Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **464,** 1306–1332 (2017)
Advance Access publication 2016 September 21

doi:10.1093/mnras/stw2396

The Large Area Radio Galaxy Evolution Spectroscopic Survey (LARGESS): survey design, data catalogue and GAMA/WiggleZ spectroscopy

John H. Y. Ching,[1] Elaine M. Sadler,[1★] Scott M. Croom,[1★] Helen M. Johnston,[1] Michael B. Pracy,[1] Warrick J. Couch,[2] A. M. Hopkins,[2] Russell J. Jurek[3] and K. A. Pimbblet[4,5]

[1] *Sydney Institute for Astronomy, School of Physics, University of Sydney, NSW 2006, Australia*
[2] *Australian Astronomical Observatory, PO Box 915, North Ryde, NSW 1670, Australia*
[3] *Australia Telescope National Facility, CSIRO, Epping, NSW 1710, Australia*
[4] *Milne Centre for Astrophysics, University of Hull, Cottingham Road, Kingston-upon-Hull, HU6 7RX, UK*
[5] *School of Physics and Astronomy, Monash University, Clayton, VIC 3800, Australia*

Accepted 2016 September 20. Received 2016 September 18; in original form 2016 April 21

ABSTRACT

We present the Large Area Radio Galaxy Evolution Spectroscopic Survey (LARGESS), a spectroscopic catalogue of radio sources designed to include the full range of radio AGN populations out to redshift $z \sim 0.8$. The catalogue covers ~800 deg^2 of sky, and provides optical identifications for 19 179 radio sources from the 1.4 GHz Faint Images of the Radio Sky at Twenty-cm (FIRST) survey down to an optical magnitude limit of $i_{\mathrm{mod}} < 20.5$ in Sloan Digital Sky Survey (SDSS) images. Both galaxies and point-like objects are included, and no colour cuts are applied. In collaboration with the WiggleZ and Galaxy And Mass Assembly (GAMA) spectroscopic survey teams, we have obtained new spectra for over 5000 objects in the LARGESS sample. Combining these new spectra with data from earlier surveys provides spectroscopic data for 12 329 radio sources in the survey area, of which 10 856 have reliable redshifts. 85 per cent of the LARGESS spectroscopic sample are radio AGN (median redshift $z = 0.44$), and 15 per cent are nearby star-forming galaxies (median $z = 0.08$). Low-excitation radio galaxies (LERGs) comprise the majority (83 per cent) of LARGESS radio AGN at $z < 0.8$, with 12 per cent being high-excitation radio galaxies (HERGs) and 5 per cent radio-loud QSOs. Unlike the more homogeneous LERG and QSO sub-populations, HERGs are a heterogeneous class of objects with relatively blue optical colours and a wide dispersion in mid-infrared colours. This is consistent with a picture in which most HERGs are hosted by galaxies with recent or ongoing star formation as well as a classical accretion disc.

Key words: catalogues – surveys – galaxies: active – radio continuum: galaxies.

1 INTRODUCTION

Over the past 15 yr, large surveys at optical, infrared and radio wavelengths have allowed us to make significant progress in understanding the typical radio properties of galaxies in the local and distant Universe. Two large-area radio surveys carried out by the Very Large Array (VLA) operated by the National Radio Astronomy Observatory (NRAO), the Faint Images of the Radio Sky at Twenty-cm (FIRST; Becker, White & Helfand 1995) and the NRAO VLA Sky Survey (NVSS; Condon et al. 1998) have been particularly influential. Both are 1.4 GHz continuum radio surveys covering a large fraction of the sky down to milli-Jansky flux densities. The high resolution and positional accuracy of the FIRST survey is complemented by the lower resolution of NVSS, which has better surface brightness sensitivity. Several studies (e.g. Sadler et al. 2002; Hopkins et al. 2003; Best et al. 2005a; Mauch & Sadler 2007; Best & Heckman 2012) have matched NVSS and FIRST radio sources to counterparts in the optical or infrared. These optical/infrared identifications, combined with spectroscopic information such as redshifts, emission line and absorption line measurements, have advanced our understanding of the physical processes responsible for radio emission from nearby galaxies.

For extragalactic radio sources, the radio continuum emission may arise from either an active galactic nucleus (AGN) or processes related to star formation. In star-forming galaxies, the

---

★ E-mail: elaine.sadler@sydney.edu.au (EMS); scott.croom@sydney.edu.au (SMC)

Published by Oxford University Press on behalf of the Royal Astronomical Society

observed radio emission is usually dominated by synchrotron emission from relativistic electrons accelerated by supernova remnants in H II regions, with a smaller contribution from thermal free–free emission (Condon 1992). The short-lived massive stars in the H II regions of star-forming galaxies photoionize the surrounding gas and produce a characteristic pattern of emission lines in the observed spectrum.

Spectroscopic studies of radio AGN reveal two main populations: those with prominent optical emission lines, and those with weak or no emission lines (Longair & Seldner 1979; Laing et al. 1994). We follow current practice and refer to the first (strong emission-line) population as high-excitation radio galaxies (HERGs) and the second as low-excitation radio galaxies (LERGs). The difference between these two populations is thought to reflect differences in the accretion efficiency of gas on to the central black hole (Hardcastle, Evans & Croston 2007). A comprehensive review of the properties of the two classes of radio AGN is given by Heckman & Best (2014).

In the current paradigm, the HERGs undergo *cold-mode* (also known as *radiative mode*) accretion, characterized by a high accretion efficiency such that gas is accreted rapidly on to the galaxy's central black hole. This allows the formation of a radiatively efficient accretion disc that photoionizes the surrounding gas to produce the observed high-excitation emission lines. The term cold-mode refers to the past temperature of the gas, which in this case has never reached the virial temperature of the halo (Kereš et al. 2009). The LERGs on the other hand undergo *hot-mode* (also known as *jet-mode*) accretion, where the gas has at least reached the virial temperature in the past and is generally cooling from a surrounding hot X-ray corona. This is an inefficient accretion process without a radiatively efficient accretion disc, so the optical spectra of LERGs show weak or no emission lines.

Hot-mode accretion is expected to occur in high halo-mass systems ($>$2–3 $\times$ 10^{11} $\mathrm{M}_{\odot}$; Kereš et al. 2009), particularly at low redshift, and cold-mode accretion in lower mass systems over a wider range in redshift (Hardcastle et al. 2007; van de Voort et al. 2011). Recent observational studies of the properties (Best et al. 2005b; Smolčić 2009; Janssen et al. 2012), environments (Best 2004; Bardelli et al. 2010; Gendre et al. 2013; Sabater, Best & Argudo-Fernández 2013) and evolution (Smolčić et al. 2009; Best & Heckman 2012) of HERGs and LERGs appear to confirm this picture, showing that HERGs are typically found in lower mass galaxies with younger stellar populations, and in poorer environments than the LERGs, which are typically in the most massive galaxies, with an old stellar population, and found in rich environments.

The most powerful radio sources are known to undergo strong cosmic evolution, with their volume density at redshift $z \sim 2$ being up to a thousand times higher than it is today (e.g. Longair 1966; Dunlop & Peacock 1990). The cosmic evolution of lower power radio AGN appears to be much less rapid (Sadler et al. 2007; Donoso, Best & Kauffmann 2009; Simpson et al. 2012), but is only just starting to be mapped out separately for the HERG and LERG sub-populations beyond the local Universe (Best et al. 2014).

Our aim in undertaking the work described in this paper was to produce a new, large and complete spectroscopic radio-source catalogue that would allow us to track the HERG and LERG populations in detail over a wide range in radio luminosity back to at least redshift $z \sim 0.8$ (i.e. a lookback time almost half the age of the Universe) as well as studying the radio galaxy and radio-loud QSO populations across a common range in redshift. There is growing evidence that the redshift evolution of the HERG and LERG populations is very different (e.g. Best & Heckman 2012; Simpson et al. 2012; Best et al. 2014), and that observed luminosity-dependent cosmic evolution of the radio luminosity function is driven mainly by the different cosmic evolution of these two populations (Heckman & Best 2014).

One key motivation for this new study arose from earlier work on the evolving radio AGN luminosity function carried out by Sadler et al. (2007) and Donoso et al. (2009). These authors used relatively large samples of radio-detected AGN (391 objects in the Sadler et al. (2007) spectroscopic sample; 14 453 objects with photometric redshifts in the larger area Donoso et al. (2009) sample) to measure radio luminosity functions in the redshift range $0.4 < z < 0.7$ with unprecedented accuracy. Both samples were photometrically selected to target luminous red galaxies (LRGs; Eisenstein et al. 2001) but exclude blue galaxies with ongoing star formation. Sadler et al. (2007) explicitly noted that the rate of cosmic evolution measured for low-power radio galaxies in their study was only a lower limit, since the LRG sample they used had a strict colour cutoff, whereas no such colour restriction was applied to the $z \sim 0$ radio galaxy sample used as the local benchmark. By compiling a new sample of distant radio AGN without any pre-selection on colour, we wanted to find out whether there is indeed a significant population of 'blue' radio galaxies in the distant Universe and (if so) how their properties compare with the better-studied population of 'red' radio galaxies.

The data catalogue presented in this paper includes over 10 000 spectroscopically observed radio sources, with a median redshift of $z \sim 0.44$ for the radio AGN which make up $\sim$85 per cent of the sample. Our sample of 2281 radio-source spectra at $0.5 < z < 1$ represents an order-of-magnitude increase over previous spectroscopic samples in this redshift range. For example, the recent Best et al. (2014) measurement of the radio luminosity function out to $z = 1$ used a catalogue of 211 radio-loud AGN at $0.5 < z < 1.0$, while the Simpson et al. (2012) measurement used $\sim$100 spectroscopically observed objects in the same redshift range (supplemented by a similar number of photometric redshift estimates). A companion paper by Pracy et al. (2016) uses the data set presented here to make new measurements of the evolving radio luminosity functions of HERGs and LERGs out to redshift $z \sim 0.8$.

We describe the optical and radio catalogues used to compile our sample in Section 2 and the radio-optical matching process in Sections 3 and 4. The spectroscopic follow-up programme is discussed in Section 5, and the related completeness analysis presented in Section 6. Section 7 describes the identification of star-forming galaxies and the classification of high- and low-excitation radio galaxies, and Section 8 presents the full data catalogue. The full sample and some sub-samples are characterized in Section 9, while Section 10 compares the properties of a matched sample of HERG and LERG host galaxies. Finally, we present a summary of the LARGESS sample properties in Section 11.

Throughout this paper we adopt the cosmological parameters $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.7$ and $\Omega_\mathrm{m} = 0.3$. All optical magnitudes are corrected for Galactic dust extinction and k-corrected using the KCORRECT code (Blanton & Roweis 2007). An analysis of a subset of the LARGESS-GAMA (Galaxy And Mass Assembly) sample (Hardcastle et al. 2013) found a mean radio spectral index (between 325 MHz and 1.4 GHz) of $\alpha = -0.7$ (where $S_\nu \propto \nu^\alpha$), and we adopt this value to calculate a radio k-correction.

Table 1. Regions covered by optical spectroscopic surveys. The coverage and overlap was calculated using the Virtual Observatory footprint service (Budavári et al. 2007, http://www.voservices.net/footprint).

	Field	R.A. (deg)		δ (deg)		Total area	FIRST-SDSS overlap		Spectral completeness to $i = 20.5$	
Survey	ID	min	max	min	max	(deg^2)	(deg^2)	Fraction	Spectrum observed	Redshift success rate
WiggleZ										
	0h	350.1	359.1	− 13.4	+1.8	136.0	44.7	0.33	55 per cent	89 per cent
	1h	7.5	20.6	− 3.7	+5.3	118.3	32.7	0.28	64 per cent	95 per cent
	3h	43.0	52.2	− 18.6	− 5.7	116.0	7.2	0.06	55 per cent	86 per cent
	9h	133.7	148.8	− 1.0	+8.1^a	137.8	136.7	0.99	71 per cent	84 per cent
	11h	153.0	172.0	− 1.0	+8.0	172.1	172.1	1.00	66 per cent	84 per cent
	15h	210.0	230.0	− 3.0	+7.0	201.7	200.0	0.99	63 per cent	88 per cent
	22h	320.4	330.2	− 5.0	+4.8	96.2	24.5	0.25	73 per cent	88 per cent
GAMA										
	9h	129.0	141.0	− 1.0	+3.0	48.2	48.2	1.00	91 per cent	94 per cent
	12h	174.0	186.0	− 2.0	+2.0	48.2	48.2	1.00	86 per cent	88 per cent
	15h	211.5	223.5	− 2.0	+2.0	48.2	48.2	1.00	86 per cent	89 per cent
2SLAQ^b										
	–	123.0	230.0	− 1.259	+0.840	325.0	301.6	0.93	65 per cent	91 per cent
	–	309.0	59.70	− 1.259	+0.840	347.9	224.3	0.64	57 per cent	93 per cent

Notes. [a]The official WiggleZ limit for this region has a maximum $\delta = 8.0$; the additional 0.1 was mistakenly put into the original search. However, since the WiggleZ pointings included this extra small area, we include these objects as well.

[b]The actual 2SLAQ regions are several small strips along the equatorial region, but for simplicity we adopt the two large pseudo-2SLAQ strips shown here.

2 OPTICAL AND RADIO CATALOGUES

2.1 The SDSS photometric sample

The Sloan Digital Sky Survey (SDSS; York et al. 2000) is a large imaging and spectroscopic survey, covering five optical bands: *ugriz* (Fukugita et al. 1996). We used the Sixth Data Release of the SDSS (SDSS DR6, Adelman-McCarthy et al. 2008), which contains images and parameters for about 287 million objects over an area of 9583 deg^2.

The SDSS 95 per cent detection repeatability for stars in the *i*-band is at 21.3 mag (Stoughton et al. 2002). We adopted a more conservative limiting *i*-band extinction-corrected magnitude limit of $i_{\rm mod} \le 20.5$ for our optical sample, since observational constraints made it difficult for us to obtain reliable redshifts for objects fainter than this.

Our optical catalogue covers the sky area defined in Table 1 and shown in Fig. 1. This region contains over 8 million SDSS DR6 sources, along with SDSS optical spectra for objects brighter than the SDSS spectroscopic survey limit of $r_{\rm pet} < 17.7$ mag. It also overlaps with several other large spectroscopic surveys that probe to fainter magnitude limits (and higher redshifts) than SDSS: 2SLAQ (Cannon et al. 2006), GAMA (Driver et al. 2011) and WiggleZ (Drinkwater et al. 2010).

In collaboration with the GAMA and WiggleZ teams, we were able to make additional spectroscopic observations (beyond the planned public surveys) for radio-selected objects in the GAMA and WiggleZ survey regions, as discussed in more detail in Section 5.

2.2 The FIRST and NVSS radio surveys

FIRST (Becker et al. 1995) and NVSS (Condon et al. 1998) are 1.4 GHz continuum surveys carried out on the Very Large Array (VLA). The FIRST survey covered over 9000 deg^2 of the Northern (8444 deg^2) and Southern (611 deg^2) Galactic caps, mainly overlapping with the SDSS coverage.

The FIRST survey used the VLA B-configuration that provides a resolution of 5 arcsec full width at half-maximum (FWHM), with a typical root-mean-square noise ($\sigma_{\rm rms}$) of ∼0.15 mJy. The positional accuracy of FIRST sources is <1 arcsec at the survey threshold. The typical detection threshold of the FIRST survey is ∼1 mJy, though co-added observations at two epochs along the equatorial region (R.A. = $21^{\rm h}\!.3$ to $3^{\rm h}\!.3$, $\delta = -1^\circ$ to $+1^\circ$) enabled the detection threshold to drop to ∼0.75 mJy. We use the 2008 July release of the FIRST catalogue, which only contains sources with peak flux density (after correcting for CLEAN bias) greater than five times the local $\sigma_{\rm rms}$ at that point (i.e. $S^{\rm FIRST}_{\rm peak} - 0.25 > 5\sigma_{\rm rms}$) and peak flux density $S^{\rm FIRST}_{\rm peak} \ge 0.75$ mJy.

The NVSS covers the sky north of $\delta = -40^\circ$. The NVSS observations were carried out in the D and DnC configurations to provide a resolution of 45 arcsec FWHM. The lower resolution provides better surface-brightness sensitivity than the FIRST survey, but with poorer positional accuracy. The typical rms noise in the NVSS images is ∼0.45 mJy beam^{-1} with a catalogue completeness limit of ∼2.5 mJy.

3 FIRST-SDSS MATCHING

The techniques for matching FIRST and SDSS sources are now well established at low redshift (e.g. Best et al. 2005a; Sadler et al. 2007; Best & Heckman 2012). Our approach is similar, except that we are matching to a fainter optical limit than earlier studies. For example, the surface density of galaxies in our $i_{\rm mod} \le 20.5$ optical sample is ∼9,300 deg^{-2}, i.e. over 50 times higher than the ∼170 deg^{-2} surface density of the Best & Heckman (2012) sample.

3.1 Identifying multi-component FIRST sources

Around 10 per cent of FIRST radio sources have complex, extended radio morphology resolved into several components in the FIRST catalogue (e.g. Ivezić et al. 2002). To identify the optical counterparts of these extended sources, we combine the collapsing technique introduced by Cress et al. (1996) and refined by

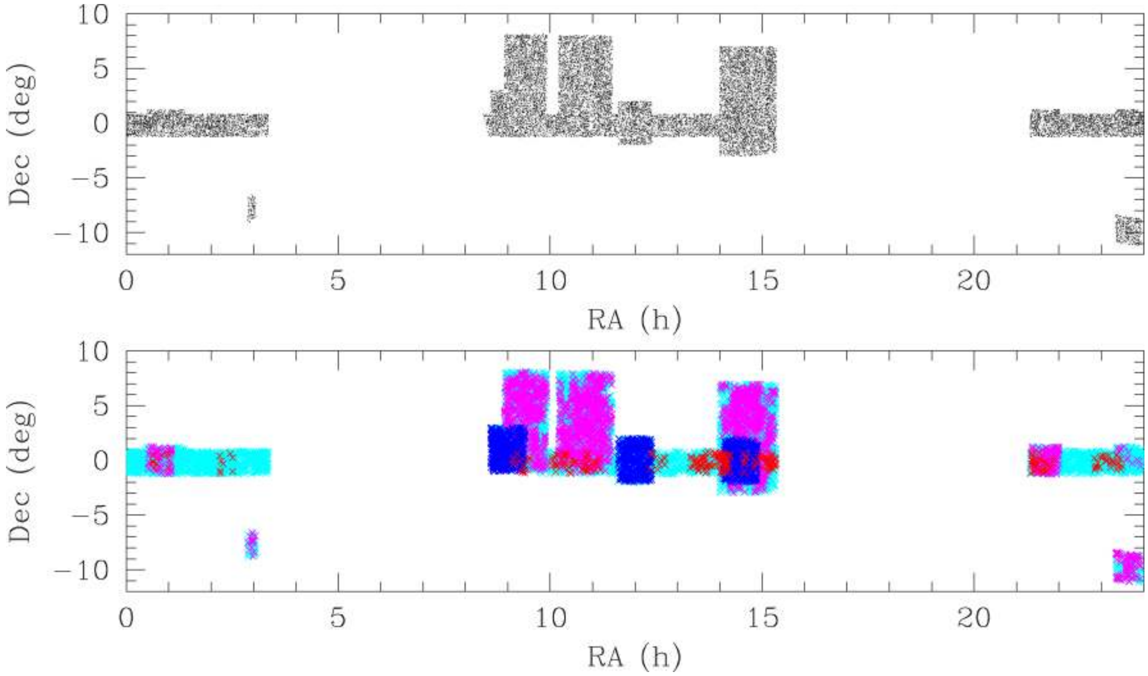


Figure 1. Sky area covered by the LARGESS radio sample: (top) sky distribution of the full photometric catalogue of 19 179 radio-sources matched to optical objects with $i < 20.5$ mag, (bottom) distribution of the 10 764 objects in the final spectroscopic catalogue that currently have good-quality optical spectra and redshifts. Points are colour-coded according to the source of the redshift measurement: GAMA (dark blue), WiggleZ (magenta), SDSS (cyan), 2SLAQ (red).

Magliocchetti et al. (1998) with the tiered algorithm used by several authors (Best et al. 2005a; Sadler et al. 2007; Donoso et al. 2009). We start by identifying the most complex multi-component sources, and then work down to simpler systems with fewer radio components, where an optical identification is more straightforward.

Cress et al. (1996) showed that about 30 per cent of all FIRST sources lie within 72 arcsec of another FIRST source, and considered these to be mainly genuine associations. Magliocchetti et al. (1998) later showed that some of the Cress et al. (1996) groups were actually unrelated sources that happen to lie close in projection on the sky. To reduce the number of spurious matches, Magliocchetti et al. (1998) applied additional constraints to decide whether or not a group of FIRST sources was part of a single system. Their constraints were motivated by known properties of radio sources, such as the ratio of the integrated flux density between the lobes and the flux-separation relation (Oort 1987) for extended sources.

Following Cress et al. (1996), we identified and grouped all FIRST sources with a separation of ≤72 arcsec on the sky (groups can span >72 arcsec in total). We then applied a range of further tests to groups of two or more sources to determine whether they were likely to be associated with a common optical counterpart. We used Monte Carlo techniques both to set appropriate selection parameters and to estimate the reliability of our final set of matches, as described in Section 3.6.

3.2 Visual matching of complex sources

We visually inspected all groups of *four or more FIRST sources*, since these are too complex for reliable automated matching. Fig. 2 shows some examples of these complex source groups. As noted below, visual matching was also used for some groups with two or three FIRST components.

The information used for visual matching included the SDSS *i*-band image, FIRST contours and/or grey-scale image, NVSS contours and/or grey-scale image, positions of FIRST sources in the field and positions of SDSS sources with $i_{\rm mod} \leq 20.5$. The user selected the most appropriate optical counterpart (which may or may not be in our $i_{\rm mod} \leq 20.5$ optical catalogue) for each FIRST source, and assigned a quality code (P-value), ranging from 1 to 4, to quantify the confidence of each match. Optical identifications with $P \geq 3$ are considered reliable enough to use in later analysis.

We used a blind test to estimate the confidence of visual matches with $P \geq 3$. To do this we took a sample of 135 FIRST radio groups and conducted a visual analysis where a random half of the sources was matched with the real sky at the position of the source and the other half matched with a random sky image at a different position. In all, we identified 25 optical counterparts with $P = 3$, and 35 optical with $P = 4$. For the visual matches with $P = 3$, 5/25 identifications came from the random sky image rather than the real one. For visual matches with $P = 4$, only 2/35 identifications were from the random image. From this, we estimate rough confidence levels of about 80 per cent (for $P = 3$) and 94 per cent (for $P = 4$) for our visual identifications of the most complex FIRST sources.

3.3 Automated matching of groups of first sources

3.3.1 Groups of three FIRST sources

Groups of three FIRST components associated with a single host galaxy are likely to contain a core and two lobe components. We chose to define the core (middle) component as the one with the smallest angular distance from other two FIRST members. The remaining two members were assumed to be lobe components (F1 and F2).

To accept the group as a single source, we required the total integrated FIRST flux densities (S_{int}) of the lobe components F1 and F2 to be within a factor of 3 of each other, i.e.

$$1/3 \leq S_{\rm int}({\rm F1})/S_{\rm int}({\rm F2}) \leq 3 \quad (1)$$

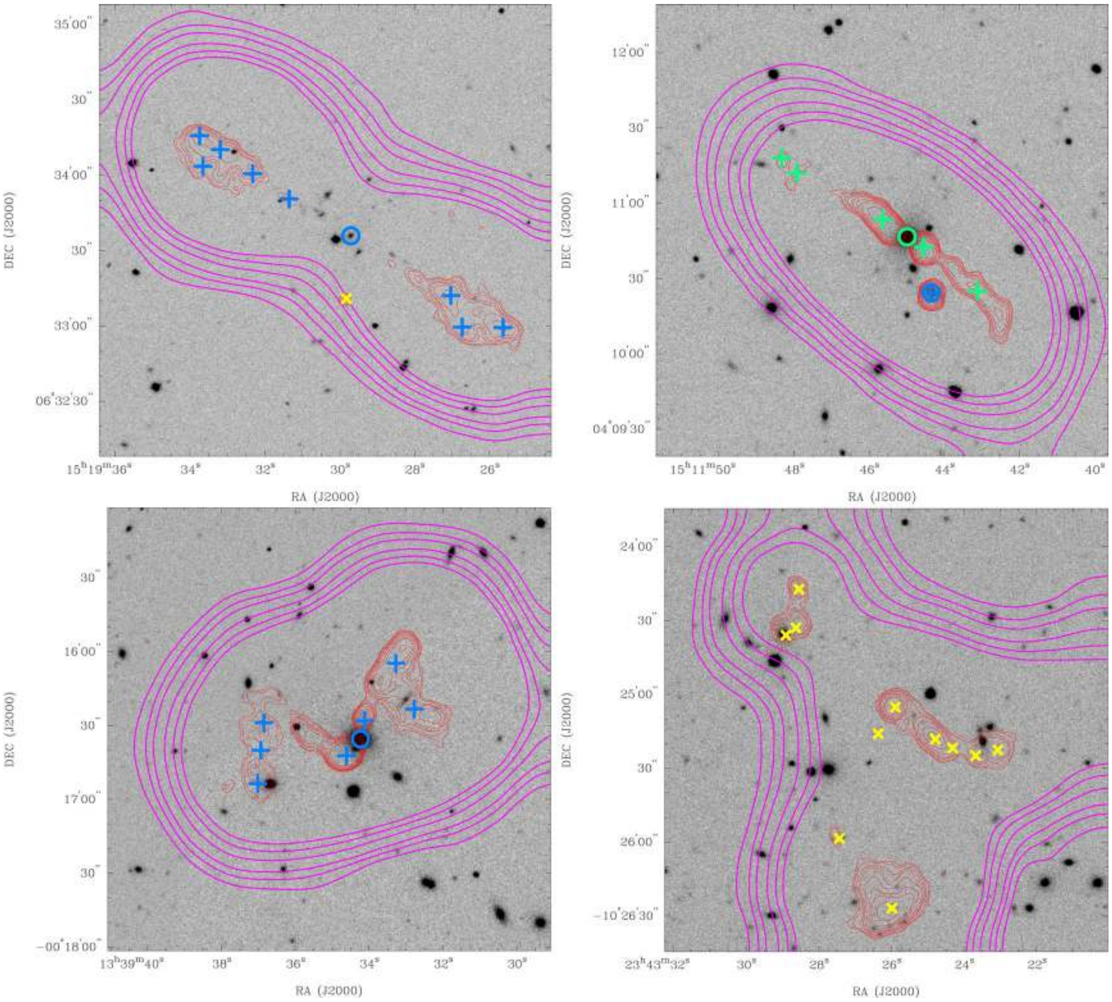


Figure 2. Four examples of complex FIRST source groups that were matched visually. SDSS i-band grey-scale images are overlaid with FIRST (red) and NVSS (magenta) contours. Open circles show the positions of reliable optical matches to the FIRST components (plus symbols), and correspond in colour if they belong to the same source group. FIRST components without an optical counterpart are shown by yellow crosses. Top left: a reliable optical match ($P = 3$, as defined in Section 3.2) to a group of eight FIRST components, with one additional unmatched FIRST source (yellow cross). Top right: two reliable optical matches ($P = 4$ for both), one to a group of five FIRST components and the other to a single FIRST source. Bottom left: an optical match ($P = 3$) to a complex group of seven FIRST sources. Bottom right: a complex group of FIRST sources where there is no optical counterpart with $i_{\rm mod} \leq 20.5$.

This is tighter than the factor of 4 limit used by Magliocchetti et al. (1998), and increases the reliability of the group as a genuine double-lobe plus core radio galaxy. For groups that satisfied this test, we assigned the position of the group as the position of the core component and matched this position to the SDSS catalogue. Groups that did not satisfy the flux-ratio test were reclassified as candidate double sources after removing the lobe component with the largest difference in flux density from the middle component. Matches to an SDSS optical object were automatically accepted at this stage if:

(i) $\theta_{\rm match} < 3$ arcsec (where $\theta_{\rm match}$ is the offset between the radio centroid and the closest SDSS object),

(ii) neither lobe component has an SDSS object within 2.5 arcsec, and

(iii) the shortest component separation (θ_c) is at least one-third of the longest value (θ_a), i.e. $\theta_c \geq 0.33 \times \theta_a$, to ensure that the core component is reasonably close to the radio centroid.

We also visually inspected all triple sources that satisfied the following slightly looser criteria:

(i) $3 < \theta_{\rm match} < 5$ arcsec and neither lobe component has an SDSS object within 2.5 arcsec; or

(ii) $\theta_{\rm match} < 3$ arcsec and neither lobe component has an SDSS object within 2.5 arcsec, but $\theta_c < 0.3 \times \theta_a$; or

(iii) $\theta_{\rm match} < 2.5$ arcsec and $\theta_c \geq 0.3 \times \theta_a$, but one lobe component has an SDSS source within 2.5 arcsec.

This visual matching added 113 triple-source matches to the 266 found by automated matching. In addition, we identified some FIRST triple groups that were genuine double lobe-core systems with an optical counterpart fainter than our survey limit of $i_{\rm mod} = 20.5$. Fig. 3 shows two examples.

3.3.2 *Groups of two FIRST sources*

Two FIRST sources associated with a single host galaxy are likely to be either a pair of lobes or a core and hotspot. We accepted pairs of FIRST sources (F1 and F2) as a genuine association if the integrated flux density ratio of the two components was within a factor of 3 (i.e. satisfied equation 1 above) and the pair

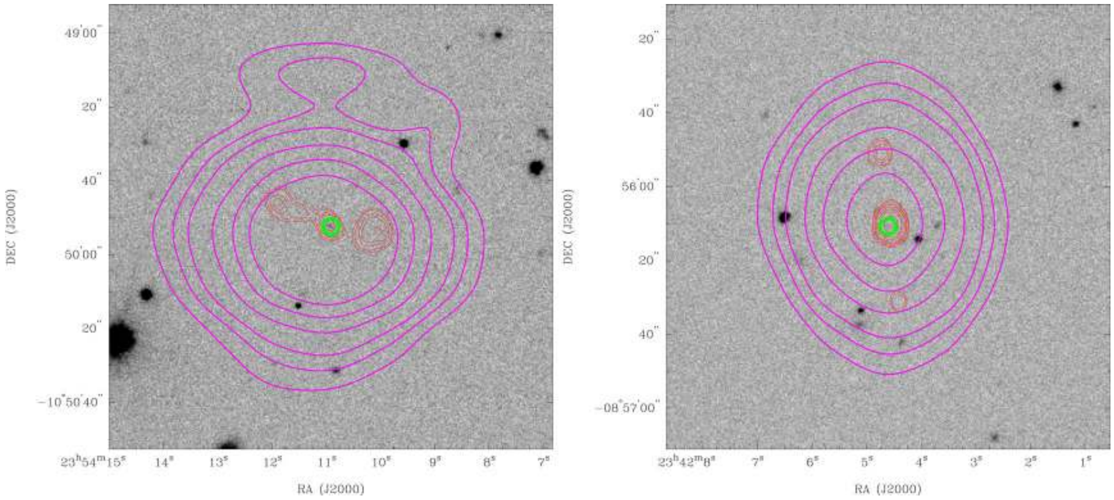


Figure 3. Two examples of FIRST groups (red) with three radio components where the optical counterpart is identified (green) but is fainter than our optical limit of $i_{\rm mod} \leq 20.5$. Objects like this are not included in the final LARGESS catalogue.

also satisfied an additional test set out by Magliocchetti et al. (1998), i.e.

$$\theta_{\rm pair} \leq 100 \times \sqrt{S_{\rm tot}/100}, \qquad (2)$$

where $\theta_{\rm pair}$ is the separation between the two FIRST sources in arcsec, and $S_{\rm tot}$ is the sum of the integrated flux densities of the two components $S_{\rm int}$(F1) and $S_{\rm int}$(F2) in mJy. Adding this constraint allows us to combine bright subcomponents even at relatively large separation, while keeping faint sources as single objects.

Matches to an SDSS optical object were automatically accepted at this stage if either:

(i) $\theta_{\rm match} < 3$ arcsec, neither FIRST component has an SDSS object within 2.5 arcsec and $\theta_{\rm match} \leq \theta_{\rm pair}/2$ (where $\theta_{\rm match}$ is the angular separation between the radio centroid and the closest optical object and $\theta_{\rm pair}$ is the separation of the two radio components), or

(ii) the matched SDSS object is within 2.5 arcsec of one FIRST component, the other FIRST component has no optical counterpart within 2.5 arcsec, and $\theta_{\rm match} \leq \theta_{\rm pair}/2$;

The first of these criteria picks out double-lobe radio galaxies, while the second identifies core-lobe systems.

We visually inspected pairs of sources where:

(i) 3 arcsec $< \theta_{\rm match} \leq 5$ arcsec, neither FIRST component has an SDSS object within 2.5 arcsec and $\theta_{\rm match} \leq \theta_{\rm pair}/2$, or

(ii) the matched optical source is within 2.5 arcsec of one FIRST components, other FIRST component has no optical counterpart within 2.5 arcsec, and $\theta_{\rm match} > \theta_{\rm pair}/2$.

This visual matching added 224 double-source matches to the 981 found by automated matching.

3.4 Automated matching of single first sources

Finally, we carried out automated matching of the large number of FIRST sources not already identified as part of a multi-component system. For these sources, we accepted the closest SDSS optical match within 2.5 arcsec. This 2.5 arcsec cutoff is more restrictive than the 3.0 arcsec value adopted by Sadler et al. (2007) and Helfand, White & Becker (2015), but was chosen on the basis of Monte Carlo tests to optimize the completeness and reliability of the final catalogue, taking into account the high surface density of optical objects down to our magnitude limit of $i_{\rm mod} \leq 20.5$.

Table 2. Number of matched optical counterparts in the final catalogue, and in the GAMA sub-region used for completeness and reliability estimates.

Number of FIRST components	Optical counterparts All	GAMA fields
One	17 163	2803
Two	1294	237
Three	454	88
Four or more	269	40

3.5 Summary of the cross-matching process

Table 2 summarizes the results of the cross-matching process for the full survey area (as well as the sub-area covered by the three GAMA fields listed in Table 1). There are 19 179 optical identifications in the final catalogue of FIRST-SDSS matches across the full survey area, with a total of 22 438 FIRST components. These 19 179 radio-source IDs all have optical photometry and morphological parameters from the SDSS in five (*ugriz*) photometric bands, and comprise the main LARGESS sample with $i_{\rm mod} \leq 20.5$ mag. The great majority of LARGESS objects (89.5 per cent) are single-component FIRST sources, with multi-component sources making up 10.5 per cent of the sample. This is similar to the fraction of FIRST-SDSS matches with complex morphology found by Ivezić et al. (2002).

For the three GAMA fields that have complete overlap between the SDSS and FIRST surveys (see Table 1), we find SDSS matches with $i_{\rm mod} \leq 20.5$ mag for 3168 radio sources made up of 3727 FIRST components.

Overall, 28.6 per cent of FIRST sources were matched with an SDSS object brighter than $i_{\rm mod} = 20.5$, and this appears consistent with the matching rate of 32.9 per cent quoted by Helfand et al. (2015) for the full SDSS photometric catalogue (which has a slightly fainter optical limit).

3.6 Completeness and reliability of the matched catalogue

We used Monte Carlo tests in the three GAMA fields (which are fully covered by all three imaging surveys: SDSS, FIRST and NVSS) to estimate the completeness and reliability of our matching technique.

We generated five pseudo-random optical catalogues by offsetting the GAMA catalogue in declination using shifts of $\Delta\delta$=[0.3, 0.5, 1.0, 1.5, 2.5] deg. Objects shifted outside the GAMA regions in this process were wrapped around the other side of each region, so that the total number and coverage of the random catalogues is the same as the test sample, and the random catalogue retains most of the projected clustering of the original catalogue. The random catalogues were matched against the FIRST radio data in the same way as the real optical catalogue in the GAMA fields, and the matching results were scaled up by the ratio of GAMA to total sky areas (see Table 1) for comparison with the full sample.

3.6.1 Reliability

The reliability R of the final catalogue, i.e. the probability that a matched radio source is genuinely associated with an SDSS object rather than being a random projection on the sky, is calculated as:

$$R = \left(1 - \left\langle N_{\rm rand}^{\rm match}\right\rangle / N_{\rm true}^{\rm match}\right), \tag{3}$$

where $N_{\rm true}^{\rm match}$ is the number of matches from the true catalogue, and $\langle N_{\rm rand}^{\rm match}\rangle$ is the average number of matches using the random catalogues.

Comparing the final number of radio sources in our catalogue (19 179) with the average number of accepted matches (~1247) from the random sample gives us an overall reliability of 93.5 per cent for the LARGESS sample.

This is lower than the value of ~98 per cent for lower redshift samples (e.g. Best et al. 2005a; Sadler et al. 2007) because we are matching to fainter optical objects than previous studies (and also matching with both galaxies and stellar objects) and the higher surface density of optical objects means that the probability of a chance association is increased. At these faint magnitudes, a higher level of reliability could only be achieved by sacrificing completeness, and our final matching strategy was chosen to give a reasonable compromise between completeness and reliability.

3.6.2 Completeness

We also used the GAMA fields to estimate the completeness of the full LARGESS sample, i.e. the fraction of all genuine associations that are identified by our matching process.

Fig. 4 shows the number of FIRST-SDSS matches for single-component sources out to 30 arcsec separation, compared to the normalized random rate from Monte Carlo catalogues. The ratio of the areas under the two curves gives a 5 per cent chance that a match with separation <2.5 arcsec is coincidental, and there is also a small excess of genuine matches out to separations of 4.8 arcsec.

The main source of incompleteness in our sample is the loss of genuine matches with radio-optical separations larger than our 2.5 arcsec matching radius. Our Monte Carlo tests show that ~140 genuine GAMA-FIRST associations will be missed by our 2.5 arcsec cutoff, implying a completeness of ~95 per cent for LARGESS sources with a single FIRST component.

For objects with multiple FIRST components, Monte Carlo tests give a slightly higher completeness level (97 per cent) for the optical matching due to the larger cut-off radius used for matches (3 arcsec, compared to 2.5 arcsec for the single sources). Set against this, the linking process used to associate multiple FIRST components may miss some genuine associations – though we expect this incompleteness to be small because of the high level of visual inspection used in checking the results. We therefore estimate that the completeness of the final sample (~95 per cent) is similar for single and multi-component FIRST sources, which is comparable to the completeness of previous radio samples (e.g. Best et al. 2005a).

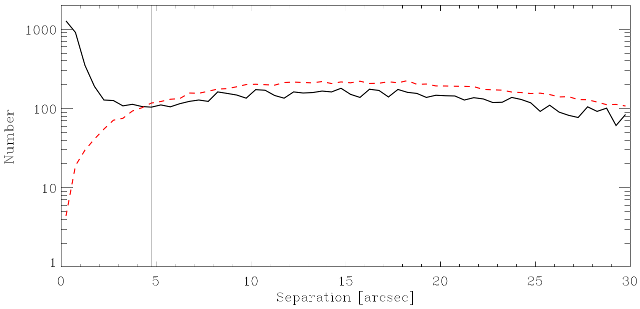


Figure 4. FIRST-SDSS matches (black; solid line) for single-component sources, compared to the average number of random matches (red; dashed line) in the GAMA fields as a function of separation out to 30 arcsec. The vertical line shows the separation at which the average number of FIRST-SDSS matches approaches the number of matches expected by chance.

4 NVSS MATCHING

We now have our final photometric catalogue of 19 179 FIRST radio sources identified with SDSS optical objects brighter than $i_{\rm mod}$ = 20.5 mag. Since the FIRST measurements may underestimate the total flux density of extended radio sources (see Section 2.2), our next step was to cross-match with the NVSS catalogue to get a more reliable measurement of the integrated radio flux density for each object.

4.1 NVSS counterparts to first sources

We used the FIRST components associated with each source in the LARGESS catalogue to cross-match between the FIRST and the NVSS catalogues, using a similar methodology to earlier studies (e.g. Best et al. 2005a; Sadler et al. 2007; Kimball & Ivezić 2008). The cross-matching process is generally straightforward for objects associated with a single FIRST source. For more complex sources, the matching was done as follows.:

For NVSS sources with two or more FIRST matches within 45 arcsec, we need to ensure that the NVSS flux density assigned to a FIRST-optical match is not artificially boosted by contributions from unrelated sources within the larger NVSS beam. To do this, we summed the integrated flux density from all FIRST components within 45 arcsec of an NVSS source. For each FIRST component, we then used its fractional contribution to the total FIRST flux density to assign a scaled proportion of the NVSS flux density to that component. For NVSS-FIRST matches where the FIRST source is associated with an optical counterpart, the NVSS flux density assigned to that FIRST source is now considered to be associated with the corresponding optical counterpart.

4.2 NVSS sources without a first match

Around 8000 NVSS sources in our survey regions did not have a FIRST source within 45 arcsec. We visually inspected the 1299 NVSS components that lay within 3 arcmin of one of our FIRST sources. As noted by Best et al. (2005a), this 3 arcmin radius is large enough to pick up any extended NVSS components, but smaller than the typical separation of unrelated NVSS sources (8–10 arcmin). We found a further 159 NVSS components associated with 121 LARGESS objects (30 with 2 NVSS components and 4 with 3 NVSS components). In addition, 259 NVSS components were found to be associated with 252 (7 with 2 NVSS components) SDSS ($i_{\rm mod} \leq 20.5$) objects that were not previously in the LARGESS catalogue.

From this, we estimate that ~20 per cent (252/1299) of the ~8000 NVSS sources without a FIRST detection in our survey area will be associated with an SDSS ($i_{\rm mod} \leq 20.5$) counterpart that is not already part of our final LARGESS sample. In other words, our survey area includes ~1600 faint radio sources that are too diffuse to be detected by the FIRST survey. A comparison with the Best & Heckman (2012) sample showed that excluding objects with an NVSS detection but no FIRST detection mainly excludes star-forming galaxies at low redshift, so does not significantly affect the completeness of our catalogue for radio AGN (HERGs and LERGs).

4.3 Comparison of FIRST and NVSS flux densities

Figs 5 and 6 compare the FIRST and NVSS flux densities for matched objects in the LARGESS sample. As can be seen from the bottom panel of Fig. 5, 95 per cent of the LARGESS sample with $S_{\rm tot}^{\rm FIRST} \geq 3.5$ mJy have an NVSS match. Any analysis requiring accurate flux densities for extended radio sources should therefore impose a $S_{\rm tot}^{\rm FIRST} \geq 3.5$ mJy limit for the LARGESS sample.

The left-hand panel of Fig. 6 shows the mean difference between the FIRST and NVSS flux densities ($\langle \Delta S \rangle = \langle S_{\rm tot}^{\rm FIRST} - S_{\rm tot}^{\rm NVSS} \rangle$) divided by the mean FIRST flux density in logarithmically spaced bins of FIRST flux density. Since the distribution of ΔS is slightly asymmetric (see Fig. 5), we also apply the same analysis using the median difference instead of the mean difference (right-hand panel of Fig. 6). Large discrepancies are only seen at the lowest flux densities, where the average difference is ~25 per cent (or ~13 per cent for the median difference) of the average FIRST flux density.

From these results, we estimate that between 5 and 25 per cent of the 1.4 GHz flux density of a typical LARGESS source is in a diffuse component detected by NVSS but missed by the FIRST survey. As can be seen from Fig. 6, the discrepancy between the NVSS and FIRST flux density measurement increases at lower flux density levels.

5 SPECTROSCOPIC DATA

In compiling the LARGESS catalogue, we aimed to achieve as high a level of spectroscopic completeness as possible across a large area of sky. To do this, we combined existing data from earlier spectroscopic surveys with new spectra obtained in collaboration with the WiggleZ and GAMA teams (mainly using 'spare' fibres not assigned to the main WiggleZ/GAMA survey targets). The final spectroscopic completeness for the catalogue as a whole is 64 per cent (i.e. 12 329 of the 19 179 objects in the LARGESS catalogue have at least one spectroscopic observation) and the redshift completeness is currently 57 per cent (10 856 objects have a reliable optical redshift). As can be seen from Table 1, the completeness varies with sub-region and is highest (>80 per cent spectroscopic completeness) in the three GAMA regions.

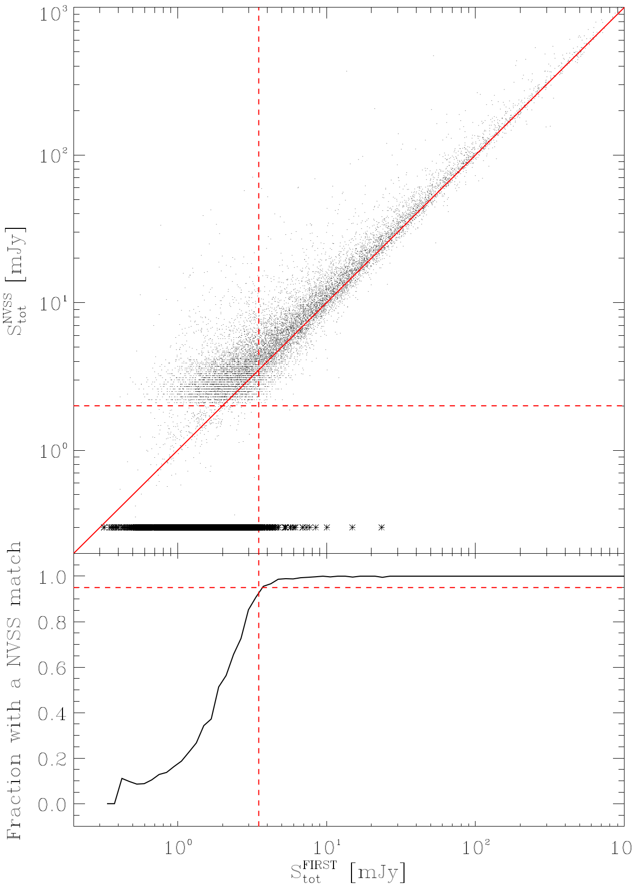


Figure 5. Top panel: comparison of FIRST ($S_{\rm tot}^{\rm FIRST}$) and NVSS ($S_{\rm tot}^{\rm NVSS}$) total flux densities for LARGESS objects. The horizontal dashed line is at $S_{\rm tot}^{\rm NVSS} = 2$ mJy – some points lie below this line because their total NVSS flux density was adjusted to take into account multiple FIRST matches as discussed in the text. FIRST sources without an NVSS match are plotted at $S_{\rm tot}^{\rm NVSS} = 0.3$ mJy, and the solid line shows a one-to-one relation in flux density. Bottom panel: the fraction of objects with an NVSS match as a function of $S_{\rm tot}^{\rm FIRST}$. The horizontal dashed line is at 0.95, and a vertical dashed line marks the value of $S_{\rm tot}^{\rm FIRST}$ at this 95 per cent matching level.

5.1 Spectra and redshifts from earlier surveys

We incorporated spectra and redshifts from earlier surveys into our catalogue by cross-matching our sample with the 2dF and 6dF QSO Redshift Surveys (2QZ and 6QZ; Croom et al. 2004), the 2dF-SDSS LRG And QSO survey (2SLAQ; Cannon et al. 2006; Croom et al. 2009) and the SDSS DR6 spectroscopic catalogue (Adelman-McCarthy et al. 2008).

We set up a uniform quality classification system for redshifts from these earlier surveys. The interactive redshift code RUNZ (Saunders, Cannon & Sutherland 2004) was designed for use in the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and also used by the 2SLAQ-LRG, GAMA and (in a modified version) WiggleZ spectroscopic surveys (Drinkwater et al. 2010). The user estimates a redshift by cross-correlating an observed spectrum with a set of template spectra and separate emission line fits, then inspects the estimated redshift and has the option to adjust the measurement. They then assign a quality flag Q to indicate the reliability of the final redshift. These surveys all used the same criteria for $Q = 1$ to 4, where higher values of Q indicate a higher confidence level for the redshift measurement. Redshifts with $Q \geq 3$ are considered to be reliable. We adopted a range of 0 to 6 for Q (GAMA only uses

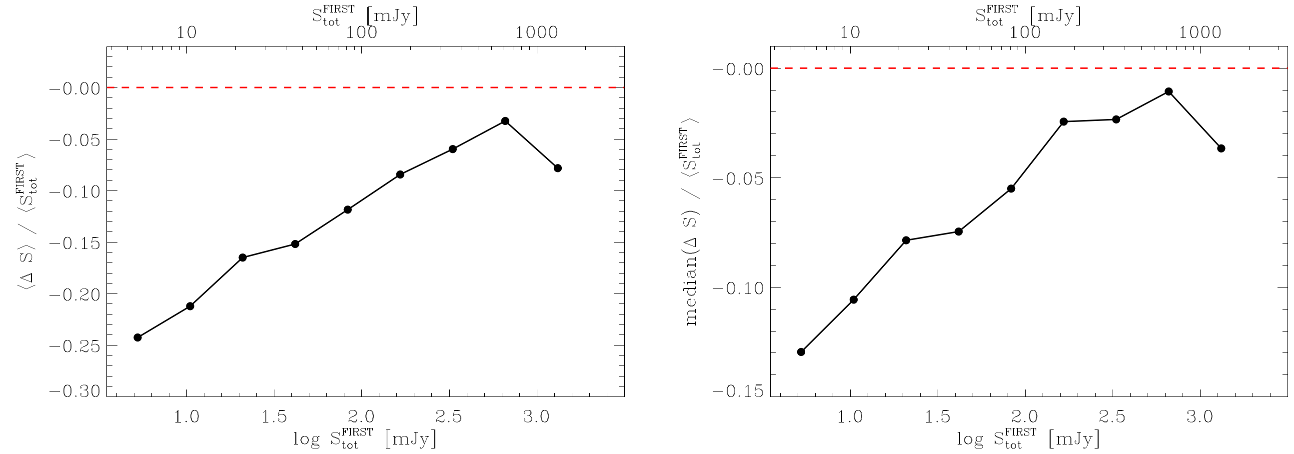


Figure 6. The fractional difference between the total flux densities of individual sources measured by NVSS and FIRST, in bins of FIRST flux density. Both plots are for sources that have $S_{\rm tot}^{\rm FIRST} > 3.5$mJy and an NVSS match. The left-hand panel shows the mean difference between the FIRST and NVSS flux densities ($\langle \Delta S \rangle = \langle S_{\rm tot}^{\rm FIRST} - S_{\rm tot}^{\rm NVSS} \rangle$) divided by the mean FIRST flux density ($\langle S_{\rm tot}^{\rm FIRST} \rangle$). The right-hand panel uses the median difference rather than the mean.

Table 3. Conversion between redshift quality codes `zconf` for SDSS and `zflag` for 2QZ/2SLAQ-QSO) and the initial quality code ($Q_{\rm initial}$) used in the LARGESS catalogue. The last three columns show the results from our re-redshifting of SDSS spectra for objects in the GAMA fields. *N*(re-redshift) is the number of SDSS spectra re-redshifted, and *N*(agree) is the number of spectra where the re-redshift and SDSS redshifts agree within 0.01 (i.e. $|\Delta z| < 0.01$) and the quality is considered reliable ($Q > 2$). The final column is the percentage that agree (*N*(agree)/*N*(re-redshift)) for each $Q_{\rm initial}$ bin.

Survey	SDSS	2SLAQ-QSO and 2QZ		SDSS re-redshift	
$Q_{\rm initial}$	`zconf`	`zflag`	*N*(re-redshift)	*N*(agree)	per cent agree
6	–	–	–	–	–
5	>0.99	–	576	575	99.8
4	>0.95 and ≤0.99	–	172	169	98.2
3	>0.80 and ≤0.95	11	143	132	92.3
2	>0.50 and ≤0.80	12, 21 (1 case) or 22	66	35	53
1	≤0.50	>22	37	8	22
0	–	–	–	–	–

$Q = 1$ to 4), where the additional $Q = 0$ identifies a poor-quality (or missing) spectrum; $Q = 5$ indicates an extremely reliable redshift from a good-quality spectrum and $Q = 6$ is reserved for spectra classified as Galactic stars.

We converted the redshift quality codes from the 2SLAQ-QSO, 2QZ, 6QZ and SDSS surveys to new values ($Q_{\rm initial}$) as outlined in Table 3. As a check, we also re-redshifted a subset of spectra from these surveys as described below. For most objects, $Q_{\rm initial}$ was unchanged after re-redshifting. Of the 6325 LARGESS sources with spectroscopic observations from one or more of these earlier surveys, 5798 were initially classified as having reliable redshifts (i.e. $Q_{\rm initial} \geq 3$).

5.2 New spectroscopic observations

Our goal was to obtain new optical spectra for all LARGESS objects with $i_{\rm mod} < 20.5$ which did not already have a reliable redshift from the spectroscopic catalogues mentioned above. To do this, we carried out piggyback observations in conjunction with two large spectroscopic observing programmes, the GAMA (Driver et al. 2011; Hopkins et al. 2013) and WiggleZ (Drinkwater et al. 2010) surveys, both of which used the 3.9 m Anglo-Australian Telescope (AAT) at the Siding Spring Observatory (SSO). All our new spectra were taken using the fibre-fed AAOmega spectrograph with the two-degree field fibre positioner (2dF). Objects that were not already part of the main GAMA or WiggleZ target sample were assigned as lower priority filler targets in the survey fields (see Driver et al. (2011) and Drinkwater et al. (2010) for priority listing).

Piggybacking on these large spectroscopic surveys allows us to obtain spectra efficiently for a large sample of radio galaxies whose surface density is too low to make effective use of the 400 fibres available in a 2dF field. We can also use the parent samples of GAMA and WiggleZ galaxies to measure environments and to build well-defined non-radio control samples to compare with.

The spectra taken by the GAMA survey team covered the wavelength range 3720–8850Å, with a typical integration time of 3000–5000 s. There were two phases to the GAMA survey, both of which are now complete. The first phase (GAMA-I) formed the basis for our spectroscopic target selection. A second phase (GAMA-II) extended the first by adding two southern fields, expanding the equatorial regions and also including objects with $r_{\rm pet} < 19.8$ mag in all regions. The GAMA-II spectroscopic targets come from the SDSS seventh data release (DR7) photometric catalogue. We did not add any additional objects to the LARGESS sample to reflect the boundary changes in GAMA-II, but our original GAMA targets remained in the GAMA-II spectroscopic target list and we have used both GAMA-I and -II spectra in our final data catalogue.

The spectra taken by the WiggleZ survey team covered the wavelength range 4700–9500Å, with an integration time of 3600 s (Drinkwater et al. 2010). The WiggleZ main survey targets were restricted to objects with r-band magnitudes in the range $20 < r < 22.5$ (the bright cutoff was applied to avoid observing low-redshift galaxies, since the WiggleZ target redshift range was $0.4 < z < 1.0$). The WiggleZ survey team observed 3674 radio targets, of which only 203 (∼6 per cent) were main WiggleZ targets.

5.3 Re-redshifting

The GAMA and WiggleZ survey teams both measured redshifts on-the-fly at the telescope after each observation. This first-pass redshift was used to select targets for observing on following nights, and in most cases (especially for WiggleZ) became the final redshift of the target. Although the first-pass redshifts and quality codes were usually reliable, they had some inhomogeneities caused by different observers (with various expertise/experience) assigning redshifts, and in the case of WiggleZ, RUNZ was optimized to measure redshifts from emission lines. To control and homogenize the data quality, we re-redshifted five sets of spectra:

(i) those observed by WiggleZ;
(ii) those observed by GAMA;
(iii) SDSS spectra in the GAMA regions;
(iv) all SDSS spectra with $Q_{\rm initial} \le 2$;
(v) all other spectra with $Q_{\rm initial} \le 2$ and SNR > 8.5.

We re-redshifted sets (i) and (ii) to homogenize the redshifts and qualities between the GAMA and WiggleZ survey. This is a different and independent re-redshifting from the one carried out by the GAMA survey team (Driver et al. 2011; Liske et al. 2015). Set (iii) was re-redshifted to compare the automatically assigned redshifts and quality codes from SDSS to those assigned by manual inspection of the spectra. Re-redshifting of the last two sets was done to identify and correct redshifts of objects that had low redshift confidence in the GAMA or WiggleZ survey, often because their optical spectra showed unusual features.

For WiggleZ observations prior to 2010, four of the authors (JHYC, SMC, EMS and HMJ) re-redshifted all the radio target spectra by eye. JHYC re-redshifted all the 2011 WiggleZ observations, as well as sets (ii) to (v) above.

5.4 Final redshifts

12 329 objects in the LARGESS sample have at least one spectroscopic observation. For each of these objects, we defined a single *best* redshift and quality (*QOP*) by comparing all available redshifts for that source as described below. Table 4 gives a breakdown of the final number of redshift measurements in each *QOP* quality bin.

The LARGESS sample is based on the SDSS DR6, which was the most recent SDSS release at the start of the project. Since then, the SDSS-II survey has been completed with the release of SDSS DR7. For objects without an existing redshift measurement, we adopted the SDSS DR7 redshift where available. The quality codes for the SDSS DR7 spectra were converted as shown in Table 3. This added an additional 1130 spectra to the final catalogue.

Almost half the sample (7595 objects) have a single spectroscopic observation and only a single redshift measurement. Most of these are located in parts of the 2SLAQ strips that do not overlap the WiggleZ or GAMA area, or are objects with a $Q \ge 3$ redshift from SDSS or an earlier survey. We automatically accepted this redshift and quality code and included it in the final spectroscopic catalogue.

1,442 objects have a single spectrum from the WiggleZ survey, but multiple redshifts for that spectrum from the re-redshifting process. In this instance, we took the redshift value with the most agreements. If there were no agreements, one of the authors (JHYC) selected a final redshift and quality.

Table 4. Number of sources in each redshift reliability bin QOP, shown separately for the GAMA/WiggleZ areas and the remaining lower completeness regions (see Table 1). Higher QOP values indicates a more reliable redshift, and $QOP \ge 3$ is taken to be reliable enough for scientific analysis. QOP = 6 is reserved for Galactic stars.

QOP	GAMA/WiggleZ regions	Other regions	Total
0	11	0	11
1	631	78	709
2	703	50	753
3	1669	428	2097
4	4422	537	4959
5	2679	1030	3709
6	88	3	91
Total	10 203	2126	12 329
Not observed	4447	2403	6850
Total including unobserved objects	14 650	4529	19179

For the 997 sources with two or more spectroscopic observations and a single redshift measurement for each observation, we started by identifying reliable redshifts that agreed with each other ($|\Delta z| < 0.01$ and $Q \ge 3$) and accepted the value with the most agreements. For redshifts with the same number of agreements, we took the set of agreements with the highest Q assigned, and within this set we selected the redshift and quality assigned to the spectrum with the highest signal-to-noise ratio (SNR) as the best redshift. If there were no agreements, we accepted the redshift with the highest Q.

There are 1165 sources with both multiple spectra and multiple redshift measurements from re-redshifting. For objects in this category, we again identified reliable redshifts that were in agreement ($|\Delta z| < 0.01$ and both have $Q \ge 3$) and defined the *best* redshift as the one with the most agreements. If there were multiple sets of redshifts with the same number of agreements, then we accepted the one with the highest Q, and if Q was the same, we adopted the redshift assigned using the highest SNR spectrum.

5.5 Redshift reliability

To assess the reliability of our redshift measurements, we used a similar technique to previous studies (e.g. Colless et al. 2001; Croom et al. 2004; Drinkwater et al. 2010) and compared our final QOP = 3–5 redshifts with a repeat observation of the same object with $Q_{\rm rep} \ge 3$ (where QOP is the quality associated with our final redshift and $Q_{\rm rep}$ is the redshift quality associated with the repeated observation).

We found that redshift measurements with $Q \ge 3$ are generally highly reliable, with implied single-measurement blunder (i.e. $|\Delta z| \ge 0.01$) rates of 8.5 per cent and 1 per cent for QOP = 3 and 4, respectively. For pairs where the two redshifts agree, the pairwise rms dispersion of redshift differences is also small (with a typical value of $|\Delta z| \le 0.0025$). The QOP = 5 pairs have a high fraction of broad emission-line QSOs, where the broad peaks sometimes make it difficult to determine a consistent redshift. As a result, the QOP = 5 redshift measurements have a slightly higher dispersion on average than the QOP = 4 measurements.

Table 5. Emission line wavelength definitions from MPA-JHU for the [O III] λ5007, [O III] λ4959 and Hβ emission lines.

Line name	Line centre (Å)	Lower bound (Å)	Upper bound (Å)
Hβ	4861.325	4851.0	4871.0
[O III] λ4959	4958.911	4949.0	4969.0
[O III] λ5007	5006.843	4997.0	5017.0

5.6 Emission-line measurements

Our catalogue includes emission-line flux measurements for galaxies with good-quality spectra observed by the WiggleZ, GAMA and/or SDSS teams. Spectra from earlier surveys such as 2SLAQ and 2dFGRS generally lack the accurate flux calibration, spectral resolution or wavelength coverage needed for reliable emission-line measurements.

5.6.1 GAMA and SDSS emission-line measurements

The GAMA survey provides emission-line measurements (internal data: GandalfSpecAnalysis v08.3; Steele et al., in preparation) using the GANDALF (Gas AND Absorption Line Fitting; Sarzi et al. 2006) code. For galaxies with SDSS spectra, we used the Max-Planck-Institut für Astrophysik – Johns Hopkins University (MPA-JHU) emission-line measurements.[1] In both cases we ensured that the redshifts used for emission-line measurements agreed with those in our final catalogue. The GANDALF and the MPA-JHU emission line measurements both apply a template fit to account for stellar absorption before measuring emission line fluxes. Thus we expect both measurements to be comparable and robust (for a more detailed comparison, see Hopkins et al. 2013).

5.6.2 WiggleZ emission lines

The WiggleZ spectra have poorer spectrophotometry than the GAMA/SDSS spectra, since the method adopted for spectroscopic curvature correction in WiggleZ spectra makes it difficult to subtract the stellar continuum accurately (the WiggleZ survey was designed to measure faint objects with strong optical emission lines and little or no visible continuum). We therefore chose to make only a single measurement of the [O III] λ5007 emission-line flux and equivalent width for the WiggleZ spectra. The [O III] λ5007 line is not significantly affected by stellar absorption features, and measuring this line allows us to distinguish between low-and high-excitation radio galaxies.

For these measurements, we used the same wavelength definitions as MPA-JHU (see Table 5) and estimated the continuum at the position of [O III] λ5007 by fitting a second-order polynomial to the local continuum. The emission-line flux was then measured by integrating over the continuum-subtracted [O III] λ5007 region. We used the covariance matrix provided by the SVDFIT routine together with the individual pixel variances to calculate the [O III] λ5007 flux error and the continuum error.

The [O III] λ5007 equivalent width (EW([O III] λ5007)) is measured by dividing the integrated line flux by the continuum flux at the line centre. The estimated continuum flux at the line centre can sometimes equal to or fall below zero due to systematic sky subtraction errors for the faintest objects, but there may still be a prominent [O III] λ5007 emission line. To avoid a non-physical value for the EW([O III] λ5007), we instead derived a minimum equivalent width by dividing the minimum value of the [O III] λ5007 flux (i.e. [O III] λ5007 flux minus the error) by the error in the continuum at the line centre.

6 SURVEY COMPLETENESS

We now quantify the spectroscopic completeness of the catalogue as a function of apparent magnitude and colour. Here, the *targeting completeness* refers to the fraction of LARGESS objects that have been spectroscopically observed, and the *redshift completeness* refers to the fraction of objects with a reliable redshift. In this section we focus on the GAMA and WiggleZ regions listed in Table 1, which have the highest completeness and so are more likely to be used for follow-up studies. These regions (which are also covered by the SDSS survey) contain a total of 14 650 catalogue sources, of which 10 203 have spectroscopic observations (see Table 4).

6.1 Targeting completeness

Fig. 7 shows the fraction of LARGESS sources for which a spectroscopic observation has been made, as a function of $r_{\rm pet}$, $i_{\rm mod}$, and ($g_{\rm mod} - i_{\rm mod}$) colour. These are separated by survey region as follows.

(i) The top row in Fig. 7 shows the full set of 14 650 objects included in either the GAMA or the WiggleZ fields (Region `G or W`). The overall targeting completeness of 70 per cent for this region closely resembles the targeting completeness for objects in the WiggleZ fields, since the WiggleZ survey area is roughly four times larger than the GAMA area and so contains many more targets. The dip in completeness seen in the left-hand panels at $18 < r_{\rm pet} < 20$ arises from the bright limit of the WiggleZ sample, as discussed in point (iii) below.

(ii) The second row shows objects in the 67 deg^2 of sky covered by the GAMA survey area but not the WiggleZ area (region `G only`), for which the overall targeting completeness is 88 per cent. These spectra come mainly from SDSS at the bright end, with GAMA observations becoming increasingly important at the faint end (the GAMA survey observed secondary targets out to $r_{\rm pet}$ = 22.5, with uniform sampling for objects with $18 < r_{\rm pet} < 21.5$ mag). The targeting rate here is fairly uniform in both $i_{\rm mod}$ magnitude and colour, but shows a gradual fall-off in r-band completeness beyond the GAMA-II main sample limit of $r_{\rm pet} < 19.8$.

(iii) The next row `W only` is for objects in the ∼540 deg^2 of sky covered by the WiggleZ survey, but falling outside the GAMA survey regions. Here, the spectroscopic observations are dominated by SDSS at the bright end ($r_{\rm pet} < 17.77$) and WiggleZ observations at the faint end. The WiggleZ survey team applied an additional bright limit of $i_{\rm mod} > 18.0$ for our piggyback targets, to reduce light contamination from brighter sources due to cross-talk between fibres on the spectrograph CCD (since the WiggleZ main survey targets were faint galaxies with $20 < r < 22.5$ mag). The combination of this bright limit and the SDSS faint limit causes a dip in the targeting completeness which can be seen in both $i_{\rm mod}$ and $r_{\rm pet}$ plots. The dip does not go to zero, because the SDSS also observed some secondary targets beyond the main survey limit (though the targeting completeness for the SDSS secondary targets drops off quickly for $r_{\rm pet} > 17.77$). We also note that bluer objects appear to be preferentially targeted in the `W only` plot of Fig. 7. This is mainly because the targeting completeness is higher for SDSS than

---

[1] http://www.mpa-garching.mpg.de/SDSS

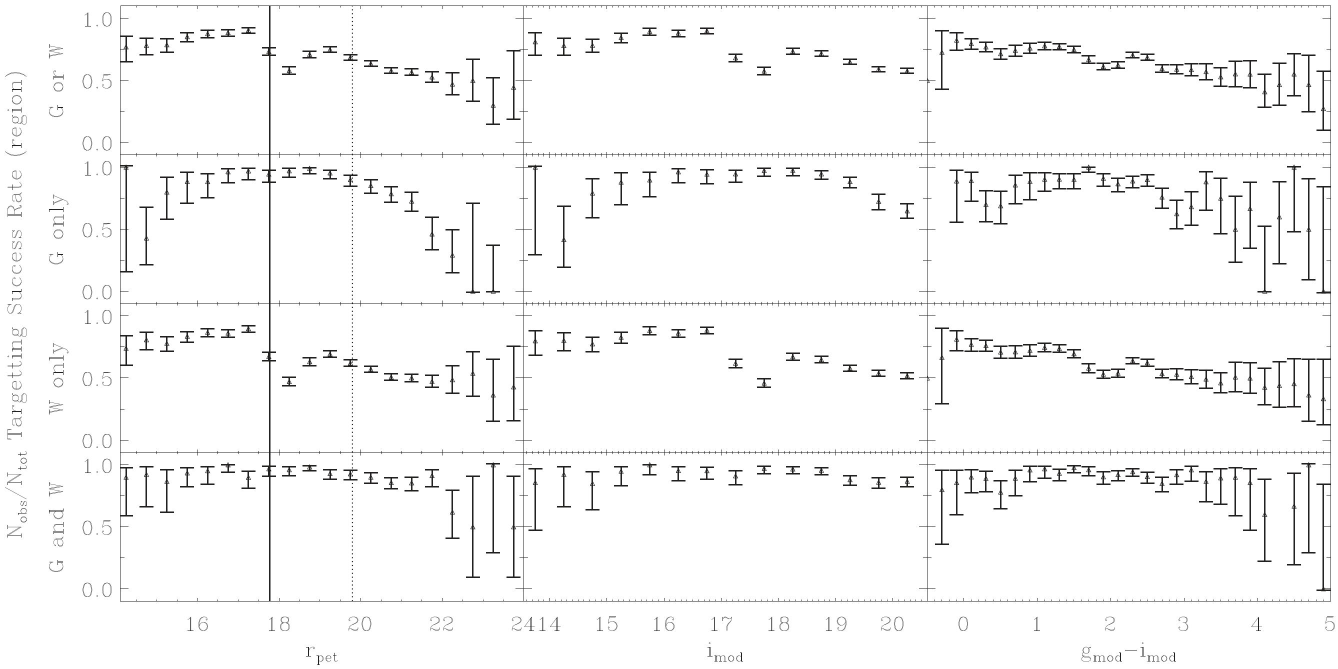


Figure 7. The spectroscopic targeting success rate ($N_{\rm obs}/N_{\rm tot}$) for LARGESS objects in the WiggleZ and GAMA regions (see Table 1) as a function of $r_{\rm pet}$, $i_{\rm mod}$ and ($g_{\rm mod} - i_{\rm mod}$). $N_{\rm tot}$ is the number of sources within a specific region, while $N_{\rm obs}$ is the number with a spectroscopic observation. We only count an object as observed if the best redshift comes from the SDSS, WiggleZ or GAMA survey. The four sets of plots are for different sky regions as described in Section 6.1 of the text. The vertical solid and dashed lines in the left-hand panel show the $r_{\rm pet}$ magnitude limits for the SDSS and GAMA-II surveys, respectively. The WiggleZ survey observed objects in the magnitude range $20 < r < 22.5$.

WiggleZ, and the SDSS observations target brighter objects that are at lower redshifts ($z < 0.3$) and so have bluer colours than the fainter objects targeted by the WiggleZ spare-fibre programme. This effect can be corrected in any follow-up analysis by taking into account the apparent magnitude completeness and limits.

(iv) The final row is for objects in the 77 deg^2 region of sky covered by both GAMA-I and WiggleZ surveys (region `G and W:` in Fig. 7; see also Table 1 and Fig. 1). Region `G and W` has the most uniform targeting completeness in all three optical parameters ($r_{\rm pet}$, $i_{\rm mod}$ and $g_{\rm mod} - i_{\rm mod}$). The combination of observations from the SDSS, GAMA and WiggleZ surveys ensures that many objects are observed, and washes out the individual magnitude and colour limits from these surveys.

6.2 Redshift completeness

Not all spectroscopic observations result in a reliable redshift measurement. The overall redshift success rate for spectroscopic observations of the LARGESS sample is 88 per cent, but this varies with target properties such as brightness, colour and the presence or absence of emission lines.

Fig. 8 shows the redshift success rate as a function of $r_{\rm pet}$, $i_{\rm mod}$, and $g_{\rm mod} - i_{\rm mod}$ colour. These plots show a drop in the redshift success rate towards fainter magnitudes, as well as a decreasing success rate for redder objects with $g_{\rm mod} - i_{\rm mod} \gtrsim 2$.

Fig. 9 shows the $g_{\rm mod} - i_{\rm mod}$ colour as a function of $i_{\rm mod}$ apparent magnitude for the full LARGESS sample. Objects with $g_{\rm mod} - i_{\rm mod} > 2$ are on average fainter (mean $i_{\rm mod} = 19.1$ mag) than objects with $g_{\rm mod} - i_{\rm mod} < 2$ (mean $i_{\rm mod} = 17.7$ mag). Therefore the lower redshift success rate for objects with $g_{\rm mod} - i_{\rm mod} > 2$ may be explained, in part, by the fact they are fainter objects. In general bluer objects are also more likely to have emission lines, which increase the chance of measuring a reliable redshift.

We can model the overall redshift completeness as a function of magnitude using the sigmoid function (e.g. Ellis & Bland-Hawthorn 2007; Loveday et al. 2012):

$$y(x) = 1/[1 + e^{a(x-b)}], \qquad (4)$$

where x is the photometric magnitude, a is the stiffness of the function and b is the magnitude at a redshift success rate of 50 per cent (i.e. $y(b) = 0.5$). The best fits are shown as red lines in Fig. 8, and the parameters are listed in Table 6.

7 SPECTRAL CLASSIFICATION

In this section, our goal is to use the optical spectra of LARGESS sources to determine the dominant physical process (either an active galactic nucleus (AGN) or star formation) responsible for the radio emission in each individual object within our sample. In other words, we are classifying the *radio source* rather than the optical spectrum itself.

In most cases, the classification of the radio source can be deduced directly from the optical spectrum. However, as discussed by Best & Heckman (2012), it is important to be able to identify objects where the optical spectrum is dominated by strong emission lines from a radio-quiet AGN[2] but the radio emission is powered mainly by star formation processes. Here, the optical star formation signature may be obscured by the AGN lines in the single-fibre spectra we are using. As discussed in Section 7.2, we make a quantitative

---

[2] In this paper, we use the terms ‘radio-quiet’ and ‘radio-loud’ to refer to objects in which the radio continuum emission is predominantly powered by star formation and AGN processes, respectively.

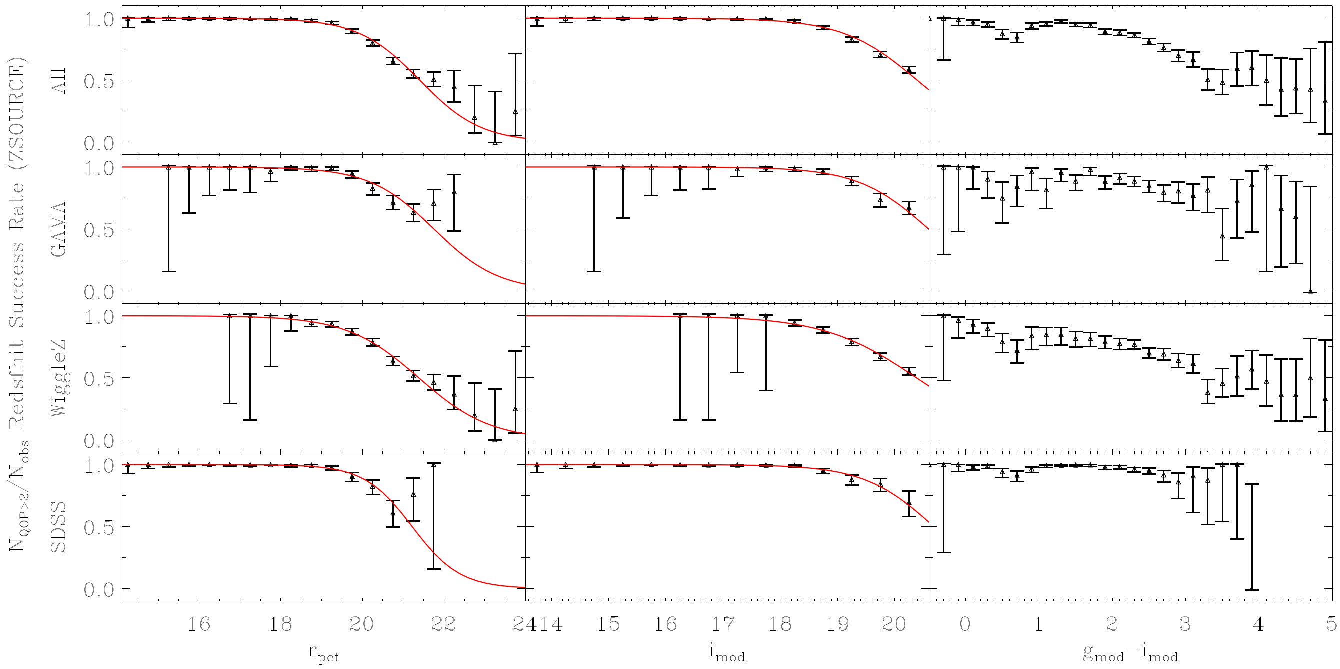


Figure 8. Redshift success rate ($N_{QOP>2}/N_{\rm obs}$, where $N_{QOP>2}$ is the number with a reliable redshift and $N_{\rm obs}$ is the number observed) as a function of $r_{\rm pet}$, $i_{\rm mod}$ and $g_{\rm mod} - i_{\rm mod}$. This is only shown for objects with ZSOURCE = GAMA, WiggleZ or SDSS. The red line is the best fit to a sigmoid function using maximum likelihood estimation, which is only applied to the apparent magnitude comparisons. Errors are estimated using the method described by Cameron (2011) for a 95 per cent confidence interval.

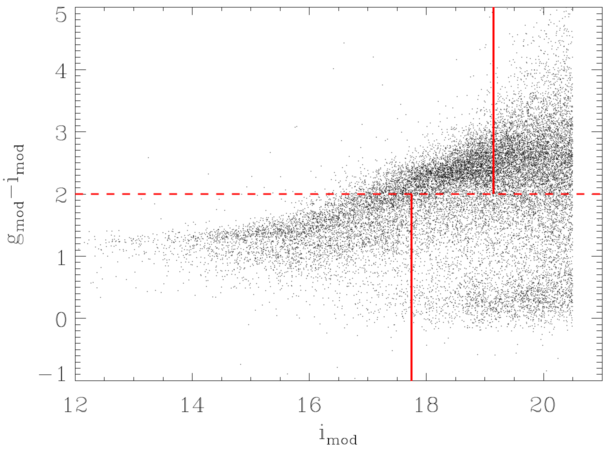


Figure 9. ($g_{\rm mod} - i_{\rm mod}$) colour versus $i_{\rm mod}$ magnitude for the full sample. The horizontal dashed line marks ($g_{\rm mod} - i_{\rm mod}$) = 2, above which the redshift success rate drops for objects with redder colour. Vertical lines show the mean $i_{\rm mod}$ magnitude for objects redder and bluer than ($g_{\rm mod} - i_{\rm mod}$) = 2.0.

Table 6. Sigmoid parameters for redshift completeness from maximum likelihood estimation, as plotted in Fig. 8 and discussed in Section 6.2 of the text.

	$r_{\rm pet}$		$i_{\rm mod}$	
ZSOURCE	*a*	*b*	*a*	*b*
All	$1.33^{+0.04}_{-0.04}$	$21.41^{+0.56}_{-0.55}$	$1.54^{+0.05}_{-0.05}$	$20.39^{+0.45}_{-0.44}$
GAMA	$1.27^{+0.08}_{-0.07}$	$21.75^{+1.43}_{-1.34}$	$1.62^{+0.13}_{-0.12}$	$20.59^{+1.06}_{-1.01}$
WiggleZ	$1.14^{+0.04}_{-0.04}$	$21.38^{+0.80}_{-0.78}$	$1.23^{+0.05}_{-0.05}$	$20.38^{+0.69}_{-0.67}$
SDSS	$1.71^{+0.12}_{-0.11}$	$21.22^{+1.16}_{-1.10}$	$1.64^{+0.10}_{-0.10}$	$20.69^{+1.17}_{-1.11}$

comparison between the observed radio luminosity and the star formation rate estimated from the Hα emission line to identify such objects.

We classified the optical spectra of LARGESS sources in two ways. A first-pass *visual classification* (described in Section 7.1) allows us to identify BL Lac candidate and emission-line objects where the lines have broad wings (class AeB below), as well as providing a useful comparison with earlier work and a series of checks on the automated classification process. We also make an automated *quantitative classification* (see Section 7.2) for objects where the [O III] λ5007 emission line falls within the GAMA/WiggleZ/SDSS spectral range.

7.1 Qualitative (visual) spectral classification

During the re-redshifting process, we assigned a spectral class based on a visual inspection (VISCLASS) for each object with a reliable redshift (i.e. $Q \geq 3$). This visual classification was based on the scheme used by Sadler et al. (2002). Example spectra of the main classes are shown in Fig. 10, and the classification criteria are as follows.

Aa. Stellar continuum with no apparent emission lines.

Aae. Stellar continuum with weak emission lines (e.g. [O III] λ5007, Hα, [N II] λ6583 etc.).

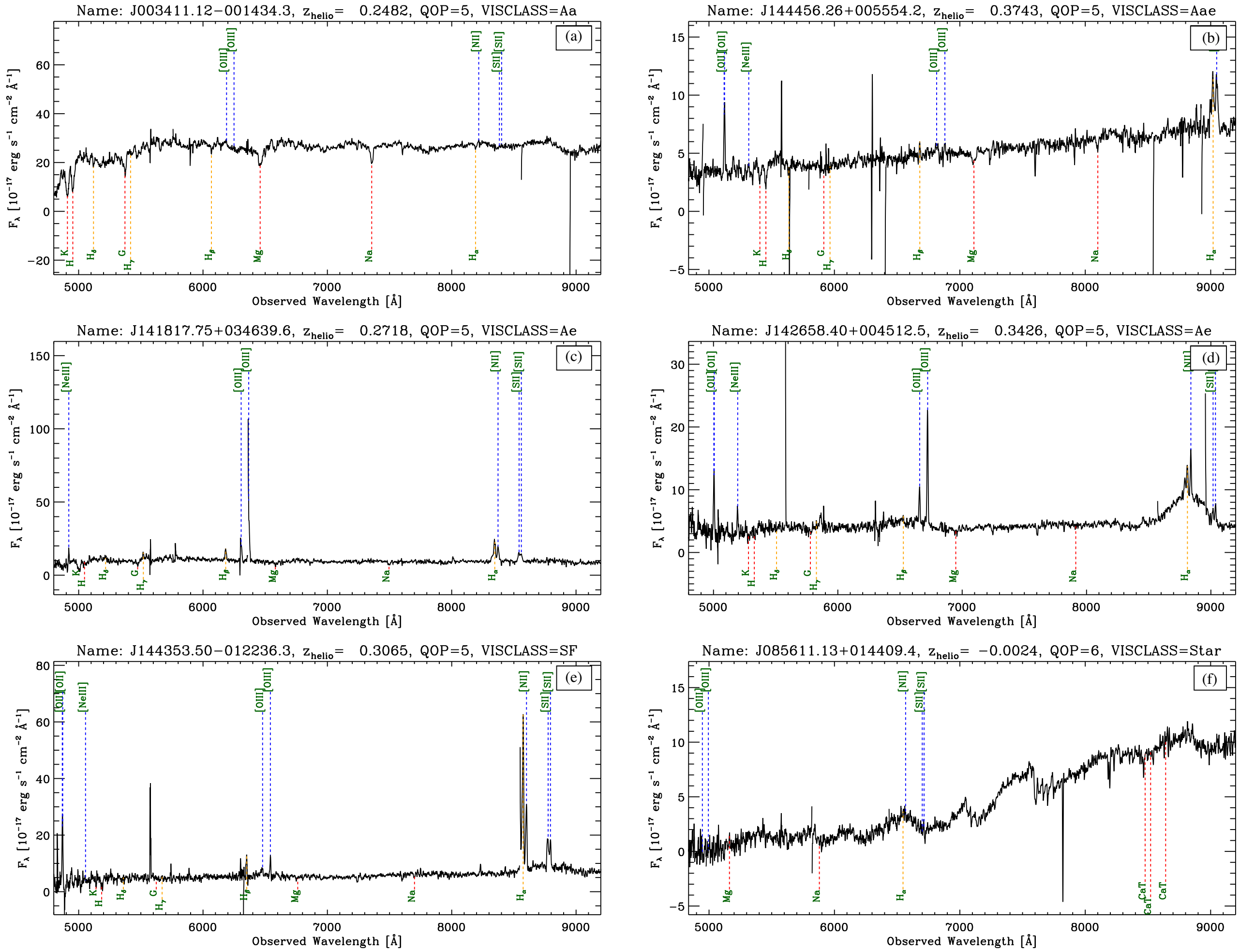
Ae. Stellar continuum with strong narrow emission lines where the [O III] λ5007 and [N II] λ6583 emission lines appear stronger than, or comparable to, the Hβ and Hα emission lines, respectively.

AeB. Broad emission lines typical of a quasar spectrum.

SF. Similar to *Ae*, but instead the Balmer emission lines appear stronger than the forbidden lines.

Star. Galactic star.

BL Lac. Featureless optical spectrum with no obvious emission or absorption lines.


Name: J003411.12−001434.3, z_{helio}= 0.2482, QOP=5, VISCLASS=Aa

Name: J144456.26+005554.2, z_{helio}= 0.3743, QOP=5, VISCLASS=Aae

Name: J141817.75+034639.6, z_{helio}= 0.2718, QOP=5, VISCLASS=Ae

Name: J142658.40+004512.5, z_{helio}= 0.3426, QOP=5, VISCLASS=Ae

Name: J144353.50−012236.3, z_{helio}= 0.3065, QOP=5, VISCLASS=SF

Name: J085611.13+014409.4, z_{helio}= −0.0024, QOP=6, VISCLASS=Star

Figure 10. Example spectra for different visual classifications (`VISCLASS`): (a) Aa, strong absorption lines and no emission lines visible; (b) Aae, strong absorption lines and weak emission lines; (c) Ae, strong emission lines with line ratios characteristic of an AGN; (d) AeB, AGN with broad emission lines; (e) SF, strong emission lines with line ratios characteristic of star-forming galaxies; (f) Star, Galactic star spectrum. Note that there are bad pixels or cosmic ray contamination in some spectra.

Unusual. Any object with a spectrum that does not fit into the categories above. This category includes some radio-source hosts with weak emission lines superimposed on a strong, featureless continuum as well as broad-absorption line quasars (BAL QSOs and a few post-starburst galaxies with strong Balmer absorption lines.

The SDSS data base also includes flags for different spectral classifications, based on the best-fitting template to each SDSS spectrum. SDSS `specClass` flags of 3 and 4 correspond to QSO and high-redshift ($z > 2.3$) QSO, respectively (Stoughton et al. 2002), and for SDSS spectra with `specClass` flag of 3 or 4, we set `VISCLASS` to *AeB*. Similarly, we incorporate the visual classification of QSO provided by the 2SLAQ-QSO and 2QZ surveys into our `VISCLASS` flag. Finally, we set the `VISCLASS` to NA (null) for the remaining sources where no visual classification was made.

7.2 Quantitative (automated) spectral classification

Baldwin, Phillips & Terlevich (1981, hereafter BPT) devised a method to distinguish between the emission lines originating from an AGN and star formation by comparing the ratio of specific forbidden lines to neighbouring Balmer lines.

We used the BPT technique to carry out an automated spectral classification for objects with *QOP* $\geq$ 3 if the [O III] λ5007 emission line fell within the observed spectral range. In practice, this imposes a redshift limit of $z < 0.768$ for GAMA spectra and $z < 0.838$ for WiggleZ and SDSS spectra. 75 per cent of the 10 856 LARGESS objects with reliable redshifts also have an automated spectral classification.

7.2.1 *Classifying WiggleZ spectra*

As explained in Section 5.6.2, we did not measure absorption-corrected emission-line ratios for objects with WiggleZ spectra because of difficulties fitting the underlying stellar continuum in a reliable way. As a result, we could not use the BPT diagram to classify our WiggleZ spectra because we were not able to correct the Balmer emission lines for any underlying stellar absorption.

For this reason, we only used the WiggleZ spectra to classify objects with $L_{\rm FIRST} > 10^{24}$ W Hz^{-1} or $z > 0.3$ – which can be assumed to be radio-loud AGN (see Section 7.2.3 below). For these objects, we used the [O III] equivalent width (which is not significantly affected by stellar absorption) to separate low-and high-excitation

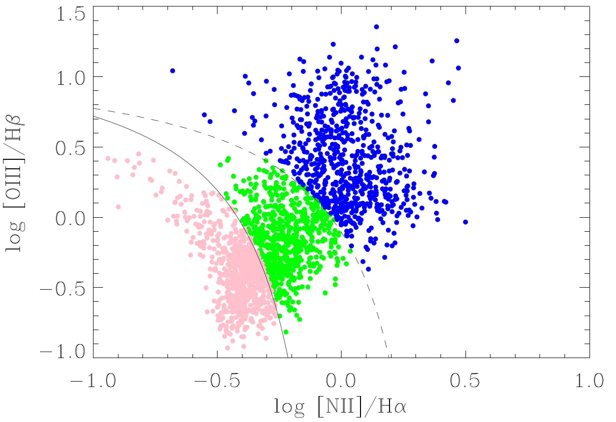


Figure 11. The BPT emission-line diagnostic diagram for LARGESS sources with SNR > 3 in all four relevant emission lines. All objects below the Kauffmann et al. (2003) limit (solid line; pink points) are assumed to have emission lines produced by ionizing radiation from star formation regions. Galaxies above the Kewley et al. (2001) limit (dashed line) have emission lines arising from gas ionized by an AGN. Galaxies in between the Kewley et al. (2001) and Kauffmann et al. (2003) limits (green points) are generally considered to be composite galaxies with some ongoing star formation.

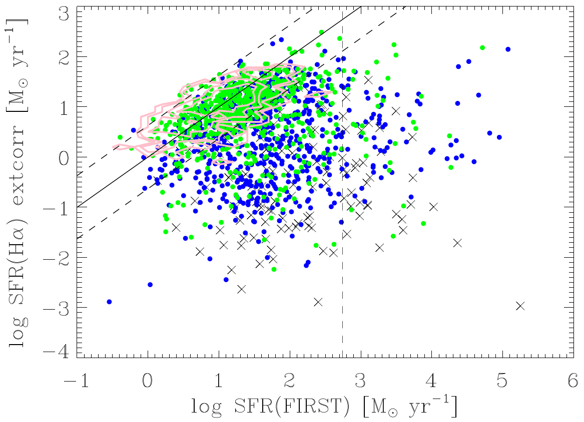


Figure 12. Comparison between the SFR derived from Hα to the SFR derived from the total FIRST flux density using the relation in Hopkins et al. (2003). The green and blue points are, respectively, AGN and composite galaxies as defined by a BPT emission line diagnostic. We overlay contours for SF galaxies in pink. The crosses indicate galaxies that were not able to go on a BPT diagram, but have significant Hα and Hβ to estimate the SFR(Hα). The solid diagonal line is a one-to-one line, and the 3σ limits to the relationship from Hopkins et al. (2003) are shown as the diagonal dashed lines. The vertical dashed line is the SFR inferred at a radio flux of 10^{24} W Hz^{-1}, beyond which star formation processes are unlikely to dominate the observed radio emission in our sample.

radio galaxies. For galaxies with spectra from GAMA and SDSS, we used the BPT diagram as described below.

7.2.2 *Galactic stars*

Objects with a reliable redshift of $z < 0.002$ are classified as Galactic stars. They may be either a genuine association of a FIRST radio source with a Galactic object or (more likely) a random superposition of a foreground star against a background radio source. They make up ∼2 per cent of LARGESS objects with a reliable redshift.

7.2.3 *Star-forming galaxies*

We expect that galaxies in our sample whose radio emission is dominated by processes related to star formation rather than an AGN will lie at redshift $z \leq 0.3$ and have 1.4 GHz radio luminosity $L_{\rm FIRST} \leq 10^{24}$ W Hz^{-1} (similar limits were chosen by Best & Heckman 2012)), since the inferred star formation rate needed to produce the observed radio emission would otherwise be unrealistically high. We also expect galaxies whose radio emission arises mainly from star formation processes to obey the Hopkins et al. (2003) relation between Hα and radio luminosity, since both these quantities are proxies for the star formation rate.

We first used the BPT diagnostic plot to identify objects with star-forming (SF) optical spectra (see Fig. 11). For galaxies with a signal-to-noise ratio (SNR) >3 in each of the [O III] λ5007, [N II] λ6583, Hα and Hβ lines, we define SF galaxies (pink points in Fig. 11) as those in the 'pure' SF region of Kauffmann et al. (2003).

For all galaxies not already classified as SF in the BPT diagnostic plot, we then compared the star formation rate estimates inferred from the Hα line with the star formation rate estimates inferred from the 1.4 GHz radio luminosity using the relations from Hopkins et al. (2003).

Galaxies where the optical spectrum was classified as an AGN in the BPT diagram were reclassified as SF galaxies if their radio luminosity placed them within 3σ of the one-to-one relation in Fig. 12, based on the methodology used in Bardelli et al. (2010). This allows us to identify the dominant process for the radio emission from SF galaxies that also contain a radio-quiet AGN. The remaining objects with an AGN spectral classification in the BPT diagram constitute a robust sample of radio-loud AGN.

7.2.4 *Radio-loud AGN: separating HERGs and LERGs*

From this robust sample of radio-loud AGN, we separated low- and high-excitation radio galaxies using a cut in [O III] λ5007 equivalent width (EW([O III] λ5007)).

We defined high-excitation radio galaxies (HERGs) as those with SNR([O III] λ5007) >3 and EW([O III] λ5007) >5Å. The choice of an EW([O III] λ5007) >5Å cutoff is based on a comparison of EW([O III] λ5007) with the visual classification (see Fig. 13), and is the same cut-off value used by Best & Heckman (2012) to separate HERGs and LERGs in their SDSS sample. All other radio-loud AGN were classified as low-excitation radio galaxies (LERGs). As can be seen from Fig. 13, this dividing line at EW([O III] λ5007) >5Å also gives results that are generally consistent with our visual Aa and Ae classification.

7.3 Comparison of the automated and visual spectral classifications

Table 7 shows a pairwise comparison between the qualitative (visual) and quantitative (automated) classifications for the 4058 LARGESS objects for which both measures are available.

For objects with weak or no emission lines, there is very good agreement between the Aa visual classification and the LERG automated classification. Some objects that were visually identified as having optical emission lines (i.e. class Aae, Ae and SF) are also classified as LERGs by the automated criteria. This is probably because the human eye is able to recognize weak emission lines that fall below the 5Å EW([O III] λ5007) limit used to separate HERGs and LERGs in the automated classification. Most visual Aae objects

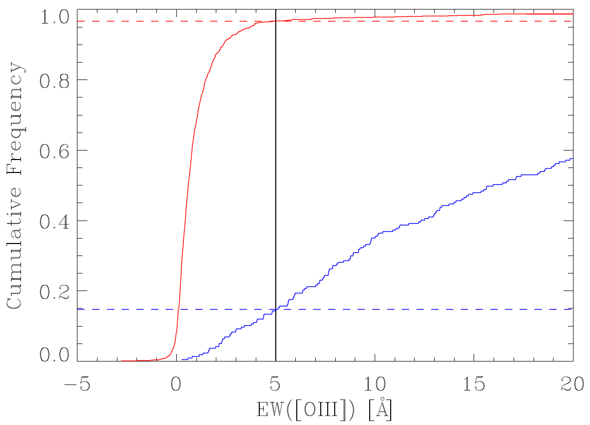


Figure 13. Cumulative distribution of [O III] λ5007 equivalent width for galaxies with `VISCLASS` Aa (absorption-line spectrum, red line) and Ae (spectrum with strong optical emission lines, blue line). The vertical line shows the 5Å EW([O III] λ5007) limit used to separate HERGs and LERGs. The horizontal red and blue dashed lines are the fraction of sources below this limit for Aa and Ae galaxies, respectively.

Table 7. Comparison between the visual spectral classification and the automated classification for the 4058 objects with both types of classifications.

Automated classification	Visual classification Aa	Aae	Ae	SF	Star	Total
LERG	2474	360	121	132	0	3087
HERG	31	78	447	108	0	664
SF	1	6	23	185	0	215
Star	0	0	0	0	92	92
Total	2506	444	591	425	92	4058

(∼81 per cent) have an automated classification as LERGs, while the great majority of visual Ae objects (∼80 per cent) are classified as HERGs. We therefore find a strong consistency between the visual and automated classifications for the optical spectra of radio AGN.

For objects with stronger emission lines, the AGN/SF classification also appears robust in most cases. Of 591 objects classified visually as emission-line AGN (class Ae), only 23 (3.9 per cent) were reclassified as SF galaxies based on the BPT diagram and Hα/radio continuum comparison. For the visual AGN sample as a whole (classes Aa, Aae and Ae combined) the fraction reclassified as SF is <1 per cent. We therefore conclude that the overall level of contamination of our AGN sample by SF objects is very low, and that our separation of AGN and SF radio sources is generally self-consistent and reliable.

7.4 Final spectroscopic classifications

The best spectral classification (`BESTCLASS`) is a combination of both visual and automated classifications. All sources flagged as either Star, AeB, Unusual or BLLac in the `VISCLASS` are also flagged as such in the `BESTCLASS`. In all other cases, the `BESTCLASS` is set to the automated classification as described above, or set to ‘NA’ if an automated classification is unavailable.

Table 8 summarizes the final spectroscopic classifications for the full sample. The combination of radio and optical flux limits for our sample means that most of the objects detected above redshift $z = 0.8$ are radio-loud QSOs (class AeB), so we also list the classifications of objects with $z \leq 0.8$ and $z > 0.8$ separately. Low-excitation radio AGN (LERGs) are the dominant population, accounting for almost 70 per cent of the objects at $z \leq 0.8$. The optical and mid-infrared properties of the spectroscopic sample are discussed in more detail in Section 9.

Table 8. Final spectroscopic classifications for the 10 856 LARGESS objects with a reliable ($Q \geq 3$) redshift measurement.

Class	Redshift All	$z \leq 0.8$	$z > 0.8$
LERG	5881	5864	17
HERG	839	827	12
AeB	1615	397	1218
SF	1415	1415	–
Star	196	196	–
BL Lac	19	5	14
Unusual	61	29	32
Unclassified (NA)	830	729	101
Total	10856	9462	1394

8 THE SPECTROSCOPIC DATA TABLE

Tables 9 and 10 present the final LARGESS spectroscopic data catalogue, ordered by Right Ascension. Table 9 describes each column of the data table, 20 lines of which are shown in Table 10. These are the first 20 objects after RA 09:00:00, an RA range that lies within one of our high-completeness GAMA fields (see Fig. 1). The table includes optical positions and unique identifiers for each radio target. We also include extinction-corrected SDSS g, r and i photometry and their errors, along with the total radio flux from FIRST (and NVSS if available) and the number of radio components associated with each optical object. The spectroscopic parameters included are the best redshift, quality code and the origin of the spectrum used for the final redshift.

As explained in Section 7, there are two spectroscopic classifications designed to separate SF galaxies from low- and high-excitation AGN. The first is a qualitative method based on visual inspection of each spectrum (`VISCLASS`), and the second is a final best classification based on both the automated and visual methods (`BESTCLASS`). Additionally we provide three flags: `HI_COMP` to indicate if a source is in a region with high spectroscopic completeness, `ZSOURCE_TARGET` to indicate if a source is a filler target explicitly observed for the LARGESS sample by the GAMA or WiggleZ team, and `DISAGREE_GAMA`, which indicates that our best redshift/quality is estimated from a GAMA spectrum, but differs from the previous GAMA AUTOZ (internal data: AATSpecAutozAllv22) redshift or quality.

9 LARGESS SAMPLE CHARACTERISTICS

We now discuss some general properties of the objects in the LARGESS data catalogue.

9.1 Optical colour versus redshift

Fig. 14 shows the observed ($g_{\rm mod} - i_{\rm mod}$) optical colour as a function of redshift for all LARGESS objects with a reliable spectroscopic redshift and classification, split into the four main spectral classes (HERG, LERG, SF and AeB). This is a key plot for the LARGESS sample, and shows the relationship between optical colour and spectroscopic class for the full range of radio-selected AGN out to redshift $z > 0.8$.

Table 9. Description of the columns in the main LARGESS data table (Table 10). The format codes are FORTRAN format descriptors. All SDSS photometric data are from the SDSS sixth data release.

Col.	Field	Format	Units	Description
1	NAME	a19	–	IAU format object name
2	SDSSID	i18	–	SDSS photometric ID
3	RA	f9.5	deg	SDSS RA J2000 in decimal degrees
4	DEC	f9.5	deg	SDSS Dec J2000 in decimal degrees
5	R_PET	f6.3	mag	SDSS Petrosian magnitude in *r* band (extinction corrected)
6	R_PET_ERR	f7.3	mag	SDSS Petrosian magnitude error in *r* band
7	I_MOD	f6.3	mag	SDSS Model magnitude in *i* band (extinction corrected)
8	I_MOD_ERR	f6.3	mag	SDSS Model magnitude error in *i* band
9	G_MOD	f6.3	mag	SDSS Model magnitude in *g* band (extinction corrected)
10	G_MOD_ERR	f6.3	mag	SDSS Model magnitude error in *g* band
11	N_FIRST	i2	–	Number of FIRST components
12	FIRST_TOT	f7.2	mJy	FIRST total integrated flux
13	N_NVSS	i1	–	Number of NVSS components; -1 = Null value
14	NVSS_TOT	f9.3	mJy	NVSS total integrated flux; -99.000 = Null value
15	Z	f9.5	–	Final best redshift; NaN = Null value
16	QOP	i3	–	Redshift reliability flag
17	ZSOURCE	a12	–	Survey source for the best redshift; NA = Null value
18	OIII_SN	e9.2	–	[O III] λ5007 SNR; NaN = Null value
19	EW_OIII	e9.2	Å	[O III] λ5007 equivalent width; NaN = Null value
20	VISCLASS	a13	–	Best visual classification; NA = Null value
21	BESTCLASS	a13	–	Best spectroscopic classification; NA = Null value
22	HI_COMP	i1	–	High-completeness region flag; 1 = in region; 0 = not in region
23	ZSOURCE_TARGET	i2	–	Radio filler target flag for GAMA/WiggleZ; 1 = filler; 0 = not filler; -1 = `ZSOURCE` is not GAMA/WiggleZ
24	DISAGREE_GAMA	i2	–	Flag to indicate whether the listed redshift and quality code agree with the GAMA AUTOZ (internal data: AATSpecAutozAllv22) estimate; 1 = disagree; 0 = agree; -1 = `ZSOURCE` is not GAMA

The broad emission-line (AeB) objects have the bluest colours at all redshifts. At redshifts above $z \sim 0.25$ their $(g_{\rm mod} - i_{\rm mod})$ colour is relatively flat as a function of redshift, and most AeB objects lie close to the track for optically selected SDSS QSOs (the dotted line in Fig. 14), implying that their optical light is dominated by a non-stellar power-law spectrum (Peterson 1997). At lower redshift ($z < 0.25$), most of our AeB objects have significantly redder colours than the track for optically selected QSOs. This is probably due to a higher contribution from the host galaxy stellar light, since the AeB objects in our sample are selected via their radio emission (and are spectroscopically defined), rather than being colour-selected like the SDSS QSOs. In summary, most of the AeB objects in the LARGESS catalogue have $(g_{\rm mod} - i_{\rm mod})$ colours similar to those of optically selected QSOs.

The LERGs in our sample are typically the reddest objects at all redshifts, with redder $(g_{\rm mod} - i_{\rm mod})$ colours at higher redshift. This is consistent with most LERG hosts being galaxies with an old, passively evolving stellar population, i.e. Luminous Red Galaxies (LRGs; Eisenstein et al. 2001). Fig. 14 also shows colour–redshift tracks for the two LRG models used by Wake et al. (2006) and derived using the Bruzual & Charlot (1993) stellar population synthesis code. Model 1 (green dashed line) is for a single 10 Gyr starburst, and evolves passively without any further star formation. Model 2 (grey dash–dotted line) has 95 per cent of the final mass in a single burst and 5 per cent as a continuous level of star formation. Most LERGs lie near or in between these tracks, though a few LERGs scatter to much redder and bluer colours (especially at higher redshift) than models 1 and 2, respectively.

The HERGs generally lie in between the colours of the AeB and LERG classes at all redshifts, with bluer colours than the LRG model 2 and redder colours than typical QSOs. There are several plausible reasons why HERG host galaxies might have bluer optical colours than LERG hosts at the same redshift.

(i) Some optical light may come from a blue AGN continuum, at a lower level than seen in the AeB objects, even though broad Balmer emission lines are not observed. In unified AGN models (e.g. Antonucci 1993), these would be objects where the central dusty torus has a smaller opening angle than in the AeB systems but some AGN light can still be seen.

(ii) The HERG host galaxies may have some ongoing star formation, and contain a substantial young or intermediate-age stellar population, even though their observed radio emission is produced mainly by the central AGN. Some supporting evidence for this comes from stellar-population studies of local (Best & Heckman 2012) and more distant (Johnston et al. 2008) radio AGN, which find consistent evidence for a younger stellar population in radio galaxies with strong emission lines. Johnston et al. (2008) also found that the composite spectra of emission-radio AGN at $z \sim 0.55$ were better fitted by a mixture of an old plus intermediate-age stellar population, rather than an old population with a non-stellar AGN continuum.

(iii) If the HERG hosts are typically lower mass galaxies than the LERG hosts (as found for low-redshift systems by e.g. Best & Heckman 2012), then they would also be expected to have lower metallicity and slightly bluer colours (Tremonti et al. 2004).

Fig. 14 does not allow us to distinguish between these possibilities, and in the next section (Section 9.2) we consider the additional information provided by mid-infrared photometry. We also discuss the relationship between HERGs and AeB objects further in Section 10.1, and in Sections 10.2 and 10.3 we compare the properties

Table 10. A list of 20 sample objects from the main LARGESS data table – the full table contains 19 179 sources. For a description of the data in each column, see Table 9.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
J090001.05-000852.8	588848899892969855	135.004 41	−0.148 02	19.507	0.063	18.814	0.020	21.270	0.084	1	4.77
J090001.28+053602.1	587732703391777311	135.005 34	5.600 60	19.676	0.059	19.120	0.027	21.212	0.079	1	4.12
J090001.85+022231.7	587727944564277289	135.007 72	2.375 49	17.587	0.332	17.288	0.010	19.749	0.036	1	3.94
J090002.65-003338.6	588848899356098923	135.011 06	−0.560 73	20.494	0.061	20.443	0.045	20.871	0.037	1	14.49
J090003.71+073056.7	587735343184937333	135.015 47	7.515 75	18.323	0.034	17.565	0.012	19.887	0.043	3	36.79
J090004.24+033318.4	588010359603463018	135.017 68	3.555 14	20.802	0.185	19.544	0.034	22.483	0.212	1	6.21
J090004.52-002548.7	588848899356099571	135.018 83	−0.430 20	20.953	0.185	19.836	0.043	23.021	0.356	1	14.36
J090004.66+000332.1	587725074990235700	135.019 43	0.058 93	18.129	0.023	17.541	0.009	19.487	0.020	1	12.89
J090005.05+000446.7	587725074990235725	135.021 06	0.079 66	15.146	0.009	14.621	0.003	15.721	0.003	1	5.33
J090005.85+073634.0	587735343185003255	135.024 39	7.609 45	21.164	0.220	20.481	0.075	24.924	1.026	2	24.37
J090005.87+072548.4	587734948047356113	135.024 48	7.430 12	17.793	0.021	17.123	0.009	18.817	0.019	1	6.92
J090006.26+021537.3	587727944564277524	135.026 12	2.260 38	20.031	0.045	19.751	0.029	19.973	0.020	1	24.36
J090006.43+022404.2	587727944564277661	135.026 81	2.401 19	19.238	0.045	18.528	0.019	20.823	0.057	2	15.75
J090008.02+033945.3	587728880868852173	135.033 42	3.662 59	20.226	0.069	19.692	0.033	21.559	0.077	1	9.86
J090010.12+023643.8	588010358529720530	135.042 18	2.612 18	19.518	0.218	19.308	0.027	21.353	0.080	2	12.06
J090010.57+080548.0	587735343721874462	135.044 04	8.096 67	20.998	0.231	19.721	0.053	22.750	0.407	1	1.06
J090011.14+050257.6	587732578298888642	135.046 42	5.049 36	20.094	0.066	18.955	0.025	21.739	0.131	1	5.02
J090012.53+023539.1	588010358529720551	135.052 21	2.594 22	17.335	0.015	16.922	0.007	18.148	0.010	1	1.08
J090013.92+024717.0	588010358529720589	135.058 03	2.788 06	19.558	0.028	19.392	0.019	19.858	0.018	1	144.76
J090014.01+053549.7	587732703391777613	135.058 40	5.597 15	21.996	0.291	20.418	0.084	23.186	0.449	1	1.45

(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
1	5.051	0.409 06	4	WiggleZ	0.52	0.57	Aa	LERG	1	1	−1
1	3.100	0.302 67	4	WiggleZ	23.20	31.90	Ae	HERG	1	1	−1
1	6.000	0.250 91	4	SDSS	2.27	0.73	Aa	LERG	1	−1	−1
1	13.700	1.007 83	4	GAMA	2.21	161.00	AeB	AeB	1	1	1
1	41.200	0.384 34	5	SDSS	1.46	0.23	NA	LERG	1	−1	−1
1	7.846	0.641 48	3	WiggleZ	− 0.73	− 3.18	Aa	LERG	1	1	−1
1	14.600	0.610 92	4	WiggleZ	1.43	–	Aa	LERG	1	1	−1
1	56.000	0.262 21	4	GAMA	6.65	3.05	Aa	LERG	1	0	0
0	− 99.000	0.053 86	4	SDSS	22.00	2.74	SF	SF	1	−1	−1
1	29.400	0.000 00	−1	NA	–	–	NA	NA	1	−1	−1
1	7.500	0.000 00	−1	NA	–	–	NA	NA	1	−1	−1
1	23.647	0.616 72	4	SDSS	–	–	AeB	AeB	1	−1	−1
1	21.600	0.349 38	4	GAMA	0.54	0.27	Aa	LERG	1	0	0
1	10.400	0.357 08	4	WiggleZ	1.39	1.17	Aae	LERG	1	1	−1
1	13.500	0.200 71	3	GAMA	1.12	–	Aa	LERG	1	0	0
0	− 99.000	0.000 00	−1	NA	–	–	NA	NA	1	−1	−1
1	5.300	0.000 00	−1	NA	–	–	NA	NA	1	−1	−1
0	− 99.000	0.200 00	4	SDSS	6.08	1.86	SF	SF	1	−1	−1
1	140.700	1.189 78	4	SDSS	–	–	AeB	AeB	1	−1	−1
0	− 99.000	0.757 97	4	WiggleZ	8.04	21.90	Ae	HERG	1	1	−1

of HERG and LERG host galaxies using sub-samples matched in stellar mass, radio luminosity and redshift.

Finally the SF class, which is only present in the LARGESS sample at low redshift ($z < 0.3$), shows a spread of $(g - i)$ values that is consistent with a range in both current star formation rate and dominant stellar populations in these systems (see e.g. Taylor et al. 2011).

9.2 *WISE* mid-infrared colours

The *Wide-field Infrared Survey Explorer* (*WISE*; Wright et al. 2010) is an all-sky mid-infrared imaging survey. It provides photometry in four bands (*W*1–4) centred at 3.4, 4.6, 12 and 22 μm, with angular resolutions (FWHMs) of 6.1, 6.4, 6.5 and 12.0 arcsec in each band, respectively.

For extragalactic objects, the *WISE* data can provide useful information about recent star formation activity (vibrationally excited polycyclic aromatic hydrocarbon emission; rest-frame 6.2, 7.7, 8.6 and 11.3 μm), silicate absorption (rest-frame 10μm) and dust heated by strong radiation fields (e.g. AGN accretion discs; rest-frame 14–16μm). At shorter rest-wavelengths (1.6–4.5μm), the *WISE W*1 and *W*2 bands typically trace the Rayleigh–Jeans ($F_\nu \propto \nu^2$) tail of the old stellar population, which may be used to estimate the stellar mass of galaxies out to moderate redshifts (Xilouris et al. 2004; Wilman et al. 2008; Hwang et al. 2012). The W3 band in particular may also be strongly affected by the power-law emission ($F_\nu \propto \nu^{-\alpha}$) of an AGN (Jarrett et al. 2011).

To examine the mid-infrared properties of the LARGESS sample, we cross-matched the full data catalogue with the all-sky All-WISE data release[3] (Cutri et al. 2013), using a 3 arcsec matching radius. The great majority of LARGESS objects (18 124, or 94.5 per cent) have a *WISE* match, and 11 814 of these (95.8 per cent of the LARGESS objects with good-quality spectra) have a reliable

---

[3] http://wise2.ipac.caltech.edu/docs/release/allwise/

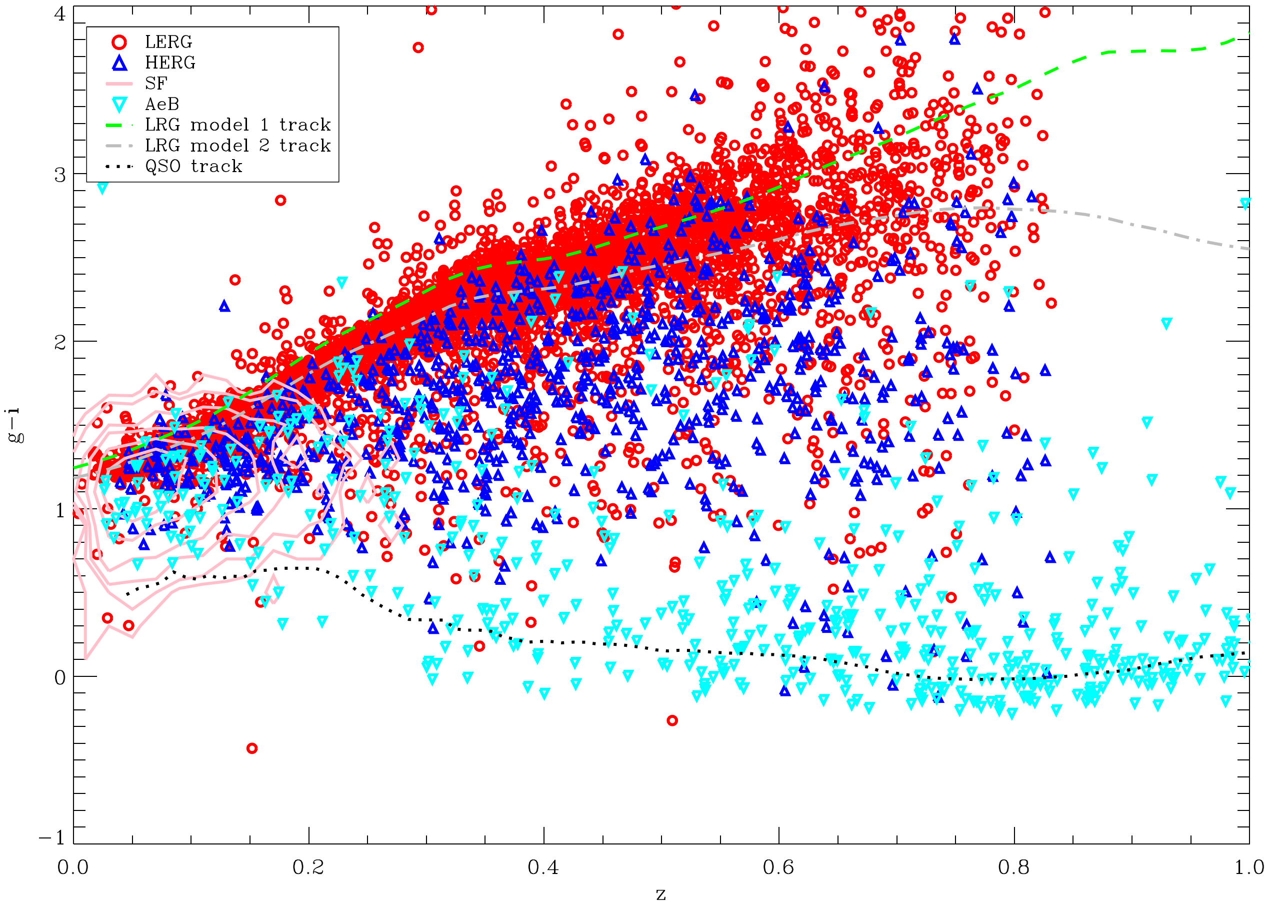


Figure 14. The ($g_{\rm mod} - i_{\rm mod}$) colour as a function of redshift for the LERGs (red open circles), HERGs (blue open triangles), AeB objects with strong, broad Balmer emission lines (cyan open inverted triangles) and contours for SF galaxies. We show the tracks of two LRG models from Wake et al. (2006, model 1, green dashed line; model 2, grey dash–dotted line; see text for details) and median observed colours for QSOs (dotted line) with $i_{\rm PSF}$ < 18.0 from the SDSS DR3 QSO catalogue (Schneider et al. 2005; Croom et al. 2009) on this plane. All three classes of radio galaxies have different spectral energy distributions (SEDs). The distribution of colour with redshift for AeB sources is quite flat with redshift suggesting a power-law SED, which is consistent with them following the QSO track. LERGs tend to follow the LRG track. The HERGs lie somewhere in between the LERG and AeB classes.

Table 11. Number of LARGESS sources with reliable *WISE* photometry in each band. This is split for: the full sample, those with spectra, and those with reliable redshifts. The final row gives the number with reliable *WISE* photometry in all four bands.

WISE band	Full sample	With spectra All	With spectra QOP ≥ 3
*W*1	15 926	10 075	8784
*W*2	17 213	11 194	9902
*W*3	7782	5609	5317
*W*4	4908	3599	3461
All	3585	2597	2503

redshift. The reliability of the *WISE* cross-matching is estimated as ∼99 per cent.

In the following analysis we only consider objects with a reliable detection in each *WISE* band, i.e. we require that the $\rm SNR_{WISE} \geq 2$, reduced chi-square of the profile fit $\chi^2_{\rm WISE} \leq 3$, and there is no contamination and confusion flag (i.e. `cc_flag` = 0), for that band. Table 11 shows the number of objects that satisfy this requirement for each *WISE* band. The highest number of reliable detections is in the *W*2 band, which is typically dominated by emission from the stellar galaxy. Although the *W*1 band is the most sensitive, it does not have the highest detection rate because the spectral energy distribution of AGN has a positive gradient (i.e. higher flux at longer wavelength) for wavelengths between 1 and 13 μm (see Assef et al. 2010), so some AGNs may be detected in *W*2 but not *W*1. The lowest number of detections is in the *W*4 band that is the least sensitive band and mainly detects emission from hot dust.

9.3 The *WISE* two-colour diagram

Since the release of the *WISE* data catalogue (Wright et al. 2010), many groups have used the unique wavelength coverage and sensitivity of *WISE* to identify and classify objects of interest for further study, including both Galactic (e.g. Cushing et al. 2011; Kirkpatrick et al. 2011) and extragalactic objects (e.g. Lake et al. 2012). For QSOs, the *WISE* data can be used to identify both obscured and unobscured objects (e.g. Jarrett et al. 2011; Edelson & Malkan 2012; Mateos et al. 2012; Stern et al. 2012; Bridge et al. 2013).

The *WISE* (*W*1−*W*2) versus (*W*2−*W*3) two-colour diagram is a powerful diagnostic tool for identifying the dominant source of MIR emission in individual objects. The diagram was first introduced by Wright et al. (2010), who mapped out the regions occupied by different types of extragalactic sources by combining data from the Spitzer Wide-area InfraRed Extragalactic (SWIRE) survey with the

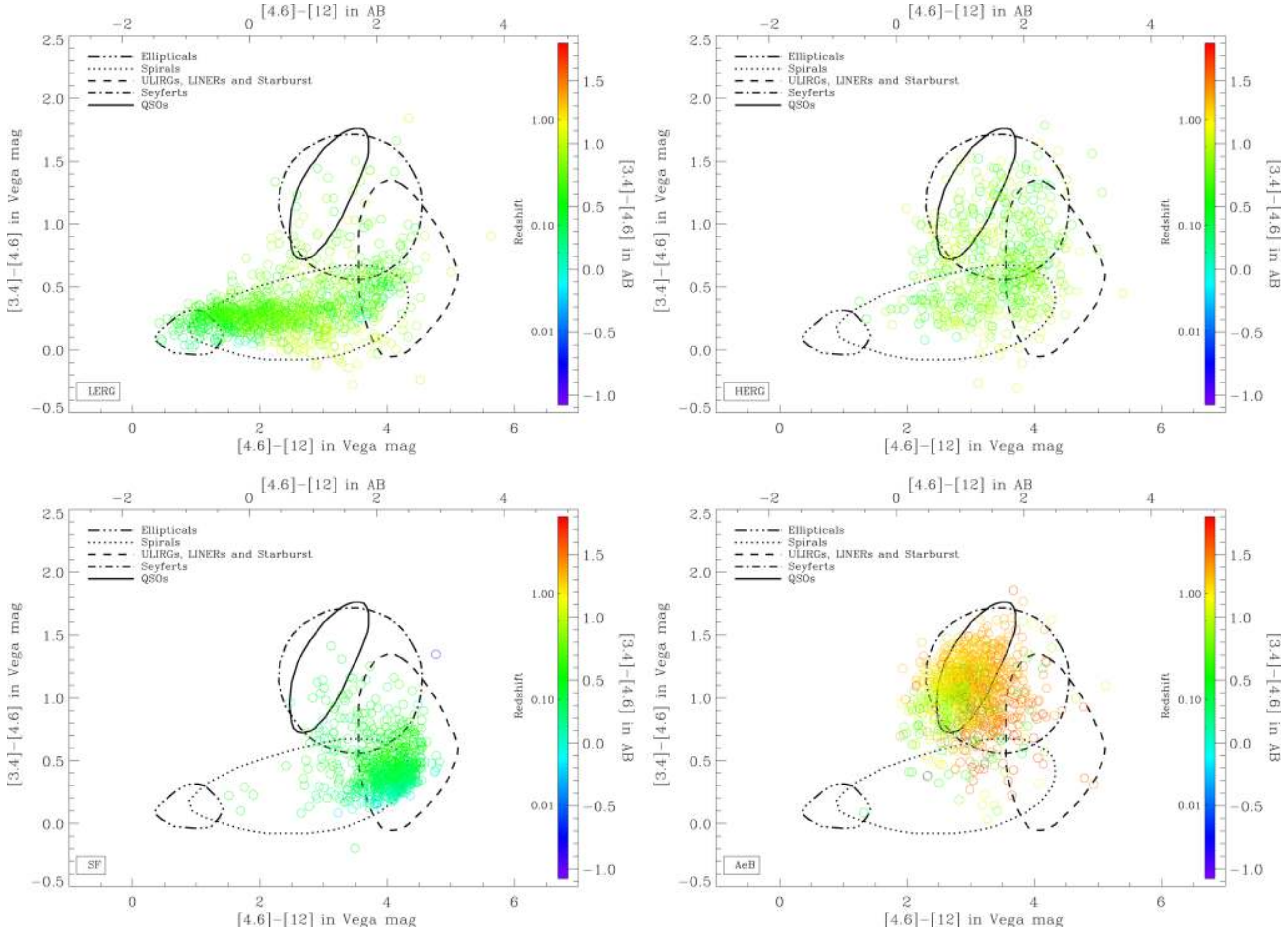


Figure 15. *WISE* mid-infrared two-colour diagrams for LARGESS objects split by spectral classification. Each plot shows the *WISE* ($W1-W2$) colour against ($W2-W3$) colour. The conversion from Vega magnitude to the AB magnitude system assumes a constant power law for the *WISE* bands (Jarrett et al. 2011). The colour of individual data points is scaled with redshift as indicted on the right of each plot, and dashed and dotted lines highlight the expected locations of different classes of astronomical objects (Wright et al. 2010).

models generated from the GRASIL (GRAphite and SILicate; Silva et al. 1998) code. These regions are highlighted by dashed and dotted lines in Fig. 15.

The power of this diagram lies in the fact that the *WISE* ($W2-W3$) colour is a useful proxy for specific star formation rate (Donoso et al. 2012), while the orthogonal ($W1-W2$) colour reflects the fractional contribution of non-stellar emission from an AGN. Stern et al. (2012) have shown that most AGN with a classical accretion disc can be reliably selected using a single colour cut of $(W1-W2)\geq 0.8$, for *WISE* detections with $\mathrm{SNR}(W2) > 10$ and $W2 < 15.05$ mag. In the following analysis, we only consider LARGESS objects with a reliable detection in each of the $W1$, $W2$ and $W3$ bands.

Fig. 15 shows the LARGESS objects in the *WISE* two-colour diagram, separated by their spectral classification and colour-coded by redshift. In these plots, LERGs are the only objects found in the 'Ellipticals' region, as expected for radio sources hosted by a massive galaxy with an old stellar population. The LERGs also extend well into the 'Spirals' region, especially at higher redshift, implying that many of them contain a warm dust component in addition to an old stellar population. This is not too surprising, since modest amounts of dust are known to be common in normal elliptical galaxies (e.g. Sadler & Gerhard 1985; Rowlands et al. 2012). A warm dust component can arise from a range of processes (e.g. weak star formation or heating by post-AGB stars) that are not dominant at radio wavelengths or related to an accretion disc, and so do not contradict the LERG classification.

The SF and AeB objects fall mainly within the 'Starburst' and 'QSO' regions, respectively. For all three of the LERG, SF and AeB classes, there is a high degree of consistency between our optical spectroscopic classification and the position of most objects in the *WISE* two-colour plot.

The situation is rather different for the HERGs. As can be seen from Fig. 15, these objects are very scattered in the colour–colour plane, and span most of the 'Seyfert', 'Spiral' and 'Starburst' regions. The diverse mid-infrared (MIR) colours of the HERG population strongly suggest that this is a heterogeneous class of objects in which the physical process dominating the optical/MIR light is not same for all members.

The broad dispersion of HERG host-galaxy colours seen in Fig. 15 is perhaps not surprising, since the $W1-W2$ colours of the Seyfert galaxies are known to depend on the relative contribution of starlight by the host galaxy compared to the AGN contribution as well as the amount of dust extinction (e.g. Stern et al. 2012, and references therein). It therefore seems likely that the position of individual HERGS in the *WISE* two-colour plot depends on both star formation rate and obscuration within the

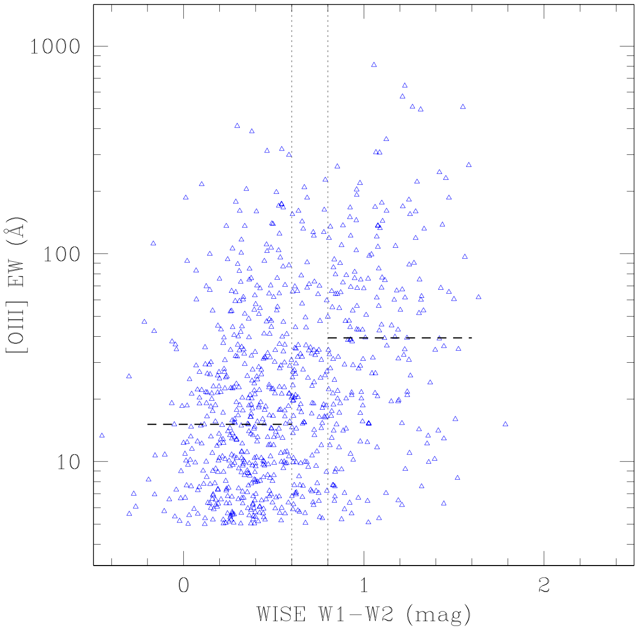


Figure 16. Comparison of the optical [O III] emission-line width and *WISE* *W*1-*W*2 colour for HERG host galaxies in the LARGESS sample, The vertical dotted lines separate objects in which the mid-IR light is dominated by the stellar galaxy (leftwards of the (*W*1-*W*2)=0.6 line) from those where the AGN dominates (to the right of the (*W*1-*W*2)=0.8 line). Horizontal dashed lines show the median value of EW [O III] for these two sub-groups.

host galaxy and the current accretion rate of gas on to the central black hole.

As in Fig. 14, LERGs are shown by red open circles, HERGs are blue open triangles and AeB objects by cyan inverted triangles. SF galaxies are plotted as pink stars.as indicated by the colour-coded labels on the left of the plot. The solid black box is the region defined by Jarrett et al. (2011) to select QSOs.

Some additional supporting evidence for this picture comes from Fig. 16. This plots the *WISE* (*W*1-*W*2) colour for HERGs in our sample against the equivalent width of the [O III] 5007 Å emission line – that can be used as a rough proxy for AGN accretion rate (Kauffmann et al. 2003; Heckman et al. 2005; Trouille & Barger 2010). In general, the light of the underlying stellar population dominates the mid-IR SED in objects with *W*1-*W*2<0.6 mag, while the AGN light dominates in objects with (*W*1-*W*2)>0.8 mag (Stern et al. 2012).

We find a median [O III] equivalent width of 15.1Å for the stellar-dominated objects and 39.4Å for the AGN-dominated objects, implying that the objects with a higher AGN accretion rate also have a higher average contribution from AGN light in the mid-infrared. As can be seen from Fig. 13, the distribution of [O III] equivalent widths in the LARGESS sample spans a broad continuum rather than having a bimodal distribution. It is therefore not surprising to see a broad diversity in the optical and mid-IR colours of HERGs, reflecting the underlying broad distribution of AGN accretion rates and the differing relative contributions of AGN and galaxy light. Since almost all the HERGs in which the galaxy light dominates have *WISE* colours characteristic of SF galaxies rather than quiescent red galaxies (see e.g. Fig. 16), it seems likely that there is some ongoing star formation in most HERG host galaxies.

10 THE HOST GALAXIES OF HIGH- AND LOW-EXCITATION RADIO AGN

We now look in more detail at the host galaxies of distant radio AGN. Earlier studies (mainly of local galaxies at redshift $z < 0.2$) have found that the host galaxies of HERGs and LERGs differ in their typical stellar mass (e.g. Smolčić 2009; Best & Heckman 2012), star formation rate/history (e.g. Herbert et al. 2010; Hardcastle et al. 2013), environment (e.g. Best 2004; Sabater et al. 2013) and many other properties (Smolčić 2009; Best & Heckman 2012; Janssen et al. 2012).

These differences cannot be accommodated within a single AGN unification model where the observed properties depend only on the orientation of a dusty torus. Instead, it is now generally accepted that a dichotomy exists between HERGs and LERGs, in which the key parameter is the accretion rate of gas on to the supermassive black hole (Hardcastle et al. 2007; Heckman & Best 2014). Since a dusty torus is not present in most LERGs, their observed optical properties are not expected to be orientation dependent. In contrast, HERG host galaxies are expected to show orientation-dependent properties and we discuss this briefly in the next section.

10.1 AGN unification and the host galaxies of HERG and AeB objects

In AGN unification models (Antonucci 1993; Urry & Padovani 1995), we would expect the strong narrow-line radio AGN in our sample (HERGs) to have the same underlying host-galaxy population as the broad-line radio AGN (AeB) and differ only in orientation. In this picture, the HERG optical spectra lack broad Balmer emission wings because the line of sight to the central broad-line region is obscured by a dusty torus. The optical–UV radiation from the accretion disc is absorbed by this torus and re-emitted at longer wavelengths (Congdon & Stein 1989). In this scenario, the accretion disc continuum will not be seen at optical wavelengths in HERGs. Instead, light from the host galaxy of HERGs is expected to dominate the emission at optical wavelengths – with the spectra of AeB objects having an additional optical contribution from the AGN accretion disc.

The LARGESS catalogue includes both HERG and AeB objects across a common range in redshift, and Fig. 17 compares the observed $i_{\rm mod}$ magnitudes of these two classes over the redshift range $0 < z < 0.8$. For redshifts out to $z \sim 0.4$, most HERG and AeB objects lie well above the $i_{\rm mod} = 20.5$ mag cutoff of the catalogue and so we can compare the two classes directly. As can be seen from Fig. 17, the magnitude difference is small at $z < 0.1$, but at higher redshifts the AeB objects start to become systematically brighter.

At $z = 0.4$, the AeB objects are typically 0.95 mag brighter than the HERGs in the SDSS *i* band. If both classes are drawn from the same parent galaxy population, this would imply that the AGN contributes around 60 per cent of the *i*-band light in the AeB objects at $z \sim 0.4$. At higher redshifts than this, a direct comparison with the HERGs becomes more difficult because lower luminosity HERGs will be excluded by the optical magnitude cut of our catalogue, but it is clear that a class of very optically bright ($i_{\rm mod} < 18$ mag) AeB objects starts to appear at $z > 0.5$.

The relative fractions of HERGs and AeB objects are similar across the whole redshift range sampled by our survey ($0.1 < z < 0.8$), with 20–30 per cent of the combined (HERG+AeB) population being AeB objects and the remaining 70–80 per cent HERGs. This is broadly consistent with unified models in which HERGs

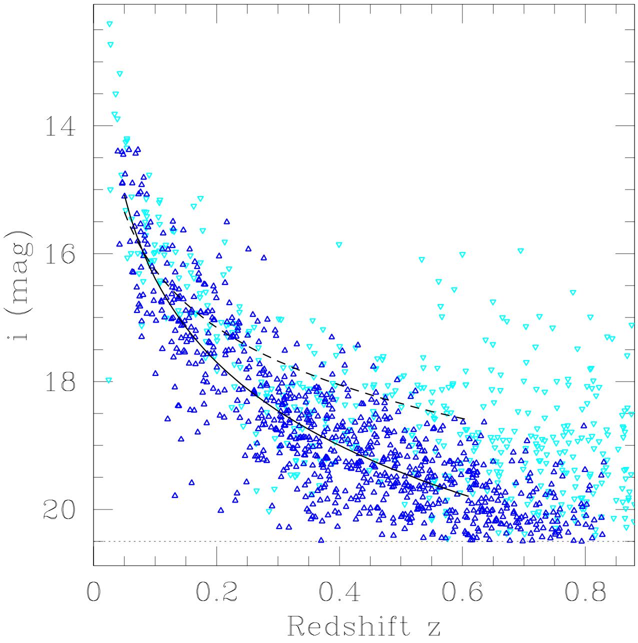


Figure 17. A comparison of the $i_{\rm mod}$ apparent magnitudes of the host galaxies of HERGs (dark blue points) and broad-line radio AGN (spectral class AeB; cyan points). The solid and dashed lines are least-squares fits to the HERG and AeB data points, respectively, and the horizontal dotted line shows the optical limit of the LARGESS catalogue at $i_{\rm mod} = 20.5$ mag.

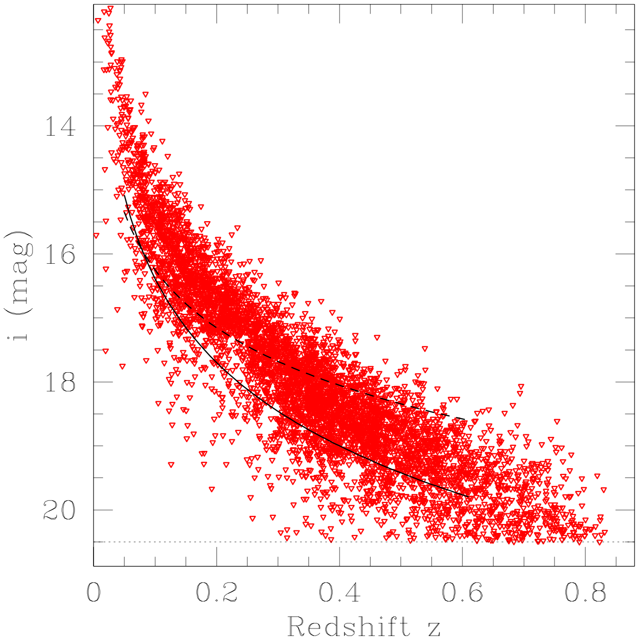


Figure 18. The $i_{\rm mod}$ apparent magnitudes of the host galaxies of LERGs (red points). For comparison with the HERG population, the solid and dashed lines are the same least-squares fits to the HERG and AeB data points shown in Fig. 17. The horizontal dotted line again shows the optical limit of the LARGESS catalogue at $i_{\rm mod} = 20.5$ mag.

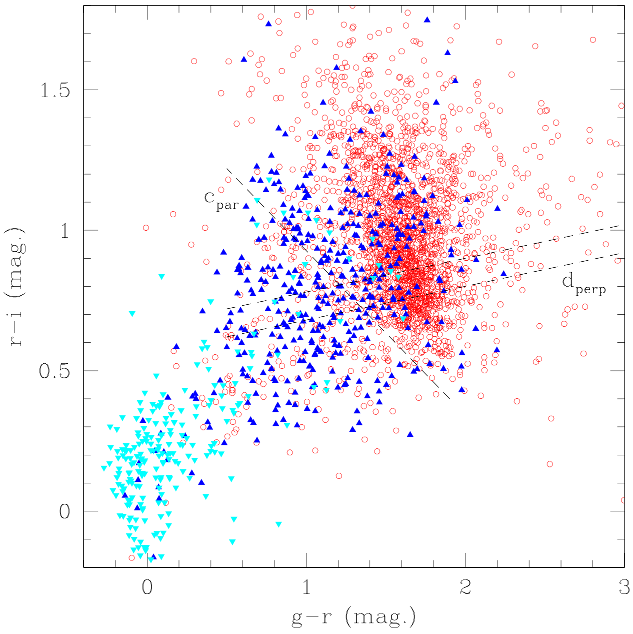


Figure 19. Optical two-colour plot for radio AGN in the LARGESS sample with spectroscopic redshifts in the range $0.4 < z < 0.8$. As in earlier plots, LERGs are shown by red dots, HERGs by dark blue triangles and AeB objects by cyan triangles. The two diagonal lines marked $C_{\rm par}$ and $d_{\rm perp}$ are the limits used by Cannon et al. (2006) (based on earlier work by Eisenstein et al. 2001) to select LRGs in this redshift range.

and AeB radio sources are members of the same underlying galaxy population seen at different orientations.

From a comparison of Figs 17 and 18, we can see that LERG host galaxies are significantly brighter (by ∼0.5 mag on average in *i*) than HERG hosts across the full redshift range, in agreement with the findings of earlier studies (e.g. Smolčić 2009; Best & Heckman 2012).

10.2 What is missed by colour-selected luminous red galaxy samples at redshift $0.4 < z < 0.8$?

We can now answer one of the questions raised in the introduction to this paper – did earlier, colour-selected studies of radio AGN in the distant Universe (Sadler et al. 2007; Donoso et al. 2009) miss a significant population of 'blue' radio galaxies?

The 2SLAQ Luminous Red Galaxy (LRG) survey (Cannon et al. 2006) was designed for cosmological studies of large-scale structure at intermediate redshifts ($0.4 < z < 0.8$), and photometrically pre-selected to contain luminous ($>2L_*$) galaxies with red colours characteristic of an old stellar population. Massive red galaxies of this kind are the hosts of almost all radio AGN in the local Universe (e.g. Best et al. 2005a; Mauch & Sadler 2007), and Sadler et al. (2007) used spectroscopy from the 2SLAQ LRG survey to compile one of the largest and most uniform spectroscopic samples of radio AGN beyond the local Universe – providing strong evidence for the cosmic evolution of low-power radio galaxies. Sadler et al. (2007) did however note that the rate of cosmic evolution measured for low-power radio galaxies in their study was a lower limit, since any additional population of blue radio galaxies at redshift $z > 0.4$ might be removed by the 2SLAQ LRG colour selection and so missed from their analysis.

Fig. 19 shows an optical two-colour diagram for LARGESS radio AGN in the redshift range $0.4 < z < 0.8$. The colour cuts used by Cannon et al. (2006) to select the luminous red galaxies (LRGs) used as the basis of the Sadler et al. (2007) radio sample are shown by the dashed lines marked $C_{\rm par}$ and $d_{\rm perp}$. The first of these ($C_{\rm par}$) divides SF galaxies from galaxies with old, passively evolving

Table 12. Photometric properties of LARGESS HERG and LERG host galaxies at redshift $0.4 < z < 0.8$.

Class	All	$C_{\rm par} \geq 1.6$ ('passive')	$C_{\rm par} < 1.6$ ('star-forming')
LERG	2236	2021	215
HERG	413	203	210
Total	2649	2224	425
HERG fraction	16 per cent	9 per cent	49 per cent

stellar populations (Eisenstein et al. 2001), while the two lines marked $d_{\rm perp}$ select out galaxies above a fixed limit in stellar mass.

There are several points to note in Fig. 19. As expected (see e.g. Fig. 14), the LARGESS AeB (QSO) objects (cyan points) have very blue colours and fall well to the left of the $C_{\rm par}$ line. Most of the LERGs in the LARGESS sample (red points in Fig. 19) fall above the $C_{\rm par}$ and $D_{\rm perp}$ lines, and so would have been included in the 2SLAQ LRG sample. However, the LARGESS sample has a fainter optical cutoff ($i < 20.5$ mag) than the 2SLAQ LRG sample ($i < 19.8$ mag), and so the LARGESS LERG distribution extends to lower mass galaxies that fall below the 2SLAQ $\mathrm{D_{perp}}$ lines.

In contrast to the LERGs, only about half the HERGs in the LARGESS sample (dark blue points) lie to the right of the $C_{\rm par}$ line. It is clear that while some HERGS at $0.4 < z < 0.8$ are hosted by luminous red galaxies (3–4 per cent of the Sadler et al. (2007) LRGs showed high-excitation [Ne III] and [Nev] lines in their optical spectra), a significant number are blue enough to be excluded by the 2SLAQ LRG colour cut.

As can be seen from Table 12, over 90 per cent of LERGs at $0.4 < z < 0.8$ satisfy the 2SLAQ $C_{\rm par}$ colour cut, but more than half of the HERG objects would be excluded. The overall completeness of the sample is only mildly affected (only 16 per cent of all radio AGN in this redshift range would be excluded by the $C_{\rm par}$ colour cut), so the Sadler et al. (2007) and Donoso et al. (2009) measurements remain a reasonable guide to the AGN radio luminosity function in the distant Universe. The effect is, however, much more dramatic for the HERG population, and it is clear that a substantial fraction of the HERG population (∼50 per cent) was missed by the 2SLAQ LRG colour cut. This underlines the importance of using a colour-unbiased sample to measure the HERG radio luminosity function (as distinct from the overall luminosity function for radio AGN) beyond the local Universe.

10.3 A matched sample of HERG and LERG host galaxies

We now turn to a comparison of the host-galaxy properties of HERGs and LERGs across the redshift range spanned by the LARGESS sample. To do this, we need to take into account the likelihood that the host galaxies of these two classes span different ranges in stellar mass.

In the following analysis, we will therefore compare the properties of samples of LERGs and HERGs that have been carefully matched in stellar mass, redshift and radio luminosity. We will also split these samples into two coarse redshift bins at $0.01 < z \leq 0.3$ and $z > 0.3$. The lower redshift bin is well matched to the median redshift of the sample used in Best & Heckman (2012), while the higher redshift bin allows us to see whether there is any evidence for redshift evolution in the properties of LERG and HERG host galaxies. The galaxy stellar masses used in this analysis were derived from rest-frame $(g - i)$ colours and i-band luminosities using the relationship defined by Taylor et al. (2011).

Table 13. Table of KS-test probability of rejecting the null hypothesis that the LERG- and HERG-matched sample are the same for each parameter. The values are the median from the 100 realizations.

Parameter	$0.01 < z \leq 0.3$	$z > 0.3$
(i) Parameters used to match the samples		
Redshift (z)	<1 per cent	<1 per cent
Stellar mass ($M_\star$)	69 per cent	12 per cent
Radio luminosity ($L_{\rm NVSS}$)	<1 per cent	6 per cent
(ii) Parameters not used for matching		
Rest-frame $(g_{\rm mod} - i_{\rm mod})_0$ colour	>99 per cent	>99 per cent
Size (R_{50})	41 per cent	86 per cent
Concentration index (C)	>99 per cent	>99 per cent

The LERGs in our sample greatly outnumber the HERGs, so for each HERG we identified three LERGs matched in stellar mass ($|\Delta \log M_\star| < 0.1$ dex), redshift ($|\Delta z| < 0.01$) and radio luminosity ($|\Delta \log L_{\rm NVSS}| < 0.25$ dex) to act as control galaxies for the corresponding HERG. Once an LERG has been assigned as a control galaxy, it is removed from subsequent matches so that the final list of controls contain a unique list of LERGs. HERGs are not used if we are not able to find three control LERGs. These limits ensure that we have a fair comparison between the properties of LERGs and HERGs, without being biased by other known differences that may influence the properties that we are interested in. We repeated the matching process to create 100 realizations so that our results are more robust. This approach is similar to the one used by Best & Heckman (2012).

Table 13 shows the median KS-test probability from the 100 realizations. These are the probability of rejecting the null hypothesis that the distribution of a property is the same between the HERG and LERG matched samples. Table 13 confirms that there is no significant difference between the HERG and LERG samples for the three parameters (stellar mass, redshift and radio luminosity) used to match them.

10.3.1 Comparing LERG and HERG host galaxies of similar stellar mass: optical properties

Fig. 20 shows the distributions of redshift, radio luminosity, stellar mass, $(g_{\rm mod} - i_{\rm mod})_0$ rest-frame colour, linear projected half light radius (R_{50}, calculated using the half-light Petrosian angular size in the r-band $r_{\rm pet,\,50}$ and the angular diameter distance equation) and the concentration index ($C = r_{\rm pet,\,90}/r_{\rm pet,\,50}$, where $r_{\rm pet,\,90}$ is the radius containing 90 per cent of the Petrosian flux), for LERGs (red) and HERGs (blue) in our matched sample. The top six plots are for the low-redshift sample and the lower six plots for the higher redshift sample.

For the three parameters that were not matched (colour, concentration index and linear size), we find that LERGs have slightly redder $(g - i)_0$ colour and a slightly higher concentration index than the HERGs. Although the differences are small, they are significant at the >99 per cent confidence level, and are similar (∼0.1 mag in rest-frame $g - i$ colour) in both redshift bins. We interpret these differences as evidence that LERGs are dominated by an older stellar population, and hosted by a higher fraction of elliptical galaxies (in the morphological sense), than HERGs of similar stellar mass. The linear sizes of HERG and LERG host galaxies were not significantly different in either redshift bin. The results shown in Fig. 20 imply that since the colour difference between the host galaxies of HERGs and LERGs persists when we compare galaxies that are matched in stellar mass, it is unlikely to be caused by metallicity

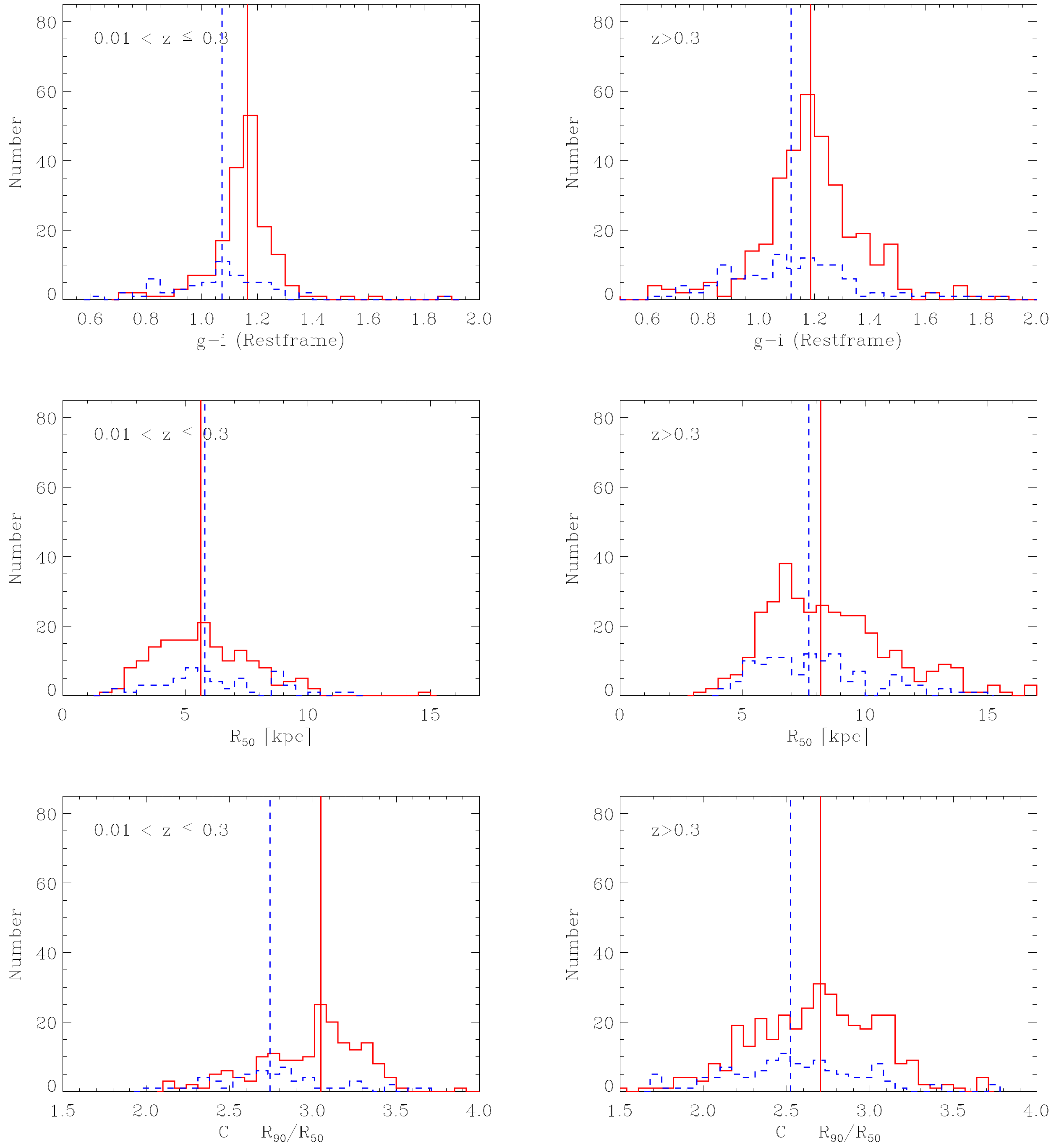


Figure 20. Comparison of properties for matched samples of LERGs (red solid line) and HERGs (blue dashed line). The samples have been selected to have the same distribution in redshift, radio luminosity and stellar mass. From top to bottom, the plots show: (i) the rest-frame $(g_{\rm mod} - i_{\rm mod})_0$ colour, (ii) the radius containing 50 per cent of the galaxy light (R_{50}; in physical units) and (iii) the concentration index ($C = R_{90}/R_{50}$, where R_{90} is the radius containing 90 per cent of the galaxy light). The left-hand plots show the results for objects with redshift $0.01 < z \leq 0.3$, and the right-hand column row is for $z > 0.3$. Vertical lines in each plot indicate the median values of the distributions.

differences and so must arise from differences in star formation rate (or possibly the presence of non-stellar continuum light in some HERG hosts).

A colour difference between HERG and LERG host galaxies has also been seen in earlier studies. Smolčić et al. (2009) found a dichotomy in many of the optical properties of weak and powerful radio AGN both in the local Universe and out to redshifts as high as $z \sim 1.3$, while Best & Heckman (2012) (but see also Janssen et al. 2012) found similar results for sample of LERGs and HERGs with $z < 0.3$.

In the picture developed from earlier X-ray studies Allen et al. (2006); Evans et al. (2006); Hardcastle, Evans & Croston (2006, 2007) the central AGN in HERGs and LERGs accrete gas at different rates, with HERGs being fuelled at relatively high rates by cold gas (some of which may be available for associated star formation) while the LERGs are fuelled at a much lower rate through

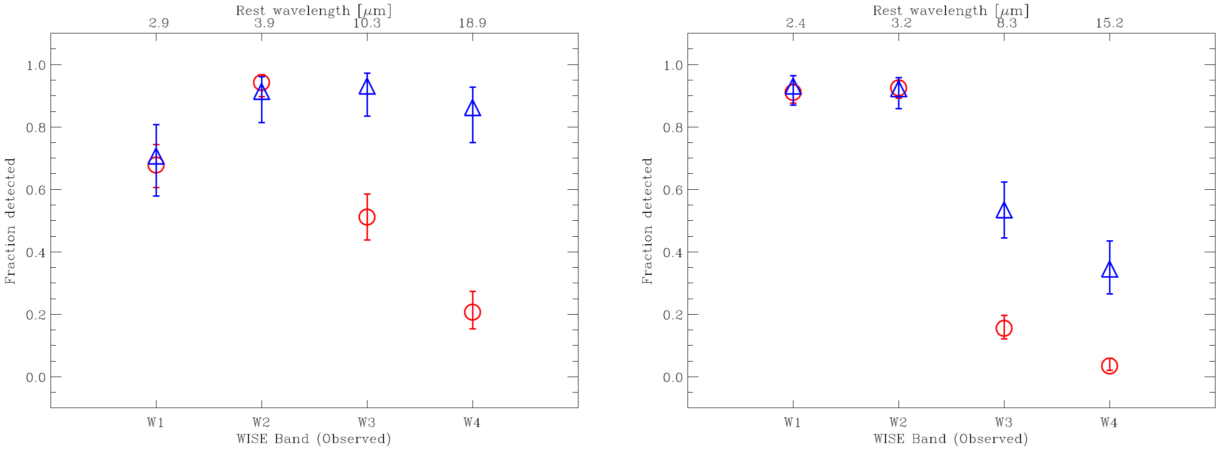


Figure 21. Fraction detected in each *WISE* band for HERGs (blue triangles) and LERGs (red circles) that are matched in redshift, radio luminosity and stellar mass. Left-hand panel is for those with $0.01 < z \leq 0.3$ and right-hand panel are for those with $z > 0.3$.

radiatively inefficient accretion flows - most likely from a surrounding hot gas halo. In the latter case, 'radio-mode feedback' (Croton et al. 2006), in which mechanical energy from the radio jet is energetically coupled to the surrounding hot gas, operates in the LERG galaxies to keep the star formation rate low. Also, because elliptical galaxies are typically found in environments (sufficiently massive haloes) where cold-mode accretion is less likely to occur, the higher fraction of elliptical for LERGs again supports the idea that they are accreting in the hot mode.

Our results are in broad agreement with this picture, though we note that in some HERGs the bluer colours may be due at least in part to the presence of a non-stellar AGN continuum rather than ongoing star formation, as discussed earlier in Section 9.3.

10.3.2 *Comparing LERG and HERG host galaxies of similar stellar mass:* WISE *mid-IR data*

Fig. 21 shows the fraction of HERGs (blue) and LERGs (red) from the matched sample with a reliable detection in each *WISE* band. As in Section 10.3.1, the two samples were matched in stellar mass, redshift and radio luminosity. The equivalent rest-frame wavelengths for each band, calculated from the median redshift, are also shown on the top axis. Once again these plots are for a single realization of our Monte Carlo simulations. The errors are estimated assuming a Beta distribution adopting a 95 per cent confidence interval (Cameron 2011).

There are large differences in the fraction detected in the *W*3 and *W*4 bands in both redshift bins, in contrast to the detection rates in the *W*1 and *W*2 bands. We use a *Z*-test to quantify the significance between two fractions. Since the errors are asymmetric, we assume the pooled error is the quadrature sum of the lower error in the larger fraction and the upper error in the smaller fraction. The null hypothesis is that the two fractions are the same (i.e. a two-tailed test). The median values of the *Z*-test from the 100 iterations are listed in Table 14.

From the values in Table 14, we conclude that the detection fractions detected in the *W*1 and *W*2 bands are not significantly different for HERGs and LERGs. As mentioned earlier, the emission in the *WISE W*1 and *W*2 bands arises mainly from the old stellar population, so is expected to correlate with galaxy stellar mass (e.g. Hwang et al. 2012; Wen et al. 2013). Since this is one of the parameters used to match the two samples, we would not expect to see a difference in these two bands.

Table 14. The median σ significance when comparing the fraction of LERGs and HERGs that have a reliable *WISE* detection in different bands.

	σ significance	
WISE band	$0.01 < z \leq 0.3$	$z > 0.3$
*W*1	0.2	0.3
*W*2	0.5	0.1
*W*3	3.3	3.9
*W*4	5.0	3.8

When we turn to the *W*3 and *W*4 bands, we find that HERGs are significantly ($\sim 3\sigma$ in *W*3 and $\sim 4\sigma$ in *W*4) more likely to have a reliable detection than LERGs in both bands. This difference is seen in both redshift bins.

The mid-infrared emission in the *WISE W*3 and *W*4 bands is most likely reprocessed emission from a starburst (*W*3 band in particular) and/or an optical-AGN (*W*4 band in particular) (e.g. Donoso et al. 2012; Hwang et al. 2012). The fact that we see differences in both *W*3 and *W*4 detection rates is consistent with the presence of a radiatively efficient accretion disc and dusty torus in HERGs, which is absent in LERGs. Hardcastle et al. (2006) found similar results in their study of X-ray properties in LERGs and HERGs. They found that HERGs have a nuclear X-ray absorption component, which was rarely observed in the LERGs, and noted that the absence of such a component in LERGs is consistent with these objects lacking a radiatively efficient accretion disc.

11 SUMMARY

We have compiled the Large Area Radio Galaxy Evolution Spectroscopic Survey (LARGESS) catalogue, a new, large data set designed to study the evolution and environments of high- and low-excitation radio galaxies (HERGs and LERGs) out to $z \sim 0.8$. The catalogue contains 19 179 radio sources detected by the FIRST survey that have an SDSS optical counterpart with $i_{\rm mod} < 20.5$. We have obtained new good-quality optical spectra for over 5000 LARGESS objects, and adding these to lower redshift archival data means that 10 856 LARGESS objects currently have a reliable optical redshift. The spectroscopic completeness varies across the survey area, and is highest (∼90 per cent) in the three equatorial GAMA fields. The median redshift of radio AGN in the LARGESS catalogue is

$z \sim 0.44$, and the bright end of the radio luminosity function ($P_{1.4} > 10^{24}$ W Hz^{-1}) is well sampled for all the main radio AGN populations over the redshift range $0 < z < 0.8$. The catalogue also contains a minority population of nearby ($z < 0.3$) SF galaxies.

The great majority of LARGESS radio sources are compact (only 10 per cent have complex or multi-component radio structures on scales larger than a few arcsec), which makes the optical matching process straightforward in most cases. We estimate that our final matched sample is 95 per cent complete and has 94 per cent reliability.

The LARGESS sources typically have between 5 per cent and 25 per cent of their 1.4 GHz radio emission in a diffuse component that is seen by the NVSS survey but missed by the higher resolution FIRST images. This needs to be taken into account in any subsequent analysis. In particular, as discussed in Section 4.3, any analysis requiring accurate flux densities for extended radio sources should impose a $S_{\mathrm{tot}}^{\mathrm{FIRST}} \geq 3.5$ mJy limit for the LARGESS catalogue.

For objects with a reliable redshift from the SDSS, GAMA and WiggleZ surveys, we made both a visual and an automated classification of the spectrum. These two classification schemes were combined to provide a single classification, and allow us to distinguish between (i) Galactic stars, (ii) radio galaxies with weak or no emission lines (LERGs), (iii) galaxies with emission lines or radio emission generated from star formation (SF) and (iv) radio galaxies where both emission line and radio emission arise from the AGN (broad emission lines, AeB and narrow emission lines, HERGs).

While the optical counterparts of most radio AGN are massive red galaxies at all redshifts out to $z = 0.8$, we find that at least 15 per cent of radio AGN at $0.4 < z < 0.8$ are hosted by blue galaxies that appear likely to have ongoing star formation. It is notable that at least half of the HERG population in this redshift range are ‘blue radio galaxies’ rather than passively evolving red galaxies. Roughly 8 per cent of the radio AGN population in the same redshift range are radio-loud QSOs with broad Balmer emission lines in their optical spectra.

We find that LERGs, HERGs and AeB (QSO) objects have very different broad-band spectral-energy distributions (SEDs). The optical and mid-IR properties of LERGs are generally similar to those of luminous red galaxies dominated by an old stellar population, while the AeB (QSO) objects are consistent with a power-law SED. In contrast, the HERGs are a much more heterogeneous population in terms of both their optical and mid-IR colours.

We used a matched sample of HERGs and LERGs to compare their physical properties, and found that that HERGs have bluer colours (probably indicative of a younger stellar population) compared to LERGs of similar mass, redshift and radio luminosity. The concentration index is also higher for LERGs than HERGs, suggesting that LERGs are hosted by a higher fraction of elliptical galaxies than the HERGs. Both these results are independent of redshift. The matched sample of LERGs and HERGs show no significant difference in the linear sizes in any redshift bin. These results confirm results of previous studies, but also extend them into a higher redshift range.

By comparing the fraction with a reliable detection in various *WISE* mid-infrared bands, we found that HERGs are significantly more likely to be detected in the *W*3 and *W*4 bands than a matched sample of LERGs, both locally and at higher redshifts. From this, we inferred the presence of a radiatively efficient accretion disc and dusty torus in the HERGs, which is absent in the LERGs. This is consistent with theories that suggest a dichotomy in the accretion mode of HERGs and LERGs.

ACKNOWLEDGEMENTS

JHYC would like to acknowledge the funding provided by the University of Sydney via an Australian Postgraduate Award, the Postgraduate Research Scholarship Scheme and the William & Catherine McIlrath Scholarship. JHYC would also like to acknowledge the support provided by the Astronomical Society of Australia.

EMS and SMC acknowledge the financial support of the Australian Research Council through Discovery Project grants DP1093086 and DP 130103198. SMC acknowledges the support of an Australian Research Council Future Fellowship (FT100100457). We thank Philip Best and Dick Hunstead for helpful comments.

GAMA is a joint European–Australasian project based around a spectroscopic campaign using the Anglo–Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including *GALEX* MIS, VST KIDS, VISTA VIKING, *WISE*, *Herschel*-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is http://www.gama-survey.org/.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS website is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium 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, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, 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.

This publication makes use of data products from the *Wide-field Infrared Survey Explorer*, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

This research has made use of data obtained from or software provided by the US National Virtual Observatory, which is sponsored by the National Science Foundation.

REFERENCES

Adelman-McCarthy J. K. et al., 2008, ApJS, 175, 297
Allen S. W., Dunn R. J. H., Fabian A. C., Taylor G. B., Reynolds C. S., 2006, MNRAS, 372, 21
Antonucci R., 1993, ARA&A, 31, 473
Assef R. J. et al., 2010, ApJ, 713, 970
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5 (BPT)
Bardelli S. et al., 2010, A&A, 511, A1
Becker R. H., White R. L., Helfand D. J., 1995, ApJ, 450, 559
Best P. N., 2004, MNRAS, 351, 70

Best P. N., Heckman T. M., 2012, MNRAS, 421, 1569

Best P. N., Kauffmann G., Heckman T. M., Ivezić Ž., 2005a, MNRAS, 362, 9

Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić Ž., White S. D. M., 2005b, MNRAS, 362, 25

Best P. N., Ker L. M., Simpson C., Rigby E. E., Sabater J., 2014, MNRAS, 445, 955

Blanton M. R., Roweis S., 2007, AJ, 133, 734

Bridge C. R. et al., 2013, ApJ, 769, 91

Bruzual G. A., Charlot S., 1993, ApJ, 405, 538

Budavári T., Dobos L., Szalay A. S., Greene G., Gray J., Rots A. H., 2007, in Shaw R. A., Hill F., Bell D. J., eds, ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 559

Cameron E., 2011, PASA, 28, 128

Cannon R. et al., 2006, MNRAS, 372, 425

Colless M. et al., 2001, MNRAS, 328, 1039

Condon J. J., 1992, ARA&A, 30, 575

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

Congdon C., Stein W. A., 1989, PASP, 101, 691

Cress C. M., Helfand D. J., Becker R. H., Gregg M. D., White R. L., 1996, ApJ, 473, 7

Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, MNRAS, 349, 1397

Croom S. M. et al., 2009, MNRAS, 392, 19

Croton D. J. et al., 2006, MNRAS, 365, 11

Cushing M. C. et al., 2011, ApJ, 743, 50

Cutri R. M. et al., 2013, Technical report, Explanatory Supplement to the AllWISE Data Release Products. Available at: http://wise2.ipac.caltech.edu/docs/release/neowise/expsup/

Donoso E., Best P. N., Kauffmann G., 2009, MNRAS, 392, 617

Donoso E. et al., 2012, ApJ, 748, 80

Drinkwater M. J. et al., 2010, MNRAS, 401, 1429

Driver S. P. et al., 2011, MNRAS, 413, 971

Dunlop J. S., Peacock J. A., 1990, MNRAS, 247, 19

Edelson R., Malkan M., 2012, ApJ, 751, 52

Eisenstein D. J. et al., 2001, AJ, 122, 2267

Ellis S. C., Bland-Hawthorn J., 2007, MNRAS, 377, 815

Evans D. A., Worrall D. M., Hardcastle M. J., Kraft R. P., Birkinshaw M., 2006, ApJ, 642, 96

Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748

Gendre M. A., Best P. N., Wall J. V., Ker L. M., 2013, MNRAS, 430, 3086

Hardcastle M. J., Evans D. A., Croston J. H., 2006, MNRAS, 370, 1893

Hardcastle M. J., Evans D. A., Croston J. H., 2007, MNRAS, 376, 1849

Hardcastle M. J. et al., 2013, MNRAS, 429, 2407

Heckman T. M., Best P. N., 2014, ARA&A, 52, 589

Heckman T. M., Ptak A., Hornschemeier A., Kauffmann G., 2005, ApJ, 634, 161

Helfand D. J., White R. L., Becker R. H., 2015, ApJ, 801, 26

Herbert P. D., Jarvis M. J., Willott C. J., McLure R. J., Mitchell E., Rawlings S., Hill G. J., Dunlop J. S., 2010, MNRAS, 406, 1841

Hopkins A. M. et al., 2003, ApJ, 599, 971

Hopkins A. M. et al., 2013, MNRAS, 430, 2047

Hwang H. S., Geller M. J., Kurtz M. J., Dell'Antonio I. P., Fabricant D. G., 2012, ApJ, 758, 25

Ivezić Ž. et al., 2002, AJ, 124, 2364

Janssen R. M. J., Röttgering H. J. A., Best P. N., Brinchmann J., 2012, A&A, 541, A62

Jarrett T. H. et al., 2011, ApJ, 735, 112

Johnston H. M., Sadler E. M., Cannon R., Croom S. M., Ross N. P., Schneider D. P., 2008, MNRAS, 384, 692

Kauffmann G. et al., 2003, MNRAS, 346, 1055

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121

Kimball A. E., Ivezić Ž., 2008, AJ, 136, 684

Kirkpatrick J. D. et al., 2011, ApJS, 197, 19

Laing R. A., Jenkins C. R., Wall J. V., Unger S. W., 1994, in Bicknell G. V., Dopita M. A., Quinn P. J., eds, ASP Conf. Ser. Vol. 54, The Physics of Active Galaxies. Astron. Soc. Pac., San Francisco, p. 201

Lake S. E., Wright E. L., Petty S., Assef R. J., Jarrett T. H., Stanford S. A., Stern D., Tsai C.-W., 2012, AJ, 143, 7

Liske J. et al., 2015, MNRAS, 452, 2087

Longair M. S., 1966, MNRAS, 133, 421

Longair M. S., Seldner M., 1979, MNRAS, 189, 433

Loveday J. et al., 2012, MNRAS, 420, 1239

Magliocchetti M., Maddox S. J., Lahav O., Wall J. V., 1998, MNRAS, 300, 257

Mateos S. et al., 2012, MNRAS, 426, 3271

Mauch T., Sadler E. M., 2007, MNRAS, 375, 931

Oort M. J. A., 1987, PhD thesis, University of Leiden

Peterson B. M., 1997, An Introduction to Active Galactic Nuclei. Cambridge Univ. Press, Cambridge

Pracy M. B. et al., 2016, MNRAS, 460, 2

Rowlands K. et al., 2012, MNRAS, 419, 2545

Sabater J., Best P. N., Argudo-Fernández M., 2013, MNRAS, 430, 638

Sadler E. M., Gerhard O. E., 1985, MNRAS, 214, 177

Sadler E. M. et al., 2002, MNRAS, 329, 227

Sadler E. M. et al., 2007, MNRAS, 381, 211

Sarzi M. et al., 2006, MNRAS, 366, 1151

Saunders W., Cannon R., Sutherland W., 2004, Anglo-Aust. Obs. Epping Newsl., 106, 16

Schneider D. P. et al., 2005, AJ, 130, 367

Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103

Simpson C. et al., 2012, MNRAS, 421, 3060

Smolčić V., 2009, ApJ, 699, L43

Smolčić V. et al., 2009, ApJ, 696, 24

Stern D. et al., 2012, ApJ, 753, 30

Stoughton C. et al., 2002, AJ, 123, 485

Taylor E. N. et al., 2011, MNRAS, 418, 1587

Tremonti C. A. et al., 2004, ApJ, 613, 898

Trouille L., Barger A. J., 2010, ApJ, 722, 212

Urry C. M., Padovani P., 1995, PASP, 107, 803

van de Voort F., Schaye J., Booth C. M., Haas M. R., Dalla Vecchia C., 2011, MNRAS, 414, 2458

Wake D. A. et al., 2006, MNRAS, 372, 537

Wen X.-Q., Wu H., Zhu Y.-N., Lam M. I., Wu C.-J., Wicker J., Zhao Y.-H., 2013, MNRAS, 433, 2946

Wilman D. J. et al., 2008, ApJ, 680, 1009

Wright E. L. et al., 2010, AJ, 140, 1868

Xilouris E. M., Madden S. C., Galliano F., Vigroux L., Sauvage M., 2004, A&A, 416, 41

York D. G. et al., 2000, AJ, 120, 1579

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table 10. A list of 20 sample objects from the main LARGESS data table.
(http://www.mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stw2396/-/DC1).

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