
Van der Heyden K., Jarvis M. J., 2010, MIGHTEE proposal to MeerKAT
Werner M. W. et al., 2004, ApJ, Suppl. Ser., 154, 1
Yuan F. et al., 2015, MNRAS, 452, 3047

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

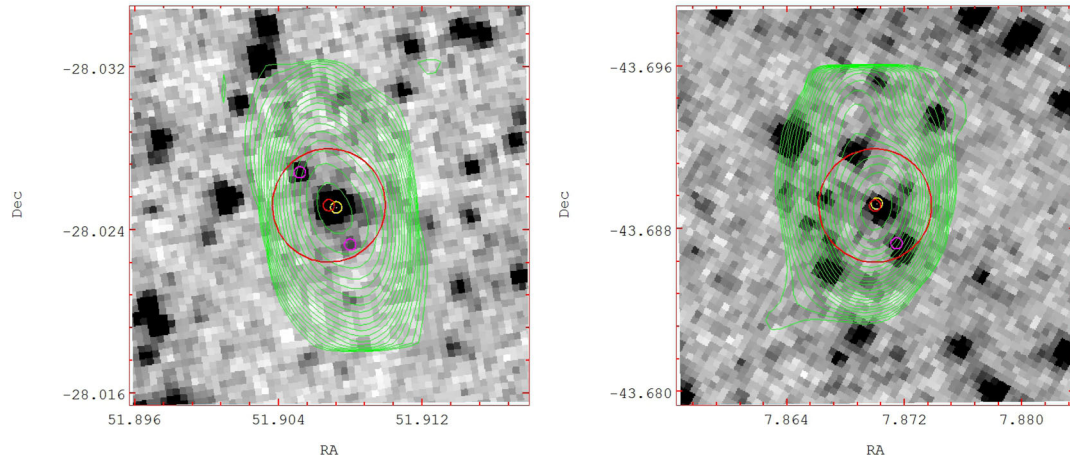## APPENDIX POSTAGE STAMP IMAGES OF XID EXAMPLES

Figure A1. Example postage stamps of the cross-identifications of ATLAS sources with the Fusion catalogue. We overlay ATLAS contours (starting at 3σ and then spaced by a factor of 2) on grey-scale 3.6 μm images centred on the radio source position to visually demonstrate the LRPY XID. The small red circle denotes the ATLAS radio candidate position and the larger red circle is the 6 arcsec search radius. The small yellow circle denotes the candidate Fusion counterpart position with the selection criteria give in Section 4.1; and the small magenta circles are other candidate Fusion counterparts within the search radius that have a reliability and LR outside the selection criteria. Each image is 70 arcsec square. On the left is ATLAS source CI0005C3 and on the right is ATLAS source EI0002C2
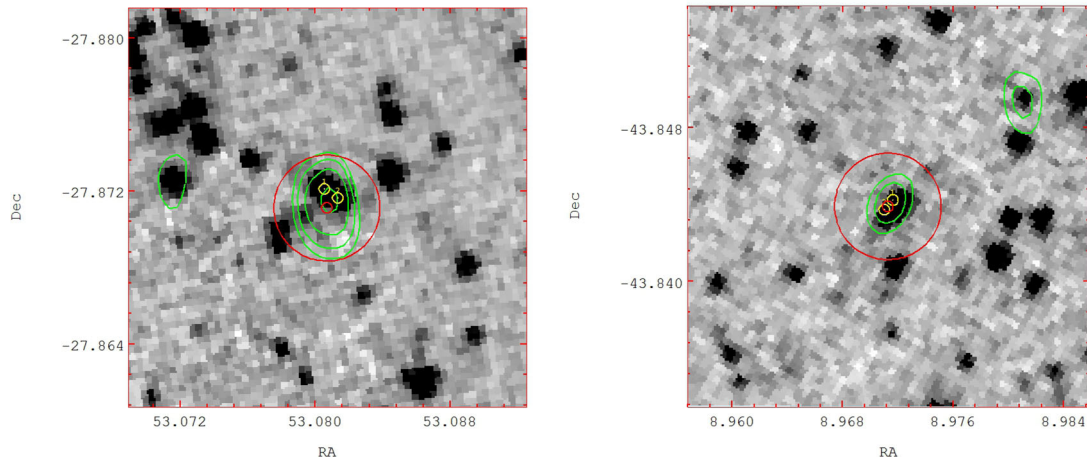

Figure A2. Example postage stamps of Infared Double (IRD) candidates with the radio contours (starting at 3σ and then spaced by a factor of 2) overlaid on grey-scale 3.6 μm images centred on an ATLAS radio source position. The small red circle denotes the ATLAS radio candidate position and the larger red circle is the 6 arcsec search radius. The small yellow circles denote the possible IRD candidate positions using the selection criteria in Section 4.2. Each image is 75 arcsec square. On the left is ATLAS source CI1036 and on the right is ATLAS source EI1034 with the two IR candidates which are in Table 5 with similar z.
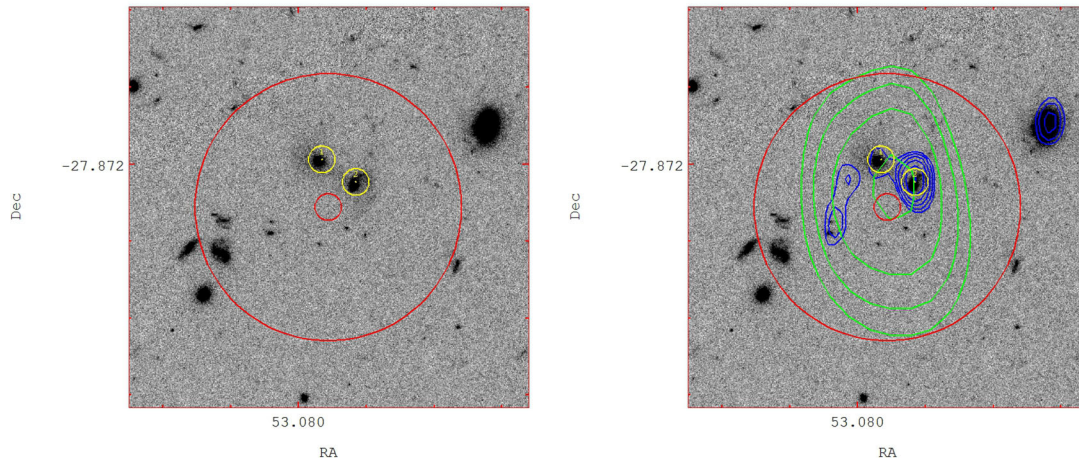
Figure A3. Postage stamp HST image centred on the ATLAS source CI1036 position, each image is 30 arcsec square. On the left-hand image, the Infared Double (IRD) candidates are marked with open yellow circles, and the ATLAS radio position with a small red open circle, the larger open red circle is the 6 arcsec search radius. The IR Source marked 1 is $z = 1.0967$ and source 2 is $z = 1.0960$, where $\delta z = 0.0007$. The right-hand image has the ATLAS radio contours (starting at 3σ and then spaced by a factor of 2) overlaid in green, in addition the VLA radio contours in blue (using the same contour spacing as for ATLAS).
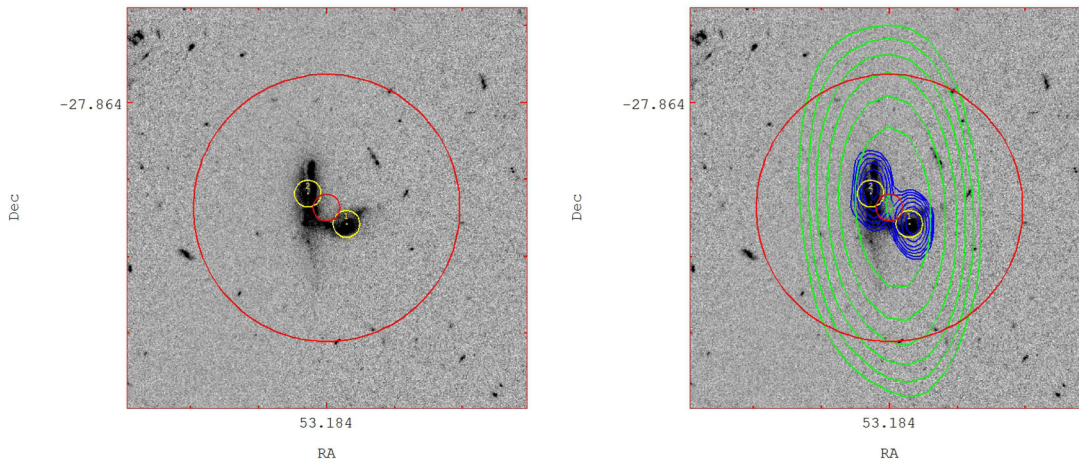

Figure A4. Postage stamp HST image centred on the ATLAS source CI0418 position; each image is 30 arcsec square. On the left-hand image the Infared Double (IRD) candidates are marked with open yellow circles, and the ATLAS radio position with a small red open circle; the larger open red circle is the 6 arcsec search radius. The IR Source marked 1 is $z = 0.2864$ and source 2 is $z = 0.2759$, where $\delta z = 0.0105$. The right-hand image has the ATLAS radio contours (starting at 3σ and then spaced by a factor of 2) overlaid in green; in addition the VLA radio contours in blue (using the same contour spacing as for ATLAS).

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