A&A 622, A12 (2019)
https://doi.org/10.1051/0004-6361/201833893

Astronomy & Astrophysics
Special issue

LOFAR Surveys: a new window on the Universe

Radio-loud AGN in the first LoTSS data release

The lifetimes and environmental impact of jet-driven sources

M. J. Hardcastle[1], W. L. Williams[1], P. N. Best[2], J. H. Croston[3], K. J. Duncan[4], H. J. A. Röttgering[4], J. Sabater[2], T. W. Shimwell[5], C. Tasse[6,7], J. R. Callingham[5], R. K. Cochrane[2], F. de Gasperin[8], G. Gürkan[9], M. J. Jarvis[10,11], V. Mahatma[1], G. K. Miley[4], B. Mingo[3], S. Mooney[12], L. K. Morabito[10], S. P. O'Sullivan[13], I. Prandoni[14], A. Shulevski[15], and D. J. B. Smith[1]

[1] Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
e-mail: m.j.hardcastle@herts.ac.uk
[2] SUPA, Institute for Astronomy, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[3] School of Physical Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
[4] Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
[5] ASTRON, The Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA Dwingeloo, The Netherlands
[6] GEPI & USN, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France
[7] Department of Physics & Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa
[8] Hamburger Sternwarte, Gojenbergsweg 112, 21029 Hamburg, Germany
[9] CSIRO Astronomy and Space Science, PO Box 1130, Bentley WA 6102, Australia
[10] Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
[11] Physics and Astronomy Department, University of the Western Cape, Bellville 7535, South Africa
[12] School of Physics, University College Dublin, Belfield, Dublin 4, Ireland
[13] Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany
[14] INAF – Istituto di Radioastronomia, Via P. Gobetti 101, 40129 Bologna, Italy
[15] Anton Pannekoek Institute for Astronomy, University of Amsterdam, Postbus 94249, 1090 GE Amsterdam, The Netherlands

Received 18 July 2018 / Accepted 22 September 2018

ABSTRACT

We constructed a sample of 23 344 radio-loud active galactic nuclei (RLAGN) from the catalogue derived from the LOFAR Two-Metre Sky Survey (LoTSS) survey of the HETDEX Spring field. Although separating AGN from star-forming galaxies remains challenging, the combination of spectroscopic and photometric techniques we used gives us one of the largest available samples of candidate RLAGN. We used the sample, combined with recently developed analytical models, to investigate the lifetime distribution of RLAGN. We show that large or giant powerful RLAGN are probably the old tail of the general RLAGN population, but that the low-luminosity RLAGN candidates in our sample, many of which have sizes <100 kpc, either require a very different lifetime distribution or have different jet physics from the more powerful objects. We then used analytical models to develop a method of estimating jet kinetic powers for our candidate objects and constructed a jet kinetic luminosity function based on these estimates. These values can be compared to observational quantities, such as the integrated radiative luminosity of groups and clusters, and to the predictions from models of RLAGN feedback in galaxy formation and evolution. In particular, we show that RLAGN in the local Universe are able to supply all the energy required per comoving unit volume to counterbalance X-ray radiative losses from groups and clusters and thus prevent the hot gas from cooling. Our computation of the kinetic luminosity density of local RLAGN is in good agreement with other recent observational estimates and with models of galaxy formation.

Key words. galaxies: jets – galaxies: active – radio continuum: galaxies

1. Introduction

Radio-loud active galactic nuclei (radio galaxies and radio-loud quasars; hereafter RLAGN) are a subset of the active galaxy population in which accretion onto the central supermassive black hole of a galaxy generates a relativistic jet of charged particles (electrons, positrons, and/or protons) and magnetic field. These jets propagate into the medium permeating and surrounding the host galaxy, inflating "bubbles" of low-density, high-pressure material containing relativistic electrons that generate the observed radio emission through the synchrotron process. Basic models of the dynamics of these objects as they interact with the external medium have been available for over 40 years (Scheuer 1974; Blandford and Rees 1974), but have been refined and improved more recently both in terms of analytical models (e.g. Kaiser & Alexander 1997; Blundell et al. 1999; Luo & Sadler 2010; Turner & Shabala 2015; Hardcastle 2018) and numerical models taking account of the known environmental properties of these objects (e.g. Reynolds et al. 2002; Basson & Alexander 2003; Zanni et al. 2003; Krause 2005; Heinz et al. 2006; Mendygral et al. 2012; Hardcastle & Krause 2013, 2014; English et al. 2016).

Radio galaxy physics has become important outside the active galactic nucleus (AGN) community over the past 20 years for two closely related reasons. The first is the role of AGN in solving the so-called cooling flow problem. This problem was

posed by observations of rich clusters of galaxies that showed that their central hot gas, emitting in the X-ray with temperatures $T \sim 10^7$ K, had cooling times ($\tau = E/(\mathrm{d}E/\mathrm{d}t)$) much less than the age of the Universe. This gas should therefore cool out of the temperature regime in which it emits X-rays and, eventually, form stars or deposit cold gas in the central cluster galaxy at a rate, for the most rapidly cooling clusters, of thousands of solar masses per year, while causing the gas to flow inwards owing to the loss of central pressure (a "cooling flow"; Fabian et al. 1984). However, these large amounts of cold gas and/or star formation were not observed, and neither, when observational advances permitted it, was the low-temperature X-ray-emitting gas that would have been predicted by the cooling flow model (e.g. Sakelliou et al. 2002). It was rapidly realised (e.g. Eilek & Owen 2006) that essentially all cooling flow clusters host a RLAGN with sufficient power to offset the cooling, and so it is now widely assumed that radio galaxies provide the "thermostat" for rich clusters of galaxies, keeping the central gas hot and rarefied. The precise mechanism by which the gas is coupled to the active nucleus, and the radio lobes to the gas, is not clear. It seems likely that at least some radio galaxies are powered by accretion of the hot phase onto the black hole (e.g. Allen et al. 2006; Hardcastle et al. 2007), although increasingly the consensus is that this is mediated by a cooling instability (Pizzolato & Soker 2005; Gaspari et al. 2013; Voit & Donahue 2015). Hot-gas accretion thus provides the connection in one direction, while the expansion of the radio lobes can do work on the hot gas in various ways (see e.g. Fabian et al. 2000 for early imaging, and McNamara & Nulsen 2012; Heckman & Best 2014 for recent reviews).

The second reason for the importance of RLAGN in recent times arises in part out of the first. A major advance in our understanding of the way all galaxies formed and evolved has come from efforts to use numerical models to predict features of the present and past galaxy population, such as the galaxy mass or luminosity function, the galaxy colour-magnitude diagram or the evolution of star formation in the Universe. Initially this work used semi-analytic models, i.e. the properties of the baryonic matter in the Universe were inferred from a hydrodynamical simulation of the dark matter (e.g. Bower et al. 2006; Croton et al. 2006). With increasing computing power, it is now possible to model the baryons and dark matter together and in a self-consistent way (e.g. Vogelsberger et al. 2014; Schaye et al. 2015) and semi-analytic modelling has also become more sophisticated (Croton et al. 2016). However, all these models agree in predicting a very different galaxy luminosity function from what is observed, if only the physics of dark matter, gas, and stars is taken into account; far too many luminous galaxies are produced, and the most luminous galaxies in the simulations are an order of magnitude more luminous than anything we observe today. Motivated in part by the observational evidence that RLAGN indeed solve the cooling flow problem and prevent the formation of massive cluster-centre galaxies in the local Universe, modellers can reproduce the observed galaxy luminosity function by introducing AGN feedback into their models. In modern models, this takes the form of an injection of energy into the baryonic matter driven by accretion onto the galactic-centre black hole. This AGN feedback takes place not just in the local Universe, but over all cosmic time, and, as it is a crucial ingredient of all modern models of galaxy formation, it is vital that the nature and energetics of the feedback predicted be tested against observations.

Cosmological models that deal with a scale large enough to reproduce the galaxy luminosity function do not simultaneously deal with the scales at which detailed AGN physics can be modelled. Even if they did, we still lack a basic understanding of what causes some AGN to have powerful radio jets. Therefore models do not predict the relative importance of radio-loud and radio-quiet AGN in heating the baryons and inhibiting star formation: the oft-quoted division by Croton et al. (2006) into "jet-mode" and "quasar-mode" AGN does not imply that radiative feedback is known observationally to terminate star formation in major mergers. There are several reasons to think that RLAGN may be important, however. Firstly, we know (as discussed above) that RLAGN, not radio-quiet ones, are responsible for the maintenance of hot cluster haloes in the local Universe: few if any of these host a luminous quasar but effectively all host a powerful radio galaxy. Secondly, RLAGN have a clear mechanism, the interaction between the jets and the external medium, for efficiently coupling the AGN output (in the form of the kinetic power of the jets) to the baryonic matter, and this is directly observed to drive hot and cold gas out of galaxies (see e.g. Morganti et al. 2005; Nesvadba et al. 2008; Hardcastle et al. 2012; Russell et al. 2017). On the other hand, radio-quiet AGN, which produce all of their energetic output as photons, can only drive outflows in dusty galaxies where the radiation from the accretion disc is efficiently absorbed before it can escape from the galaxy, meaning that, for example, almost all optically selected quasars cannot be efficiently optically coupled to their host galaxies. The true answer to the question of which AGN are implicated in feedback processes can only be provided by observation. In order to understand the contribution of RLAGN to these processes, we need the ability to measure the kinetic power, and thus the kinetic luminosity function, of large, well-constrained samples of RLAGN.

At present, although significant advances have been made in recent years, this is still a difficult undertaking even in the local Universe. Two approaches to measuring the jet power from the radio luminosity are commonly used. Firstly, analytic models of the source can be used to predict the radio luminosity for a given jet power (e.g. Willott et al. 1999). Secondly, estimates of the jet kinetic power can be derived from X-ray observations that show cavities in the hot gas inflated by the radio lobes are used to infer the $p\Delta V$ work done to inflate the cavity, which can be combined with some estimate of the source age to infer the jet power; these cavity powers can then be empirically related to the radio luminosity (e.g. Bîrzan et al. 2004; Cavagnolo et al. 2010). Both methods have significant problems. The cavity power method relies on a poorly known source age and can only work when cavities are observed, which rules out the use of this approach in the case of the most powerful classical double AGN, in which typically the lobes are brighter in inverse Compton than their surroundings (see Hardcastle & Croston 2010 for a discussion of why this is so). This method is, moreover, biased towards small sources in rich cluster environments (Bîrzan et al. 2012) and relies on expensive X-ray observations that are not available for large samples of sources. Therefore there is at least some possibility that the relationships that are derived from cavity estimates are biased for the population in general. It is not even clear whether the correlations between radio luminosity and cavity power that are observed in these samples are driven by physics rather than a common correlation with distance (Godfrey & Shabala 2016). On the other hand, a single conversion based on a theoretical model giving the radio luminosity is also unrealistic, since it is clear on simple physical grounds that the radio luminosity must depend strongly on the source age (since the luminosity depends on the energy density in the lobes, the lobe volume and the magnetic field strength in the lobes, all

of which evolve with time), on environment, and on redshift (due to inverse-Compton losses). Both numerical and analytical models of this evolution exist, at least for certain types of RLAGN (e.g. Kaiser & Alexander 1997; Blundell et al. 1999; Mocz et al. 2011; Turner & Shabala 2015; Hardcastle 2018), but they have generally not been applied to large numbers of sources in a consistent way to infer jet powers.

To measure kinetic powers in the local Universe, large sky areas are needed, but large-area statistical studies of RLAGN have been hindered in the past by the capabilities of previous-generation radio instruments. Existing very wide-area radio surveys that have had a resolution high enough to allow adequate identification of RLAGN with their host galaxy or quasar have not simultaneously had the range of short baselines necessary for high-fidelity imaging of extended structures. To date the highest resolution wide-area radio survey is the Very Large Array (VLA) survey Faint Images of the Radio Sky at Twenty-Centimetres (FIRST), with a resolution of 5 arcsec (Becker et al. 1995). As this is insensitive to structures on scales larger than around 1 arcmin, however, it is not possible to generate a complete sample from FIRST alone, and in the past it has been necessary to combine catalogues from the NRAO VLA Sky Survey (NVSS; Condon et al. 1998) and FIRST to achieve this (e.g. Best et al. 2005; Hardcastle et al. 2012; Best & Heckman 2012). With this approach, though it is possible to obtain flux densities and optical identifications for radio sources, it is not possible (without a great deal of work on the archival FIRST and NVSS uv data) to make fully spatially sampled high-resolution images of them; this means that insufficient information about, for example, source size, a proxy of age, is available for jet power inference.

The LOw Frequency ARray (LOFAR; van Haarlem et al. 2013) is in the process of solving this problem. The LOFAR survey of the northern sky, the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al. 2017), when complete, will provide an unrivalled resource for wide-area low-frequency (144 MHz) selection of extragalactic samples, both of star-forming galaxies (hereafter SFG) and of RLAGN[1]. At optimal declinations for LOFAR LoTSS is approximately ten times deeper than FIRST for typical observed spectral indices ($\alpha \sim 0.7$), while having a similar resolution (6 arcsec) and, crucially, possessing the short baselines necessary to image all but the largest scale structures in the radio sky[2]. Low-frequency selection for RLAGN is extremely valuable because it minimizes the effect on the total flux density of flat-spectrum beamed structures such as the core, jets, and hotspots: at low frequencies emission from a RLAGN is dominated by the much more isotropic large-scale lobes. Thus, although the forthcoming Evolutionary Map of the Universe (EMU) survey (Norris et al. 2011) with the Australian Square Kilometre Array Precursor (ASKAP) will cover a larger sky area at comparable (slightly lower) resolution to LoTSS and very similar sensitivity to typical sources at its operating frequency of 1.3 GHz, LoTSS as a low-frequency survey will remain competitive until the (currently hypothetical) long-baseline extension of the low-frequency Square Kilometer Array (SKA) itself.

The present paper is concerned with the properties of RLAGN selected from the LoTSS survey of the *Hobby-Eberly* Telescope Dark Energy eXperiment (HETDEX; Hill et al. 2008) Spring field (hereafter the HETDEX survey; Shimwell et al. 2019), the first full-quality data release of LoTSS (DR1). We investigate what can be learned about RLAGN physical properties, and in particular their effect on their environments, from the LOFAR-detected RLAGN population without spectroscopic information other than what is provided by the Sloan Digital Sky Survey (SDSS; Eisenstein et al. 2011). We begin by constructing an RLAGN sample based on the spectroscopic data where available and on photometric redshifts and WISE colours otherwise. This allows us to construct a very large sample of objects with radio luminosity and (projected) physical size information. We then show that a simple model of the RLAGN lifetime function, essential input into an inference of jet power from radio observations, adequately explains the observed distribution of source sizes for luminous sources. Furthermore, there is no evidence for any difference in host galaxy properties as a function of physical size, which is consistent with a simple model in which the powerful radio galaxies are a single physical population observed at different times in their life cycle. This conclusion allows us to carry out bulk jet power inference using a dynamical model of radio source evolution and to construct a jet kinetic luminosity function in the local Universe whose integral can be compared to the current radiative output of groups and clusters. Throughout this paper we use a cosmology in which $H_0 = 70\,\mathrm{km\,s^{-1}}$, $\Omega_\mathrm{m} = 0.3$ and $\Omega_\Lambda = 0.7$. The spectral index α is defined in the sense $S \propto \nu^{-\alpha}$.

2. The data

2.1. Radio data used in this paper

This paper is based on DR1 of the LoTSS survey, which covers 424 deg^2, i.e. about 2% of the total planned northern sky coverage. As described by Shimwell et al. (2019), we have devised an observation and imaging strategy for this area that permits high-fidelity imaging over wide areas down to a typical rms noise level of 70 μJy beam^{-1} at the full 6 arcsec resolution of the Dutch LOFAR baselines[3]. Williams et al. (2019) describe the processing of the raw catalogues derived from the Python Blob Detector and Source Finder (PyBDSF) software (Mohan & Rafferty 2015) to give a sample of 318 520 radio sources that are believed to be real (i.e. not artefacts from the limited dynamic range of the survey) and physical (i.e. lobes of radio galaxies are associated and unassociated sources are de-blended). These authors also describe the combination of the radio images and catalogues with the available optical and near- to mid-infrared data from PanSTARRS DR1 (Chambers et al. 2016) and All WISE (Wright et al. 2010; Mainzer et al. 2011), a process that gives plausible optical/IR counterparts for 72% of these objects (231 716). The vast majority of these sources are derived from likelihood-ratio cross-matching with a combined optical/IR catalogue (for simplicity we refer to these as optical counterparts in what follows). Finally Duncan et al. (2019) describe the algorithms used to estimate photometric redshifts for these optical counterparts; 162 249 sources (51% of the input catalogue and 70% of those objects with optical identifications; IDs) have some kind of redshift estimate, using spectroscopic redshifts where available (principally from the SDSS; Eisenstein et al. 2011) and photometric redshifts otherwise.

[1] See http://lofar-surveys.org/

[2] In the imaging that supports this paper we use a short-baseline cut of 100 m, allowing good imaging of structures on scales up to ∼1°. In practice, we are limited in imaging such structures by surface brightness sensitivity rather than short baselines.

[3] The component of the International LOFAR Telescope (ILT) located in the Netherlands has a maximum baseline of 120 km. The observations of the HETDEX field did not include the longer baselines of LOFAR, using telescopes in international partner countries, although they are generally present in other data for the LoTSS survey.

Table 1. Samples considered in this paper.

Name	Description	Number of objects
Full	Complete sample of Williams et al. (2019)	318 520
FC	Flux-complete, flux cut at 0.5 mJy	239 845
O	Optical ID exists	231 716
FCO	Intersection of FC and O	172 898
Z	Some redshift estimate exists	162 249
ZG	"Good" photometric redshift exists	89 671
FCOZG	Intersection of FCO and ZG	71 955
FCOZGM	Cross-match of FCOZG with the MPA-JHU sample	12 803
FCOZGM RLAGN	RLAGN selected from FCOZGM	3706
FCOZGM SFG	SFG selected from FCOZGM	9097
RLAGN	RLAGN selected from FCOZG	23 344
SFG	SFG selected from FCOZG	41 998

Our starting point in this paper is the 318 520 sources in the "value-added" radio and optical catalogue of Williams et al. (2019). We describe this as the value-added catalogue because it contains optical, infrared, and redshift information that is not present in the raw radio catalogues. Shimwell et al. (2019) and Williams et al. (2019) describe the measurement of the radio properties of these objects, but it is worth briefly summarising these properties here. In essence, objects in the sample fall into two categories: objects for which we adopt the PyBDSF properties of an original radio detection, and objects in which a number of original PyBDSF sources have been amalgamated (or, in very rare cases, where one PyBDSF source has been split into components) after human visual inspection. In the former case (simple sources), the flux density is the result of a Gaussian fit or fits to the image data by PyBDSF, and we adopt a largest angular size for the source that is twice the full width at half maximum (FWHM) of the deconvolved fitted Gaussian (this is roughly correct for uniform-brightness projected spherical or ellipsoidal sources). In the latter case (composite sources) the total flux density of the resulting source is taken to be the sum of the total flux densities of all the components used, and the largest angular size is taken to be the maximum distance across the convex hull enclosing the elliptical regions with semi-major and semi-minor axes corresponding to the deconvolved major and minor axes (FWHM) of the fitted Gaussians. This definition has the property that it would be consistent with the simple-source definition if there were only one Gaussian in the composite source. In order to permit the convex hull to be calculated, unresolved sources that are part of a composite object are given a very small size (0.1 arcsec).

Our definition of composite source size differs from that of (Hardcastle et al. 2016; hereafter H16), which was the previous largest area AGN survey with LOFAR. In their work, H16 used the maximum pairwise distance between the centres of all components of a composite source. However, visual inspection of sources from H16 established that, while summing the flux densities of composite components gives results that are consistent with flux-density measurements from hand-drawn regions, the H16 size definition tends to systematically underestimate true source sizes. Our present definition is likely to be closer to the truth than that of H16 in many cases and is good enough for the purposes of the present paper. More computationally complex size definitions will be discussed in other papers.

Shimwell et al. (2019) give as the criterion for deciding whether a simple source is genuinely resolved a relationship between peak and integrated flux density: a source is unresolved if

$$\frac{S_{\rm int}}{S_{\rm peak}} > 1.25 + 3.1\left(\frac{S_{\rm peak}}{\rm RMS}\right)^{-0.53}, \quad (1)$$

where RMS is the local RMS noise level and the coefficients of the relationship are best-fitting parameters of an envelope that encompasses 95% of the apparently compact LoTSS-DR1 sources, checked by comparison with the properties of bright FIRST sources. We adopt this definition and apply it to both simple and composite sources, with two additional criteria: we say that sources are always resolved if they are composite sources with two or more components, and that they are never resolved if they are less than 1 arcsec in size (this catches composite sources with one bright unresolved component). By these criteria, there are 38 230 resolved sources and 280 290 unresolved sources in the catalogue.

Throughout the rest of the paper we refer to the LOFAR observing frequency as 150 MHz, for ease of comparison with the many other surveys that have used this observing frequency. The small difference between 150 MHz and the true central frequency of around 144 MHz at the pointing centre has no effect on the scientific interpretation of the data. The effective frequency varies slightly across the field because in a given LOFAR observation the lower frequencies, corresponding to a larger station primary beam, contribute more to the image at large off-axis distances. Given the mosaicing strategy described by Shimwell et al. (2019), this gives rise to only a small effect on the data.

2.2. Catalogues

We generated catalogues for further study by imposing cuts on the complete catalogue of 318 520 sources, which generates new samples. For reference, a list of all the samples considered in this work is tabulated in Table 1.

The first point to consider is the flux completeness of the survey. Shimwell et al. (2019) show that the survey is better than 99% complete for point sources having flux densitities greater than 0.5 mJy at 150 MHz. As PyBDSF selects sources above a 5σ detection threshold and the worst rms noise levels in the mosaiced images are around 100 μJy beam^{-1}, this number seems reasonable and adopting it is equivalent to adopting a uniform noise floor across the survey. If we apply a flux density cut at 0.5 mJy the total number of sources is reduced to 239 845 (hereafter the FC sample), with a very similar optical identification

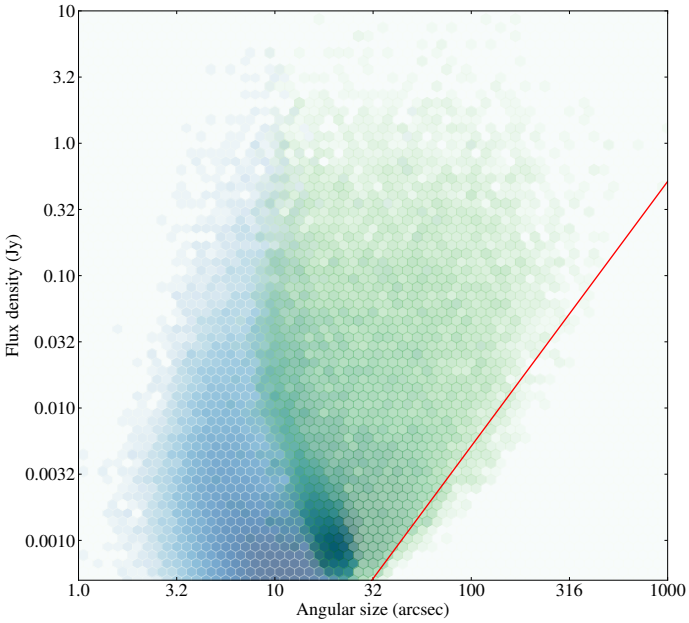

Fig. 1. Total flux density of sources above the point-source completeness cut as a function of their total angular size. The density plot shows the distribution of the FCO sample sources with measured angular sizes; blue sources are unresolved and green sources resolved. The red line shows an empirically normalised line of $S \propto \theta^2$ as expected for a surface-brightness limited sample. There are 280 290 unresolved and 38 230 resolved sources; the two colour scales are adjusted to make both populations visible.

and redshift fraction; basing our analysis on this sample ensures that we can make unbiased statements about the fractions of sources as a function of measured flux density or some associated quantity. It is important to remember that the survey is not complete for resolved sources at this level, however, as it is surface-brightness limited rather than flux limited. The fact that surface-brightness limitations affect us is clearly seen in Fig. 1, where we plot the total flux density of sources above the point-source completeness cut as a function of their total angular size. A boundary to the right of this plot imposed by surface-brightness limitations is visible and shows approximately the expected slope. However, we can also see that we are sensitive to at least some comparatively large sources even at the lowest flux densities. This is a consequence of the fact that many sources are not uniform in surface brightness. The dependence of this observational limit not just on the average surface brightness but on its distribution is an insuperable problem for this sort of survey (in absence of a much more sensitive survey from which we can estimate the incompleteness) and its effects must be borne in mind in what follows.

The next set of criteria to be applied is on optical identifications and redshifts. For any study of the physical nature of these sources we need an optical identification, so at this point we have to restrict ourselves to the 231 716 sources with an optical counterpart in the WISE or PanSTARRS data (sample O). The nature of the remaining objects cannot be determined at this point; a large fraction of these sources are expected to be high-redshift galaxies, but they will also include low-redshift objects where the optical identification is ambiguous or the radio structure is not clear enough to permit an ID. Further investigation of this population is important but is beyond the scope of this paper. If we restrict ourselves to sources that are both in FC and in O, we obtain 172 898 sources (sample FCO).

We also require a redshift, and so we needed to make a decision on the quality of photometric redshifts that we were prepared to accept. In total 162 249 sources have either a spectroscopic redshift or some photometric redshift estimate (sample Z). The smaller size of Z compared to O is essentially because an optical detection, which gives us matched photometry across all of the optical bands, is required to derive a photometric redshift and a large number of the detections in O are in WISE only. The errors on some of the redshift estimates are large. We chose to use $\Delta z/(1+z)$ as our figure of merit for photometric redshifts, where $\Delta z = (z_{1,\mathrm{max}} - z_{1,\mathrm{min}})/2$ is the half-width of the 80% credible interval defined by Duncan et al. (2019), and is therefore slightly larger than the 1σ error of standard error analysis. "Good" photometric redshifts then have $\Delta z/(1+z)$ less than some threshold value[4]. For example, 89 671 sources (sample ZG) have either a spectroscopic redshift or a photometric redshift with $\Delta z/(1+z) < 0.1$. The relative numbers of sources with different redshift quality as a function of optical brightness are shown in Fig. 2, which also shows the effect on the outliers of applying this cut on Δz. Generally the effect is to reduce the number of sources with grossly discrepant redshifts, although a small number of sources remain (in the bottom right of Fig. 2) with photometric redshifts much less than their spectroscopic redshifts. These objects are all high-redshift quasars and are discussed by Duncan et al. (2019); other quasars are well fitted by the photometric redshift code and the issues that affect these particular objects include very bright broad emission lines that affect the optical spectral energy distribution (SED), or lines of sight with particularly low absorption due to intervening inter-galactic medium (and hence weak Lyman break features). Because bright quasars are very likely to be selected as such by SDSS spectroscopy, it seems unlikely that they represent a significant contaminating population at low redshift. There are a total of 71 955 sources in the flux-complete catalogue that also have an optical ID and a good redshift. From this population (FCOZG) we can start to select samples of RLAGN.

3. AGN selection

The separation of RLAGN from SFG is one of the biggest problems faced by this and all other current-generation extragalactic radio surveys in which SFG are present in significant numbers (i.e. any survey, like LoTSS, with the equivalent of a sub-mJy flux limit at 150 MHz). Emission due to the stellar population is always going to be present at least in cases in which we do not have the ability to separate this emission spatially from RLAGN activity, which requires resolution substantially better than the spatial scales of the galactic disc. Therefore a perfect RLAGN selection would involve selecting as AGN all those, and only those, galaxies whose radio emission significantly exceeds the level expected from star formation or other stellar processes (Hardcastle et al. 2016; Calistro Rivera et al. 2017; Smolčić et al. 2017). It should be noted that this is significantly different from other AGN selection methods and produces a different population. Many radiatively efficient AGN, selected as such using X-ray emission, SED fitting (e.g. Calistro Rivera et al. 2016) or traditional emission-line classifications, appear to lie on the star-forming main sequence, perhaps with no significant radio emission that is not due to star

[4] We note that the error estimates do not take into account some systematic effects. For example, contamination of the photometry by emission lines has a complex, redshift-dependent effect that it is difficult to model and remove.

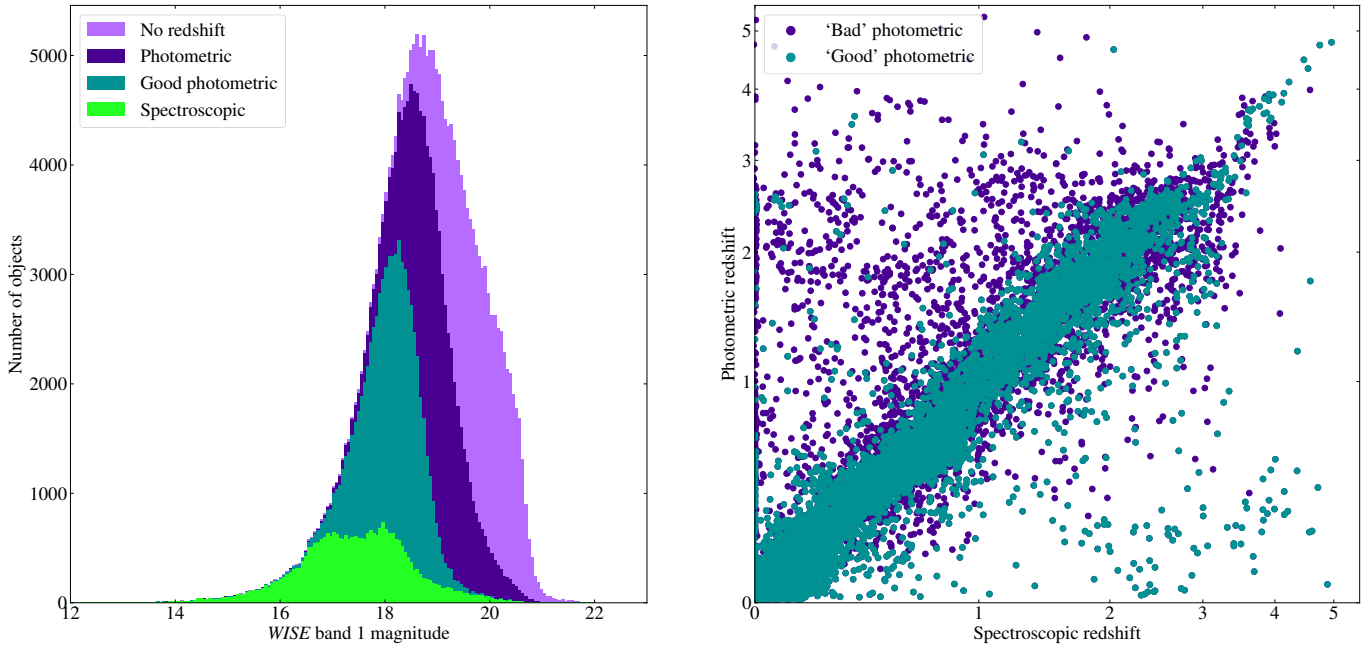

Fig. 2. *Left panel*: histogram of the distribution of WISE band-1 AB magnitudes for optically identified objects in the sample (sample O), colour-coded by the quality of available redshifts (spectroscopic, good photometric with $\Delta z/(1+z) < 0.1$, any photometric, or none). The plot shows 218 600 sources with WISE detections. *Right panel*: photometric vs. spectroscopic redshift for sources where both are available, showing the distribution of all photometric redshifts and of the good sample.

formation (Mingo et al. 2016; Gürkan et al. 2018, 2019), while many RLAGN have little radiative nuclear output and would not easily be selected as AGN in any band other than the radio. Similarly, AGN selections using mid-infrared colour/colour criteria (e.g. Assef et al. 2010; Jarrett et al. 2011; Stern et al. 2012; Mateos et al. 2012; Secrest et al. 2015) cleanly select sources dominated by very luminous (quasar-like) AGN; these selections, however, have been shown to under-represent the radio-loud population, as they are biased against lower luminosity and higher redshift AGN, both of which are preferred hosts for radio sources (e.g. Gürkan et al. 2014; Rovilos et al. 2014; Mingo et al. 2016).

There are two problems in practice with selection based on the expected level of emission from stellar processes. Firstly, the relationship between radio emission and star formation is still poorly understood. It may depend not just on star formation but on a number of galaxy parameters (Gürkan et al. 2018) and, because of the complex chain of physical processes and timescales connecting low-frequency radio emission to star formation, it certainly has a good deal of irreducible, intrinsic scatter that will always act to blur the distinction between strong star formation and weak AGN activity. Thus there are physical reasons why there will never be a unique right answer for objects on the SFG/AGN boundary, irrespective of the accuracy of the available star formation rate estimates. Secondly, in our particular case, we do not have good information about the star formation rates of most of the HETDEX host galaxies. Working in the H-ATLAS NGP field, H16 were able to make use of the *Herschel* data to select RLAGN using the radio/far-infrared relation; Gürkan et al. (2018) in the same field expanded this to select radio-excess AGN candidates based on star formation rates inferred from spectral fitting to the broad-band far-infrared through to optical photometry for the (low-z) galaxies in their parent sample, using the MAGPHYS code (da Cunha et al. 2008) in a manner similar to that described by Smith et al. (2012). However, we do not have *Herschel* data for HETDEX and inference of star formation rates from SED fitting is much less robust without it.

One approach is simply to apply a luminosity cut. However, starburst galaxies with star formation rates of $\sim 10^3\ M_\odot\ \mathrm{yr}^{-1}$ would have LOFAR luminosities of $\sim 10^{25}\ \mathrm{W\,Hz^{-1}}$ at 150 MHz according to the radio to star formation rate relation of Gürkan et al. (2018), although it should be noted that this is an extrapolation as such extreme objects do not exist in their sample. A simple cut in luminosity thus needs to be placed at relatively high luminosities to avoid contamination. For example, cutting FCOZG at $10^{25}\ \mathrm{W\,Hz^{-1}}$, which should remove most star-forming objects, leaves 6660 sources – still a large sample but less than a tenth of the parent population. Many low-luminosity RLAGN would be excluded by such a cut.

A subset of the objects in the FCOZG sample (12 803 objects: sample FCOZGM) have emission-line measurements and estimates of host galaxy properties from SDSS, provided by the MPA-JHU catalogue[5]. Data available for the FCOZGM objects include spectroscopic source classifications and estimates of star formation rates using the methods of Brinchmann et al. (2004), which combine emission-line and continuum (4000 Å break) information. For these objects, which are typically at low redshift given the requirement for SDSS spectroscopy, it would in principle be possible to follow Gürkan et al. (2018) and select as RLAGN sources that lie significantly above the locus for SFG in a plot of star formation rate versus radio luminosity. Such a plot (Fig. 3) indeed appears to show a good division between two distinct populations. However, a problem with this is that star formation rates for quiescent galaxies may be underestimated relative to, for example, the SED-fitting results of Gürkan et al. (2018), as we have verified by considering the same plot using the

[5] https://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/

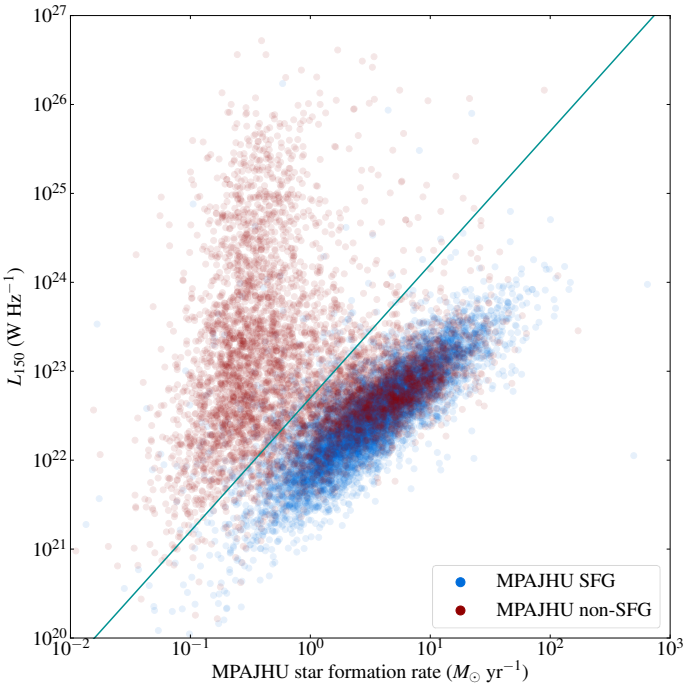

Fig. 3. Radio luminosity as a function of star formation rate for the sources in the FCOZGM sample. Objects are colour-coded according to whether they are classed as star forming in the MPA-JHU catalogue (classifications "STARBURST" or "STARFORMING"); some objects not classified as SFG in this way are clearly nevertheless on the radio to star formation rate relation for star-forming objects. All objects not so classified (including unclassified objects) are placed in the "non-SFG" sample. The line shows a plausible by-eye selection of a division between the two classes.

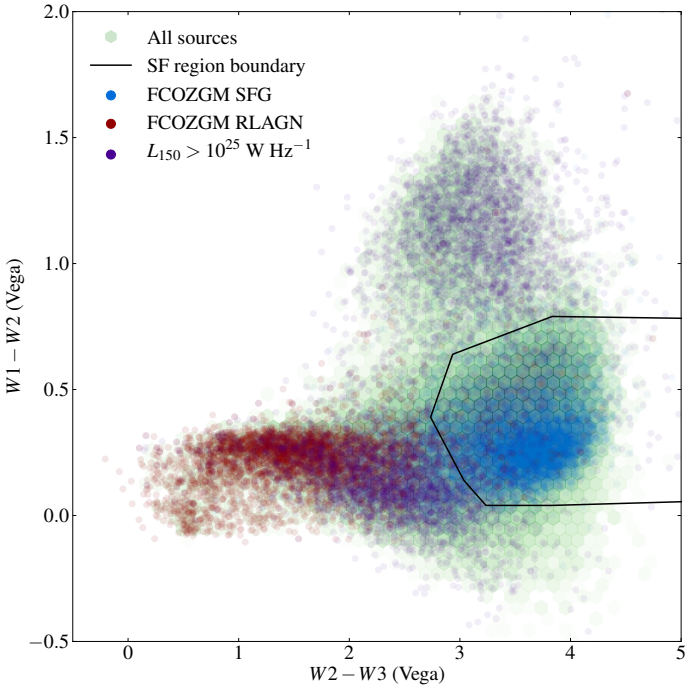

Fig. 4. Observational WISE colour–colour diagram for the FCOZG sample. Overlaid on the green density plot showing the full sample are the locations of FCOZGM RLAGN, FCOZGM SFG, and luminous radio sources. Lines indicate the locus populated by SFG and avoided by RLAGN discussed in the text.

H-ATLAS NGP data. Use of the MPA-JHU star formation rates could artificially accentuate the differences between sources at low star formation rates. We therefore do not use this method directly. Instead, we use the classification scheme developed by Sabater et al. (2019), which builds upon the work of Best et al. (2005), Best & Heckman (2012). In brief, Sabater et al. (2019) consider four different diagnostic diagrams to separate radio AGN from galaxies whose radio emission is primarily powered by star formation. These are (1) the comparison between the 4000 Å break strength and the ratio of radio power per unit stellar mass, developed by Best et al. (2005); (2) the widely used BPT emission line ratio diagnostic diagram (Baldwin et al. 1981; Kauffmann et al. 2003; Kewley et al. 2006); (3) the radio luminosity versus Hα line luminosity; and (4) the *W2–W3* WISE colour (as used by e.g. Wright et al. 2010; Mateos et al. 2012; Gürkan et al. 2014; Herpich et al. 2016). The first and third of these diagnostics are based on the same principle as the use of the radio/far-IR relation: the two parameters are expected to be related for SFGs as they both broadly trace specific star formation rate (diagnostic 1) or star formation rate (diagnostic 3); the RLAGN are identified as those sources offset from this relation due to an additional (jet-related) contribution to the radio luminosity. Diagnostic 2 is well established to separate AGN from SFGs in galaxies with measured emission lines, but fails to distinguish radio-quiet from RLAGN. Diagnostic 4 is less precise, but provides a valuable discriminant where the other diagnostics give contradictory results. Sabater et al. (2019) then combine the results from these four diagnostics to produce an overall AGN/SFG classification, using a comparison with the classifications determined by Gürkan et al. (2018) for the H-ATLAS NGP sample to optimise this combination. Using this classification scheme, 3706 of the FCOZGM sources are classified as being radio-loud AGN, and 9097 are classified as SFGs (where the latter category may include radio-quiet AGN).

This leaves us with the problem of classifying the remaining sources that do not have this spectroscopic information. For this purpose we considered only the WISE data, as WISE data are available for almost all the FCOZG sample (only 2600 sources do not have WISE photometry); the three bands we used, *W1*, *W2*, and *W3*, correspond to 3.4, 4.6, and 12 μm and so sample the rest-frame near- and mid-IR wavelengths for the redshifts of our sample. To plot this diagram in the traditional way we converted the catalogued AB WISE magnitudes for our sources into Vega magnitudes. Figure 4 shows a density plot for the whole FCOZG sample with the classified FCOZGM sources overlaid. As expected, the hosts of FCOZGM objects lie in very different locations depending on their classification as RLAGN or SFG. Moreover, when we add in luminous ($L_{150} > 10^{25}$ W Hz^{-1}) sources, we see that these also tend to avoid a well-defined location in the colour-colour diagram around the location of the FCOZGM SFG. We therefore exclude objects that have WISE colours consistent with the SFG locus, defined as lying in a polygonal region in colour-colour space chosen to give the best separation between SFG and other objects, as shown in Fig. 4.

At high redshifts quasars present a particular problem. Although these are AGN by construction, they need not show any excess radio emission over the expectation from star formation (Mingo et al. 2016). Indeed, Gürkan et al. (2019) argue that the majority of LOFAR-selected quasars have radio emission consistent with star formation if we assume that star formation scales with AGN power as observed at low redshift. If this is the case we should exclude these objects from the RLAGN sample, which we do by making an empirical cut in radio luminosity/absolute magnitude space. We can very easily select quasars

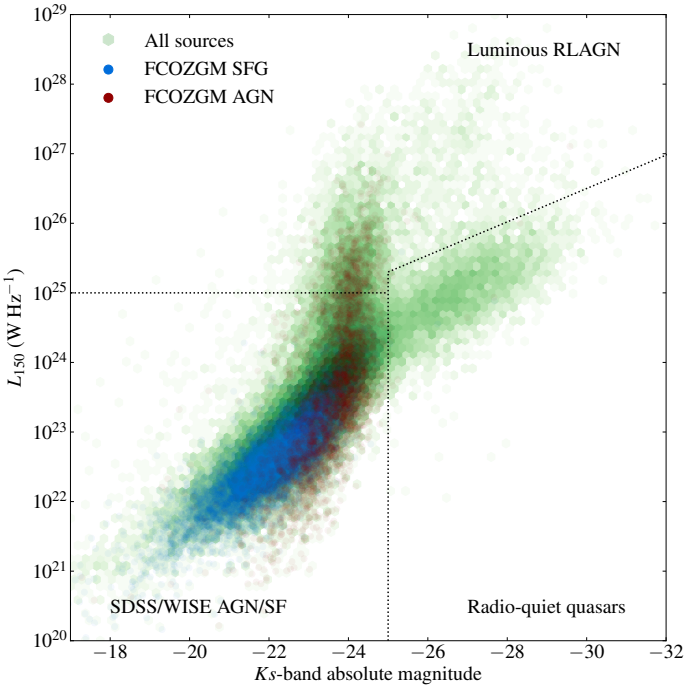

Fig. 5. Radio luminosity vs. K_s-band absolute magnitude for the FCOZG sample. Overlaid on the density plot showing the full sample are the locations of FCOZGM RLAGN and FCOZGM SFG. Lines show the divisions on the plot used to select optically classified or radio-luminous AGN.

by their bright rest-frame magnitudes; anything with K_s-band absolute magnitude <-25 is likely to be a quasar.

The full selection method is as follows. Starting from the FCOZG sample,

1. Sources with K_s-band rest-frame magnitudes outside the range $-33 < K_s < -17$ are removed. This disposes of sources with outlier absolute magnitudes that presumably indicate aberrant redshifts.
2. Sources classed as SFG from FCOZGM are removed.
3. Sources with WISE colours in the SFG locus of Fig. 4, or with no available WISE data, are removed unless either:
   - They are classified in FCOZGM as RLAGN
   - Their luminosity is $>10^{25}$ W Hz^{-1} and their Ks-band magnitude is >-25 (non-quasars), or
   - Their K_s-band rest-frame magnitude is <-25 (quasars), and their radio luminosity is such that $\log_{10}(L_{150}) > 25.3 - 0.06(25 + K_s)$.

The motivation for the cuts used in radio and optical luminosity is illustrated in Fig. 5, which shows a plot of the sample in radio/optical luminosity space indicating the positions of the FCOZGM-classified objects. After these cuts are applied we are left with 23,344 sources, which form our RLAGN sample.

Clearly there are a number of ways in which this selection is not ideal. The WISE colours have errors and so classifications cannot be exact at the boundary. The FCOZGM RLAGN and SF overlap to some extent in the WISE colour space and so we know that it does not provide an exact separation between the populations. We are using apparent colours and therefore the precise boundary between populations should in principle be redshift dependent, but we have not attempted to take this into account in any way. Some high-excitation RLAGN[6] with intermediate nuclear absorption are expected to lie in the SFG colour location, and these are excluded from the RLAGN sample. And, most obviously, we are essentially selecting based on the colour of the host galaxy and not on the radio properties of the source, such that, for example, we cannot select as AGN strongly SFG that also host RLAGN unless their radio luminosity is very high. For all these reasons our RLAGN sample is likely to be neither clean nor complete, but it represents the best sample we are able to construct with the available data given that we lack the data to select radio-excess sources directly. It should be noted that the RLAGN luminosity functions of Sabater et al. (2019) and Williams et al. (in prep.), which use respectively the FCOZGM and the full RLAGN sample, agree well with those of H16, which used a radio-excess selection method. Thus we can be confident that the necessarily more complex selection used in this work is not significantly biasing the RLAGN selection.

Figure 6 illustrates the differences between the 23 344 objects selected as RLAGN from FCOZG (hereafter the "RLAGN sample") and the 41 998 objects selected as SFG on the basis of FCOZGM classifications or WISE colours; the 3460 candidate SF-dominated quasars are excluded from both plots as they are neither RLAGN nor typical SFG. We see that, as expected, RLAGN are generally more luminous and at higher redshift and that resolved SFG have a characteristic size of tens of kpc. A small tail of very large (>100 kpc) SFG must either indicate misclassification, misidentification, incorrect size measurements, or incorrect redshifts and visual inspection of some of these sources shows that all of these factors are involved, and at least some of the SFG show RLAGN-like structures on scales larger than those of the host galaxy. On the whole, however, these plots show that the separation gives the expected behaviour in terms of physical properties of the radio sources.

4. Results and modelling

4.1. Powers and linear sizes of RLAGN

Figure 7 shows the sampling of the luminosity-redshift plane by objects in the RLAGN sample. The sample luminosity spans over nine orders of magnitude due to the wide range in redshift present in the data. However, the high-luminosity objects are dominated by quasars due to the requirement for an optical or WISE detection. Only below a luminosity of 10^{27} W Hz^{-1} do we have large numbers of galaxies, which occupy the space below $z \approx 0.8$. Below a luminosity of around 10^{24} W Hz^{-1} the sample is limited by the radio flux density limit rather than optical detectability in the sense that objects below this radio luminosity cannot be seen at all $z < 0.8$.

The power-linear size plane or $P-D$ diagram for RLAGN (Baldwin 1982; Kaiser et al. 1997; Blundell et al. 1999; Turner et al. 2018) is analogous to the Hertzsprung-Russell diagram for stars or the stellar mass/star formation rate plot for galaxies, in the sense that the location of a source is indicative of both its initial conditions and its evolutionary state and tracks in the diagram can be associated with particular phases of evolution. In the $P-D$ diagram, objects with particular properties describe tracks on the plane that are defined purely by source physics, while remnant sources in which the jets have switched off follow a distinct set of evolutionary tracks (Godfrey et al. 2017; Hardcastle 2018). However, the interpretation of the position of a particular source on the $P-D$

[6] These are objects with radiatively efficient nuclei and thus strong optical emission lines, including quasars and broad- and narrow-line radio galaxies (see Hardcastle et al. 2009 and references therein) and are contrasted with low-excitation radio galaxies (LERG), which have colours and emission-line properties more typical of ordinary ellipticals.

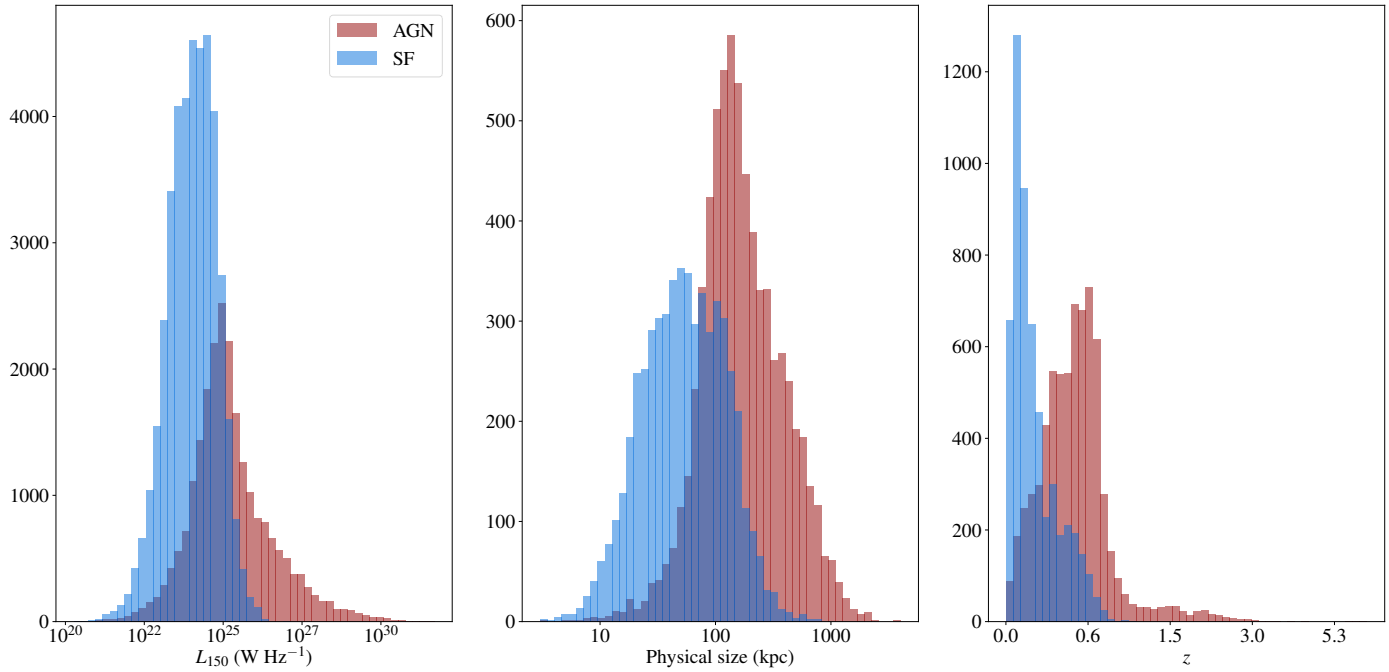

Fig. 6. Distributions of radio luminosity (*left panel*), physical size (*middle panel*), and redshift (*right panel*) for objects selected and not selected as RLAGN from the FCOZG sample using the criteria described in the text. All sources are shown in the left- and right-hand histograms, whereas in the middle only resolved sources (as defined in Sect. 2.1) are plotted.

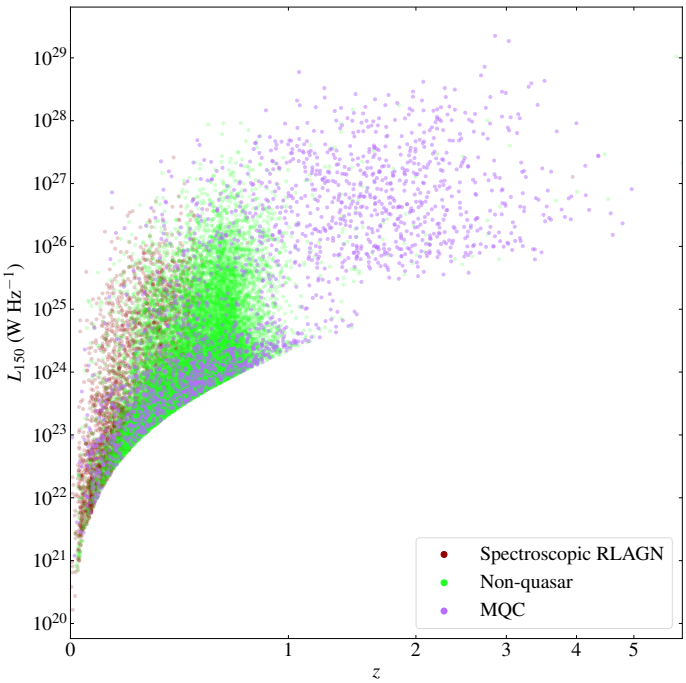

Fig. 7. Sampling of the redshift/luminosity plot by the RLAGN sample. The figure distinguishes between FCOZGM RLAGN ("spectroscopic AGN"), objects classified as quasars in the Million Quasar Catalogue (MQC; http://quasars.org/), which are flagged as such in the value-added catalogue (Duncan et al. 2019) and non-quasar AGN selected using the other criteria discussed in Sect. 3. We note that the x-axis shows $\log(1+z)$, labelled linearly.

diagram is complicated for several reasons. Firstly, the RLAGN environments have a non-negligible effect on their tracks in the $P-D$ diagram (see Hardcastle & Krause 2013, 2014). Secondly, radio galaxies are not spherical, with the effect that the position of real sources on the $P-D$ diagram is dependent on unknown observing factors such as Doppler boosting and the source angle to the line of sight. Thirdly, the theoretical tracks used to interpret the diagram tend to assume that there is a single phase of constant-jet-power evolution followed by a phase of zero-jet-power evolution, whereas we know observationally both that activity of sources can stop abruptly and restart and that optical AGN activity can vary on very short timescales, so that there is no reason to suppose that the jet power Q cannot vary with time on a wide range of timescales. Nevertheless, the $P-D$ diagram remains one of the key tools for interpreting the evolution of populations of RLAGN.

Figure 8 shows this plot for the 23 344 sources of the RLAGN sample, which represent by far the largest sample to have been interpreted in this way at the time of writing, along with the 3CRR sources of Laing et al. (1983) for comparison[7]. For LOFAR sources, resolved and unresolved sources are plotted; for the unresolved sources we take as an upper limit on size the measured deconvolved major axis plus three times the formal error on the major axis; this value is plotted on the density plot for these sources rather than the best estimate of the size (which is zero in many cases). Another feature of the $P-D$ diagram is that it is strongly affected by surface-brightness limitations, as noted by H16. Physically large, low-luminosity (and therefore low-redshift) sources cannot be detected and catalogued even by LOFAR because their surface brightness falls below the detection threshold for our full-resolution imaging. Only for luminosities around 10^{26} W Hz^{-1} and above does this limitation have a negligible effect on the observed size distribution. It can be seen, in spite of this bias, that the LOFAR data span a far wider range in luminosity than the 3CRR sources, while covering much the same range in linear size.

Also overplotted on Fig. 8 are theoretical evolutionary tracks from the models of (Hardcastle 2018; hereafter H18). These,

[7] Data from https://3crr.extragalactic.info/.

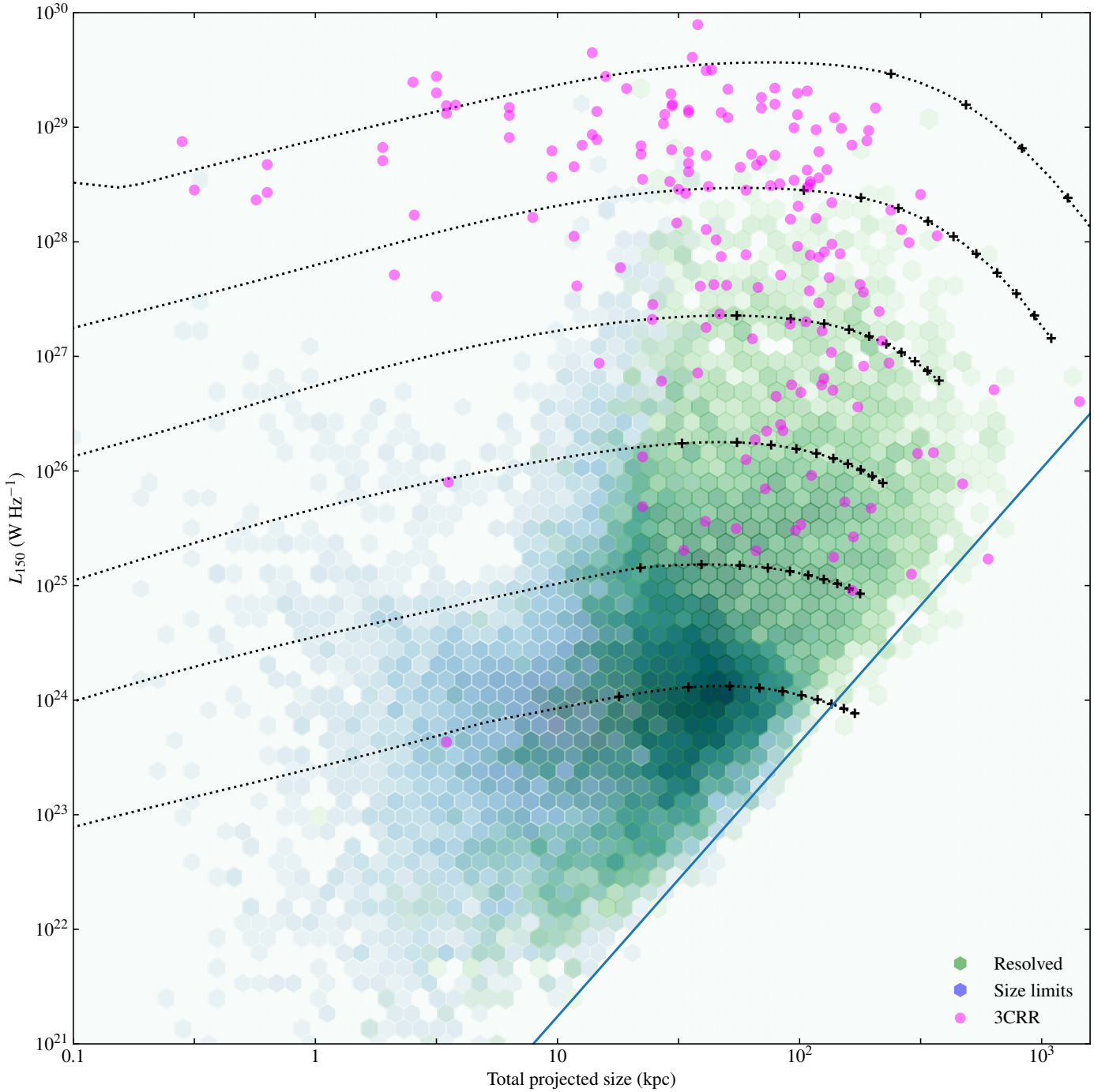

Fig. 8. Power/linear size plane ($P-D$ diagram) for the RLAGN sample. Sources that are resolved as defined in Sect. 2.1 are shown in the green density plot; unresolved sources, where the sizes are upper limits, are in blue; and the 3CRR sample (Laing et al. 1983) is overplotted for comparison. There are a total of 6850 resolved and 16 494 unresolved sources on the plot; the colour scales are adjusted so that both groups can be seen. The diagonal blue line shows (qualitatively) the area of the plot in which surface brightness limitations become important, following the analysis of H16. Overplotted are theoretical tracks for $z = 0$ sources lying in the plane of the sky in a group environment ($M_{500} = 2.5 \times 10^{13}\,M_\odot$, $kT = 1$ keV) for two-sided jet powers (from bottom to top) $Q = 10^{35}, 10^{36}, \ldots, 10^{40}$ W; see the text for details. Crosses on the tracks are plotted at intervals of 50 Myr, where linear size increases monotonically with time; each track lasts for 500 Myr in total.

in common with a number of other models in the literature discussed in Sect. 1 are derived from a model that predicts the time evolution of both luminosity and physical size in a given environment and for a given jet power Q (defined as the two-sided power, i.e. the total kinetic power of both jets). To simplify the plot we use a single environment, a group with $M_{500} = 2.5 \times 10^{13}\,M_\odot$ (corresponding to an X-ray gas temperature of ∼1.0 keV), and evolve sources with jet powers $Q = 10^{35}, 10^{36}, \ldots, 10^{40}$ W for a lifetime of 500 Myr assuming $z = 0$; the choice of redshift affects the radiative losses due to inverse-Compton emission. Looking just at the normalisation of the tracks, we can see that the powerful 3CRR sources in these models correspond to jet powers $\gtrsim 10^{39}$ W, while the LOFAR survey is dominated by sources with jet powers $\lesssim 10^{38}$ W. The

Table 2. Numbers and fractions of real and simulated sources with size <100 kpc as a function of radio luminosity

Luminosity range	Real sources			Simulation (model i)	Simulation (model ii)
W Hz^{-1}	Total	Small	Small fraction	Small fraction	Small fraction
10^{26}–10^{27}	566	124	0.22 ± 0.02	0.26 ± 0.01	0.38 ± 0.02
10^{25}–10^{26}	944	2683	0.35 ± 0.01	0.33 ± 0.01	0.49 ± 0.02
10^{24}–10^{25}	4443	7457	0.59 ± 0.01	0.38 ± 0.01	0.54 ± 0.02

positions of the time evolution markers on the tracks show that, if all RLAGN have long lifetimes, we expect them to spend most of their lifetime with (unprojected) sizes between a few tens and a few hundreds of kpc and that these predictions seem to be qualitatively consistent with the size distribution of LOFAR sources with luminosities $\gtrsim 10^{25}$ W Hz^{-1}; at lower radio luminosities this is much less clearly the case, with many smaller sources being present. The H18 work showed that the expected size distribution is sensitive to the lifetime function, i.e. the fraction of sources in the population that have active jet lifetimes less than some limiting value. To investigate the lifetime distribution in the current sample we have to carry out more detailed modelling.

4.2. Modelling the linear size distribution

There are two possible approaches to trying to infer population properties of RLAGN by combining models and data. In the first, we would try to estimate the physically interesting parameters of each source (such as jet power Q and source age t) from the available data for that source. Since the easily available observables (radio luminosity and linear size) depend not just on Q and t but also on the unknown source environment, the angle of the source to the line of sight θ, and redshift, the inference of Q is a poorly constrained inverse problem and necessarily will not produce particularly accurate answers for any given source. Better results would be achieved with per-source environmental measures if they were available. We begin by trying the second approach, which is to forward-model subsets of the whole population using known constraints on the distribution of environments, redshifts, and angles to the line of sight. This approach has the advantage that observational limitations like the surface brightness limit can easily be taken into account, but the disadvantage that it is computationally expensive and cannot provide a full exploration of all the underlying distributions. However, it is well suited to the current goal of understanding whether the observed projected linear size distributions are consistent with models.

In order to investigate the implications of the size distribution we first restrict ourselves to sources with $z < 0.8$, motivated by Fig. 7. Above this redshift the sample becomes increasingly dominated by quasars, which are biased in their angle to the line of sight: excluding high redshifts also makes us insensitive to the treatment of radio-quiet quasars discussed above. We then consider three slices in the P–D diagram in the luminosity ranges 10^{24}–10^{25} W Hz^{-1}, 10^{25}–10^{26} W Hz^{-1}, and 10^{26}–10^{27} W Hz^{-1}. As Fig. 8 shows, the last of these should be basically unaffected by surface brightness limitations and thus allows us to constrain the upper end of the lifetime function. These three luminosity ranges sample similar redshift ranges, limited by the optical data (see Fig. 7) and therefore results can be compared without worrying excessively about the cosmological evolution of the population.

Simulated samples were created as described by H18, but we drew the distribution of redshifts from the observed redshift distribution in each luminosity bin, smoothed using a Gaussian kernel density estimator (KDE) with bandwidth 0.05. In general cluster masses can be described by a mass function, which conventionally gives the number of clusters above a given mass as a function of mass (e.g. Reiprich & Böhringer 2002). We took cluster masses from the mass function of Girardi and Giuricin (2000), who show that at $z = 0$ a single Schechter function can describe the local mass function of both groups and clusters. Of course the mass function of RLAGN-hosting clusters and groups may be different from that of clusters and groups in general, but the approach we used should give us a reasonable approximation; we drew environments in the mass range 10^{13} to $10^{15} M_\odot$ from their distribution, which of course implies a strong bias towards the sort of group-mass environments that RLAGN are known to tend to favour based on optical clustering and X-ray studies (e.g. Lilly et al. 1984; Prestage & Peacock 1988; Hill & Lilly 1991; Hardcastle & Worrall 1999; Harvanek et al. 2001; Best 2004; Ineson et al. 2015). We took the probability of a source having a given jet power $p(Q) \propto Q^{-1}$, motivated by the slope of the steep end of the RLAGN luminosity function (see also below, Sect. 4.5). For the trial lifetime functions, we followed H18 and adopted two possibilities: (i) lifetimes are distributed uniformly in linear space between 0 and 1000 Myr and (ii) lifetimes are distributed uniformly in log space between 1 and 1000 Myr. Starting times were distributed uniformly between 0 and 1200 Myr before the time of observation, and remnant sources were included in the models, as they are presumably present in the data; as noted by H18, however, they are expected to constitute only a small fraction of the total for powerful objects. We simulated 10 000 sources for each luminosity range, tuning the range of input jet powers simulated to be appropriate for the luminosity range, and then simulated observations that matched the completeness flux cut of our RLAGN sample and the surface brightness limits that applied to the real data (Fig. 1) and the appropriate luminosity cuts.

Results are shown in Fig. A.1. As was already implied by the P–D diagram presented in Fig. 8, we see that model (i), the uniform-lifetime model, reproduces extremely well the linear size distribution of the most powerful sources ($L_{150} > 10^{25}$ W Hz^{-1}). It perhaps slightly underpredicts the number of very large sources but we have not attempted to adjust the maximum lifetime to fit the observations. The differences in models are clearest when we compare the numbers of small sources (where we define "small" as <100 kpc to include all the upper limits on size in this bin) and so in Table 2 we compare real fractions of small sources as a function of radio luminosity with simulated sources. We see that model (i) agrees very well (to within a few per cent) with the fraction of small sources observed above 10^{25} W Hz^{-1}, but is not at all consistent with the fraction of small sources in the low-luminosity bin. By contrast we see that model (ii) substantially overpredicts the number of small sources in the

more luminous subsamples, while doing a better job with the numbers in the lowest luminosity bin. Model (ii) also substantially underpredicts the number of very large sources observed in the two higher luminosity bins while overpredicting the numbers of large sources in the lowest luminosity bin.

These results have several interesting implications. Firstly, the fact that we can reproduce the size distribution of the most powerful sources with such a simple model as model (i) is striking. Equally, it is clear that the data for the most luminous sources are not consistent with a model, like model (ii), where there are many more short-lived objects than there are long-lived objects. While the very youngest sources are expected to be affected by absorption effects that are not included in the analytical model, this is only relevant for a small fraction of the lifetime of a source (consistent with the small fraction of sources with a low-frequency spectral turnover detected by Callingham et al. 2017) and cannot explain the low numbers of small, luminous sources seen in the LOFAR samples. If the models are anywhere near correct, we must assume that the typical lifetime of a powerful radio galaxy is long, of the order of several hundred Myr at least, such that most of these systems spend most of their lifetimes extended on $\gtrsim$100 kpc scales.

We can then ask why the results are so different at lower luminosities, particularly for the $10^{24} < L_{150} < 10^{25}$ W Hz^{-1} sample. This difference cannot be a redshift-dependent effect, partly because the redshift distributions for the three samples are not very different (Fig. 7) and partly because the modelling takes account of the different redshift distribution of each sample. Several possible explanations may be considered:

- SFG contaminate the samples at low luminosities. This is likely to be the case at some level given the limitations on the colour selection that we discuss above and the lowest luminosity range we consider is such that moderately powerful SFGs might well be present, although we cannot say in what numbers. However, if this is the case then the WISE colour selection must be failing badly for a large population of SFG. Alternatively, some other less obvious contaminating population that generates low-luminosity, compact sources may be present.
- Identifications are worse at low luminosities. This seems unlikely to be the case since the contaminating population are mostly compact sources that usually have a good identification with a nearby galaxy.
- There is a genuine luminosity (or rather jet-power) dependent difference in the lifetime function of low-power and high-power sources, such that low-power sources are genuinely more short-lived and have a lifetime function more like that of model (ii). One possibility is that this difference is related to the different fuel sources available to RLAGN; perhaps sources powered by accretion from the hot phase of the inter-galactic medium have a significantly different lifetime function. Testing this model requires more environmental and AGN accretion mode information than we currently have for this sample; Croston et al. (2019) show that most objects in the sample are not members of the available optical group and cluster catalogues.
- The models get the source physics wrong at low luminosities. To some extent we expect this to be the case; the model overpredicts the radio luminosity of FRI-type sources, which should dominate the lowest luminosity bin, where a significant amount of the energy input of the jet appears to go into non-radiating particles (Croston et al. 2018). But it is difficult to see how this solves the problem; if we are overpredicting radio luminosities in this régime then the jet powers in this luminosity band should actually be higher than in the models and the sources in the simulated sample, if corrected for this, correspondingly larger.
- The models get the environment wrong in a way that induces a luminosity dependence. There are several ways in which this might be possible. For example, the models do not contain the dense, cold central gas that is invoked in "frustration" models of compact steep-spectrum sources, and such a component would have a larger effect on sources of lower jet power. Other, more subtle luminosity-dependent effects include a tendency for lower luminosity sources to lie away from their host group or cluster centre and a dependence of radio luminosity on host environment (Ineson et al. 2015; Ching et al. 2017; Croston et al. 2019).
- The measured sizes are wrong. This is very likely to be the case in faint sources in the low-power, FRI regime, since the surface brightness of lobes or plumes drops off rapidly with distance from the nucleus. We may simply lack the surface brightness sensitivity to map extended structures in many of these sources (cf. Shabala et al. 2017). The H18 model is based on the dimensions of the shocked shell driven out by the momentum flux of the jet, which may well extend beyond the limits of any observable jet for FRI sources, while it is almost always going to be close to the hotspots of resolved FRIIs. If this is the sole explanation for the large number of apparently compact RLAGN then we would expect deeper LOFAR observations still to start to reveal extended structures around many RLAGN that are compact at our current observational sensitivity. Existing surveys at higher frequencies, even with high sensitivity, are likely to be less sensitive to extended structure than LOFAR and would also miss this extended emission. Such an explanation will be testable with "Tier 2" LOFAR surveys data with sensitivities of tens of μJy, or with deep surveys with MeerKAT (Jarvis et al. 2016) or the SKA.

4.3. RLAGN host properties with size

In the models discussed in the previous section, which successfully describe powerful radio sources in the RLAGN sample, large physical size is just a marker of a long-lived source rather than indicating something special about the host galaxy or its environment. The H18 models produce a very few extreme giants (high-power sources in low-density environments) but generally giant radio galaxies are expected to be a natural consequence of observing normal powerful sources towards the end of their lives. The RLAGN sample contains 126 objects with projected physical size >1 Mpc in our adopted cosmology, satisfying the classical definition of a giant radio galaxy (GRG); as noted by H16, LOFAR's combination of low-frequency selection (GRGs are likely to have steep radio spectra) and excellent surface-brightness sensitivity makes it a very productive instrument for studies of such large sources. The sky density of candidate GRGs in the HETDEX survey (about 1 per 4 square degrees) exceeds even that reported by H16 by a factor ~5 thanks to the improved image fidelity, uniform sensitivity, and better optical data of the HETDEX survey. We emphasise that these are giant candidates only, as their sizes have been measured automatically and many of the redshifts are photometric; O'Sullivan et al. (2019) report a case in which the use of a newly obtained spectroscopic redshift instead of the photometric redshift used in this work reduces the projected size of one of these objects from 4 Mpc to 3.4 Mpc. However, a substantial fraction of the GRG redshifts are spectroscopic and there is no reason to

suppose that a large fraction of them will be reclassified below the 1 Mpc threshold either because of their redshifts or because of their automatically measured angular sizes.

The RLAGN sample therefore provides an excellent opportunity to test the hypothesis that the hosts of these objects are not special and that they merely represent the late-time evolution of normal powerful radio galaxies. In this hypothesis properties of the host galaxies, such as their colours and absolute magnitudes, should be close to independent of source projected physical size[8] Fig. A.2 shows such a test. We divided the RLAGN sample into the three luminosity bins of the previous section and then binned in projected linear size, taking the average of rest-frame K_s-band magnitude and WISE band 2/band 3 colour (see Figs. 5 and 4 for distributions of the whole sample in these parameters). The upper limits on physical size are treated as measurements for purposes of binning in these plots; as almost all of these limits are less than 100 kpc (Fig. A.1) there is very little ambiguity in the binning. A tiny minority of sources without WISE photometry are ignored.

What we see in the first panel of Fig. A.2 is that the absolute magnitudes of all three samples show very little variation with physical size, barring a slight deviation from the mean in the 200–500 kpc bin for the lowest luminosity sources for which we have no explanation. Broadly this plot is consistent with the idea that all powerful RLAGN hosts have an absolute magnitude around −24.0, and scatter of a few tenths of a magnitude at most irrespective of their radio luminosity or size. This is consistent with what is seen for the whole population in Fig. 5 and this standard infrared magnitude is of course the basis of the well-known K−z relation for radio galaxies (Lilly & Longair 1984). Sabater et al. (2019) discuss in more detail the distributions of the host galaxy masses of RLAGN.

The second panel of Fig. A.2 shows that the mean WISE colour of the highest luminosity sample is constant with length, that of the intermediate-luminosity sample deviates from a constant value in the lowest size bin, and for the lowest luminosity sample the colour is very strongly dependent on projected linear size over the whole range of sizes studied. It is very striking that the population that shows such a deviation from the hypothesis that all RLAGN hosts are the same is precisely the population that we previously suggested may be contaminated by some other type of source, such as SFG. The colour deviations seen in this figure are in the sense that sources move closer to SFG colours as their sizes get smaller. We emphasise that the average colours never become as extreme as colours that we expect from SFG, which would be impossible given the WISE colour selection we used for the RLAGN sample, and that type 1 and type 2 quasars and Seyfert galaxies also have higher $W2 - W3$ colours due to the torus. We conclude that it is plausible that the low-luminosity RLAGN sample contains more than one population. However, the constancy of host galaxy colours and masses as a function of size for the highest luminosity bins provides strong evidence that powerful RLAGN are homogeneous: there is no evidence that the largest, oldest RLAGN have different hosts from their smaller counterparts. Investigation of the related question about environment – some relationship between size and environment is a prediction of the models – will require a data set with more environmental information than is currently available.

[8] For powerful radio sources there is evidence (Best et al. 1997) that the early stages of radio galaxy evolution are associated with an aligned, blue component in the host galaxy, which may be connected to, for example jet-induced star formation, and which disappears later in the lifetime of a source. However, this effect is much less obvious in the infrared bands that we use for this test, and as this effect is also seen in sources much more powerful than those in our sample, we neglect it here.

4.4. Bulk inference of jet power

Noting that tracks of constant jet power Q describe characteristic curves in the P−D diagram for a given environment and redshift (Fig. 8), we can now investigate a simple model-dependent method for inferring jet power Q from the observed redshift, L_{150} and projected linear size D for the RLAGN sample. We do not have direct measurements of environmental richness for most of these objects (see Croston et al. 2019 for a discussion of the available constraints) and similarly almost no information about the angle to the line of sight for a given source; other potentially useful parameters such as the axial ratio of the lobes or their integrated spectral index (H18) have not yet been measured. Thus we focus on what can be inferred from z, L_{150} and D.

Our approach, as in Sect. 4.2, is to generate populations of simulated sources that match the LOFAR observations in terms of observational selection criteria and populate the observable regions of the P−D diagram using the models of H18. In the absence of any environmental information we assume the same distribution of source environments as earlier and the same distribution of angles to the line of sight. We can then estimate the jet power corresponding to any particular position in the P−D diagram by looking at the mean jet power of simulated sources that lie close to that location: the uncertainty in the inference comes from the distribution of the local simulated sources. This method automatically takes into account the unknown angle to the line of sight and the unknown environment, as long as the distributions we use are approximately correct. To take into account the strong redshift dependence of radio luminosity as a result of inverse-Compton losses, we generate populations for a number of redshifts in the range $0 < z < 0.8$ where we have a uniform population of RLAGN, and interpolate between the nearest one or two for any given source.

In detail, we take a set of redshifts $(0.05, 0.15, \ldots, 0.75)$ and, for each redshift, populate a P−D diagram using jet powers in the range $10^{34} < Q < 10^{40}$, where we assume a uniform distribution of Q in log space to make sure that all of the luminosity range is populated. We take the lifetime function to be a uniform distribution of lifetimes in linear space, as in model (i) of Sect. 4.2. We apply the LOFAR observational selection criteria to the simulated sources, giving us of order 5000 sources per redshift slice. A plot showing the binned mean Q as a function of position in the P−D diagram for the stacked simulated sample, and the dispersion in inferred Q introduced by different environments, projection angles, ages, and redshifts, is shown in Fig. A.3.

We then restrict the RLAGN sample to $z < 0.8$ and $L_{150} > 10^{23}$ W Hz^{-1} giving us a total of 18 948 objects; below that luminosity we regard the jet models as uncertain and linear sizes above 100 kpc are not expected to be present. Then, for each resolved object in the restricted sample, we take the Gaussian-weighted mean Q in log space of all of the simulated points within 3σ of the position of the real object in P, D space, where we define the width of the weighting Gaussian $\sigma = 0.04$ dex, corresponding to a fractional error of 10%. This is reasonable at least for the luminosities, where the absolute flux calibration uncertainty is probably of this order: we have no real constraints on the uncertainties on projected physical size but a 10% uncertainty seems plausible. For unresolved objects we instead use the upper limit on size from earlier in this section and consider all

simulated sources consistent with that limit and within 3σ of the position defined by the radio luminosity. In both cases an error on Q can be estimated by bootstrapping from the sample of simulated sources: this automatically accounts for the uncertainties on inference in parts of the $P-D$ plane that can be populated by a large range of jet powers. Typically the errors estimated in this way are of the order of 10% in Q, which is reasonable given the assumed input uncertainties on L_{150}. In a few cases the errors are much larger (>0.5Q) or there are not enough points in simulated $P-D$ space for the estimation or bootstrap process to work: in this case we flag the measured values of Q as bad. In total 19 356 objects have a good estimated jet power.

Figure A.4 shows the relationship between radio luminosity and jet power that we infer for the RLAGN sample. We overplot for comparison the relations derived by Cavagnolo et al. (2010), Daly et al. (2012), Heckman & Best (2014), Ineson et al. (2017). We see that the Ineson et al. (2017) relation agrees well with our inference, which is not surprising since its methods are closest to the assumptions of the H18 models. The inference in this work is for slightly higher jet powers for a given radio luminosity when compared to the relations of Ineson et al. (2017) or H18. But this is probably a result of the assumptions that we make regarding environment – the typical environment of our simulated sources is poor and so less radio emission is produced for a given jet power – and partly due to the higher average redshift, $z \sim 0.5$ of the RLAGN sample, which gives rise to lower radio luminosity for a given jet power compared to $z = 0$ because of stronger inverse-Compton losses. Our results are similar to, but generally predict slightly higher powers than, the results of Daly et al. (2012), which are based on powerful FRIIs; their method (O'Dea et al. 2009) uses spectral ages involving minimum-energy magnetic fields and so would be expected to underestimate both the age and energetic content given the observed sub-equipartition field strengths (Hardcastle et al. 2002; Kataoka & Stawarz 2005; Croston et al. 2005; Ineson et al. 2017), but clearly these effects cancel to some extent in practice. There is much less good agreement with the cavity-based relations of Cavagnolo et al. (2010) or Heckman & Best (2014) at low luminosities. For the most powerful sources in our sample, with luminosities $L_{150} \approx 10^{26}\,\mathrm{W\,Hz^{-1}}$, however, all jet power estimates are of the same order of magnitude (see discussion by Heckman & Best 2014) and the cavity relations are actually reasonably consistent with our inference, although it should be noted that the cavity relations are not generally supported by much data at these radio luminosity values. For low-luminosity sources, as already noted, the H18 model is likely to overestimate the radio luminosity for a given jet power and the truth is likely to lie somewhere in between our inferred values and the cavity models.

Better environmental information for our sample would improve our inference process and decrease the uncertainties on the inferred jet powers. Without this information – or other information that we might be able to make use of, such as constraints on angle to the line of sight for individual sources – these jet powers are still only estimates that are not expected to be particularly accurate for any given source. Nevertheless this work demonstrates the feasibility of bulk estimation of Q without resorting to simple scaling relationships based on radio luminosity.

4.5. Jet kinetic luminosity function

We can use the inferred jet powers from the previous subsection to construct a jet kinetic luminosity function that represents our current best estimate of this quantity for the LOFAR RLAGN sample, bearing in mind that we excluded very low-luminosity radio sources and those with poor jet model fits and that the inclusion of these sources would probably slightly increase the normalisation of the luminosity function. To do this we simply apply the standard V_{max} method (Schmidt 1968; Condon 1989) to the jet power Q, calculating the volume based on the combined radio and optical constraints. Only sources within $0.01 < z < 0.7$ are considered and an i-band limit of 21.5 mag is imposed to ensure photometric redshift completeness. The radio V_{max} is calculated as $\int d_{\mathrm{max}} \mathrm{d}A$, where the completeness function, dA, is determined from the LoTSS rms map. For the optical, V_{max} is calculated from our imposed i-band limit of 21.5 mag after the optical magnitudes are corrected for the Galactic reddening calculated by Schlafly & Finkbeiner (2011) and K-corrected based on the rest-frame magnitudes calculated by Duncan et al. (2019). Our approach automatically takes into account the unknown RLAGN duty cycle since only the density of LOFAR-detected, luminous sources is calculated. The results are shown in Fig. A.5. We see, as expected, that the kinetic luminosity function appears very like the radio luminosity function, in that it is flatter at low jet powers and steepens at higher powers; there is little evolution in the kinetic luminosity function with redshift, as expected since the dominant LERG population is known not to evolve strongly (Best et al. 2014; Williams et al. 2018).

The kinetic luminosity function $\rho(Q)$ is a physically important quantity in that the integral $\int Q\rho(q)\mathrm{d}\log(Q)$ tells us the energy per comoving volume injected by all RLAGN jets into their host environments, and this work represents the first attempt to construct $\rho(Q)$ from bulk inference of jet powers for a large sample. Integration of our kinetic luminosity function over the observed range gives a total RLAGN kinetic luminosity density, including the effects of all the sources in our luminosity range not excluded by surface-brightness selection, of $7 \times 10^{31}\,\mathrm{W\,Mpc^{-3}}$. The integral is dominated (Fig. A.5, right-hand panel) by powerful sources, peaking at jet powers around 10^{38} W, and so is not sensitive to the uncertain jet powers of low-power objects; indeed, the peak lies in the region in which inference, FRII power estimates, and cavity power estimates all give similar results (Fig. A.4). The integral of the kinetic luminosity function may be compared to the total radiative (cooling) luminosity density of groups and clusters. Integrating the Schechter function fitted by Böhringer et al. (2014) to the local cluster luminosity function between cluster luminosities of 10^{42} and $10^{46}\,\mathrm{erg\,s^{-1}}$, we obtain a cooling luminosity of $2 \times 10^{31}\,\mathrm{W\,Mpc^{-3}}$, a result that is insensitive to the limits of integration because of the form of the Schechter function. Thus the RLAGN population found in this work can in principle completely offset, in statistical terms, all the local radiative cooling of the environments that they are expected to occupy, even allowing for the fact that some of the kinetic luminosity goes into cosmic rays that may not play much of a role in heating the thermal plasma in groups and clusters[9]. This is strong support for feedback models in which RLAGN provide the "maintenance mode" required to prevent the hot phase of the environment of their host galaxies cooling back onto the central galaxy and reinvigorating star formation.

[9] In numerical models, the fraction of the jet kinetic power that heats the external environment is a little over 0.5 during the active source lifetime (Hardcastle & Krause 2014; English et al. 2016), but much of the energy stored in the lobes is then lost to the large-scale environment in the remnant phase (English et al., in prep.); the fraction of energy that remains in cosmic rays at late times is not well known, although it is an important ingredient in models of cluster evolution.

We emphasise that this is only the first step towards the construction of a truly reliable kinetic luminosity function. A key problem is the effect on the kinetic luminosity function of the large number of low-luminosity sources that probably in reality do not lie on the relation between Q and L_{150} that is implied by the H18 models (see Sect. 4.2). Their radio luminosities for a given Q are expected to fall significantly below the model expectations, since some of their internal pressure is provided by a non-radiating particle population, and so their inferred jet powers should actually be higher than the values we used. In addition, if larger versions of these sources exist, it is entirely possible that we are systematically missing numbers of them due to the surface-brightness limitations on our survey, which feed through into limitations on the sources that may be observed in the power/linear-size diagram (Fig. 8). Modelling of these two effects will be important in order to make progress, though as we noted above, these sources will have only a small effect on the integral of the luminosity function unless their jet powers are ∼2 orders of magnitude higher than we infer them to be. On the other hand, powerful sources are likely to live in richer environments than we have modelled (Ineson et al. 2015) and to have lower jet powers than we infer, which would have a stronger effect, because of the shape of the luminosity function, on the integrated kinetic power input we find. It is entirely possible that some of the curvature in the observed radio luminosity function is due to these combined effects of radio galaxy physics, observational selection, and environment.

Earlier calculations of the kinetic luminosity density were carried out by for example Best et al. (2006), Smolčić et al. (2017), and these compare very well to the results we derive from integrating the kinetic luminosity function. Best et al. (2006) computed jet powers for the $z \approx 0$ SDSS/FIRST-based sample of Best et al. (2005), i.e. a very comparable, although smaller, sample, using a version of the cavity power estimates discussed in Sect. 1, which gives a very flat jet power/radio luminosity dependence compared to ours. They nevertheless obtained a luminosity density of 4×10^{31} W Mpc^{-3}, which agrees with ours to within a factor 2. Smolčić et al. (2017) used deep VLA data from the COSMOS field with excellent multiwavelength counterparts and so were able to probe out to much higher redshifts than we can achieve, allowing them to investigate the cosmic evolution of the kinetic luminosity function. On the other hand their sampling of the local Universe was necessarily limited by the small volume available to them. They made use of the Willott et al. (1999) radio luminosity/kinetic luminosity relation, which, as discussed in Sect. 1, can only ever be an approximation; however, as shown by H18, suitable choices of normalisation of the Willott relation can bring it into agreement with more sophisticated models for large, mature sources at a particular redshift, and the values adopted for the "uncertainty parameter" of the Willott relation by Smolčić et al. (2017) span the range that would be appropriate for powerful, mature sources at $z = 0$ in the H18 models. Given these differences in the model and the data, there is excellent agreement between our kinetic luminosity density of 7×10^{31} W Mpc^{-3} at $z < 0.7$ and their estimates ranging between ∼2 and $\sim 5 \times 10^{31}$ W Mpc^{-3} for their preferred uncertainty parameter over the same redshift range. We caution, however, that in the H18 models the uncertainty factor is a function of environment and redshift, and therefore it is not safe to assume that it is constant over the lifetime of the Universe. Smolčić et al. (2017) further estimated the luminosity density required by the SAGE model of Croton et al. (2016) to be $\sim 7 \times 10^{31}$ W Mpc^{-3} (roughly constant or slightly declining over the range $0 < z < 1$), which is again in excellent agreement with our calculation, although it somewhat exceeds the observed group/cluster X-ray cooling luminosity density. Combining our work with that of Best et al. (2006), Smolčić et al. (2017), we can conclude that estimates of the effects of RLAGN on their local environment are in remarkably good agreement with both X-ray observations and models. Sabater et al. (2019), using cavity-based jet power estimates, come to a similar conclusion in their study of the nearby AGN population, showing that the jet power of RLAGN is more than sufficient to offset the cooling of gas in their host ellipticals.

5. Summary and future work

In this paper we have constructed a sample of RLAGN from the value-added catalogue drawn from the LoTSS survey of the HETDEX Spring field, based on a combination of radio properties, spectroscopic information where available, and WISE colour information or radio luminosity otherwise; this is not a true radio-excess sample of the type selected by H16 or Smolčić et al. (2017), but is expected to be very comparable to such a sample. Although only a small fraction of the total radio catalogue can be classed robustly as RLAGN using our methods because many objects are SFGs or do not yet have good enough optical identifications or redshifts to be classified, this process still yields one of the largest homogeneous RLAGN data sets in existence, and one from which many interesting individual objects can be drawn.

In the current paper we have focussed on new conclusions that can be drawn about the properties of the RLAGN population using this large sample. To do this we require a model of radio galaxy evolution, which can give us observable quantities such as radio luminosity and total linear size from model inputs such as jet power, redshift, environment, and time. We chose to work with the models of H18 but it is important to note that this is not the only radio source model available; different analytical models make different approximations and it will be important in future to cross-calibrate these models and to see what differences the use of a different model makes to the inference of population properties.

Comparison with the H18 model tracks in the power/linear-size (P–D) plot showed that the distribution of source sizes in the luminosity range best sampled by the LOFAR data was perhaps surprisingly consistent with a model in which most sources in the luminosity range $10^{25} < L_{150} < 10^{27}$ W Hz^{-1} are long-lived objects in relatively poor (group-like) environments. The critical unknown distribution in this case is the lifetime function, the distribution of total lifetimes of RLAGN. We showed that a uniform distribution in the range 0–1000 Myr reproduced well the distributions of projected linear sizes of the powerful sources; there are relatively few physically small sources in this radio luminosity range. On the other hand, at the lower end of the LOFAR luminosity range there are many more small sources, even when surface-brightness selection effects are taken into account as we are able to do with our modelling. This low-luminosity, compact population has been noted previously (e.g. Sadler et al. 2014; Baldi et al. 2015; Whittam et al. 2017) but we find large numbers of these objects, requiring either a very different lifetime distribution at low luminosities, a breakdown of the underlying models, or some contamination by a separate population of objects. Based on analysis of the colour and magnitude of host galaxies binned by linear size, we show that it is entirely plausible that the luminous LOFAR RLAGN are a homogeneous population in which large sizes (including those of the many >1 Mpc giants in our sample) are simply an effect of old age; but at low luminosities the strong dependence of colour on physical

size suggests that more than one population is present. This may be the result of contamination by star-forming objects or it may indicate that more than one RLAGN population, perhaps with different large-scale fuelling mechanisms, is present.

Finally, we used the H18 models to attempt to infer the jet powers Q for LOFAR sources based only on their positions on the P–D diagram, marginalising over the unknown environments and angles to the line of sight of the LOFAR sources. This is a proof of principle for bulk inference of Q and maybe other source parameters from large volumes of data. Again, the results are model-dependent and also dependent on our assumptions, particularly relating to environment. It seems likely, as noted by H18, that observations, for example those of Ineson et al. (2015), require some intrinsic relationship between jet power and environmental richness that is not present in the models used in this work. Nevertheless we are able to derive jet powers that agree reasonably well with results already present in the literature and allow us to construct the first large-scale jet kinetic luminosity function based on inference of jet powers rather than simple scaling relations with radio luminosity. The distribution of jet powers, and the integral of the jet kinetic luminosity function, are key parameters in models of galaxy formation and evolution and, as shown by Smolčić et al. (2017), it is now possible to compare these quantities to the assumptions made in such models. Integration of the existing luminosity function, which will be substantially refined in future, suggests that the energy input from RLAGN is more than adequate to offset all of the observed X-ray radiative cooling of the group and cluster population in which we assume the RLAGN to lie; the value we obtain is consistent both with independent observational estimates of the kinetic luminosity density by Best et al. (2006), Smolčić et al. (2017) but also with galaxy evolution models.

Forthcoming developments in LoTSS observations and ancillary data will allow substantial improvements to be made in all of these areas in the near future. Star formation/RLAGN separation, as well as the quality of redshifts and thus luminosities and physical sizes, should be greatly improved by the WEAVE-LOFAR project (Smith et al. 2016), which will provide both spectroscopic redshifts and emission-line diagnostics for large numbers of LOFAR sources, including those at high redshift for which we currently have little information. High-resolution images using the LOFAR international baselines will help with source size measurements, optical identifications, and RLAGN/star formation separation; the Very Large Array Sky Survey (VLASS[10]), when complete, will also be very useful for the identification of flat-spectrum cores in LOFAR objects and for resolving bright, compact sources. A key missing ingredient in our bulk inference in this paper is information on the environments of the RLAGN. Environmental information can be obtained, for example from SDSS, at low redshifts (Croston et al. 2019) but the HETDEX sky area is too small to obtain a representative sample of powerful AGN. The much larger sky areas provided by the full LoTSS survey, which will reach 10 000 deg^2 of coverage in the next two years, will allow us to probe a larger range of radio luminosities and environments at low redshift and to take full account of environmental information both from SDSS and from the forthcoming *e-ROSITA* X-ray survey. The lessons learned from this and the planned subsequent LoTSS work will inform the even larger surveys that will be carried out with the SKA, but, as the present paper demonstrates, the era of big data for RLAGN surveys is already here.

[10] https://science.nrao.edu/science/surveys/vlass

Acknowledgements. We would like to thank Sarah Needleman (University of Cambridge: summer student in 2016) and Sinan Hassan (University of Hertfordshire: BSc investigation in 2018) for their contributions to this work. MJH and WLW acknowledge support from the UK Science and Technology Facilities Council (STFC) [ST/M001008/1]. JHC acknowledges support from the STFC under grants ST/R00109X/1 and ST/R000794/1. KJD acknowledges support from the ERC Advanced Investigator programme NewClusters 321271. PNB and JS are grateful for support from STFC via grant ST/M001229/1. MJJ acknowledges support from the Oxford Hintze Centre for Astrophysical Surveys which is funded through generous support from the Hintze Family Charitable Foundation. GG acknowledges a CSIRO OCE Postdoctoral Fellowship. SM acknowledges funding through the Irish Research Council New Foundations scheme and the Irish Research Council Postgraduate Scholarship scheme. FdG is supported by the VENI research programme with project number 1808, which is financed by the Netherlands Organisation for Scientific Research (NWO). IP acknowledges support from INAF under PRIN SKA/CTA "FORECaST". RKC is grateful for support from STFC. SPO acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG) under grant BR2026/23. LOFAR, the LOw Frequency ARray designed and constructed by ASTRON, has facilities in several countries, which are owned by various parties (each with their own funding sources), and are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy. The ILT resources have benefited from the following recent major funding sources: CNRS-INSU, Observatoire de Paris and Université d'Orléans, France; BMBF, MIWF-NRW, MPG, Germany; Science Foundation Ireland (SFI), Department of Business, Enterprise and Innovation (DBEI), Ireland; NWO, The Netherlands; the Science and Technology Facilities Council, UK; Ministry of Science and Higher Education, Poland. Part of this work was carried out on the Dutch national e-infrastructure with the support of the SURF Cooperative through grant e-infra 160022 & 160152. The LOFAR software and dedicated reduction packages on https://github.com/apmechev/GRID_LRT were deployed on the e-infrastructure by the LOFAR e-infragroup, consisting of J. B. R. Oonk (ASTRON & Leiden Observatory), A. P. Mechev (Leiden Observatory) and T. Shimwell (ASTRON) with support from N. Danezi (SURFsara) and C. Schrijvers (SURFsara). This research has made use of the University of Hertfordshire high-performance computing facility (http://uhhpc.herts.ac.uk/) and the LOFAR-UK computing facility located at the University of Hertfordshire and supported by STFC [ST/P000096/1]. This research made use of Astropy, a community-developed core Python package for astronomy (Astropy Collaboration 2013) hosted at http://www.astropy.org/, of Matplotlib (Hunter 2007), of APLpy, an open-source astronomical plotting package for Python hosted at http://aplpy.github.com/, and of topcat and stilts (Taylor 2005). The Pan-STARRS1 Surveys (PS1) have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, and the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), and the Los Alamos National Laboratory. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the US Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University. The National Radio Astronomy Observatory (NRAO) is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. 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, and NEOWISE, which is a project of the Jet Propulsion Laboratory/California Institute of Technology. WISE and NEOWISE are funded by the National Aeronautics and Space Administration.

References

Allen, S. W., Dunn, R. J. H., Fabian, A. C., Taylor, G. B., & Reynolds, C. S. 2006, MNRAS, 372, 21
Assef, R. J., Kochanek, C. S., Brodwin, M., et al. 2010, ApJ, 713, 970
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Baldi, R. D., Capetti, A., & Giovannini, G. 2015, A&A, 576, A38
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Baldwin, J. 1982, IAU Symp., 97, 21
Basson, J. F., & Alexander, P. 2003, MNRAS, 339, 353
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., Longair, M. S., & Röttgering, H. J. A. 1997, MNRAS, 292, 758
Best, P. N., Kauffmann, G., Heckman, T. M., et al. 2005, MNRAS, 362, 25
Best, P. N., Kaiser, C. R., Heckman, T. M., & Kauffmann, G. 2006, MNRAS, 368, L67
Best, P. N., Ker, L. M., Simpson, C., Rigby, E. E., & Sabater, J. 2014, MNRAS, 445, 955
Bîrzan, L., Rafferty, D. A., McNamara, B. R., Wise, M. W., & Nulsen, P. E. J. 2004, ApJ, 607, 800
Bîrzan, L., Rafferty, D. A., Nulsen, P. E. J., et al. 2012, MNRAS, 427, 3468
Blandford, R. D., & Rees, M. J. 1974, MNRAS, 169, 395
Blundell, K. M., Rawlings, S., & Willott, C. J. 1999, AJ, 117, 677
Böhringer, H., Chon, G., & Collins, C. A. 2014, A&A, 570, A31
Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Calistro Rivera, G., Lusso, E., Hennawi, J. F., & Hogg, D. W. 2016, ApJ, 833, 98
Calistro Rivera, G., Williams, W. L., Hardcastle, M. J., et al. 2017, MNRAS, 469, 3468
Callingham, J. R., Ekers, R. D., Gaensler, B. M., et al. 2017, ApJ, 836, 174
Cavagnolo, K. W., McNamara, B. R., Nulsen, P. E. J., et al. 2010, ApJ, 720, 1066
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05560]
Ching, J. H. Y., Croom, S. M., Sadler, E. M., et al. 2017, MNRAS, 469, 4584
Condon, J. J. 1989, ApJ, 338, 13
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
Croston, J. H., Hardcastle, M. J., & Birkinshaw, M. 2005, MNRAS, 357, 279
Croston, J. H., Ineson, J., & Hardcastle, M. J. 2018, MNRAS, 476, 1614
Croston, J. H., Hardcastle, M. J., Mingo, B., et al. 2019, A&A, 622, A10 (LOFAR SI)
Croton, D., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 111
Croton, D. J., Stevens, A. R. H., Tonini, C., et al. 2016, ApJS, 222, 22
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
Daly, R. A., Sprinkle, T. B., O'Dea, C. P., Kharb, P., & Baum, S. A. 2012, MNRAS, 423, 2498
Duncan, K., Sabater, J., Rottgering, H., et al. 2019, A&A, 622, A3 (LOFAR SI)
Eilek, J. A., & Owen, F. N. 2006, in Heating vs. Cooling in Galaxies and Clusters of Galaxies, eds. H. Böhringer, G. W. Pratt, & P. Schuecker (Heidelberg: Springer-Verlag), 0612111
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
English, W., Hardcastle, M. J., & Krause, M. G. H. 2016, MNRAS, 461, 2025
Fabian, A. C., Nulsen, P. E. J., & Canizares, C. R. 1984, Nature, 310, 30
Fabian, A. C., Sanders, J. S., Ettori, S., et al. 2000, MNRAS, 318, L65
Gaspari, M., Ruszkowski, M., & Oh, S. P. 2013, MNRAS, 432, 3401
Girardi, M., & Giuricin, G. 2000, ApJ, 540, 45
Godfrey, L. E. H., & Shabala, S. S. 2016, MNRAS, 456, 1172
Godfrey, L. E. H., Morganti, R., & Brienza, M. 2017, MNRAS, 471, 891
Gürkan, G., Hardcastle, M. J., & Jarvis, M. J. 2014, MNRAS, 438, 1149
Gürkan, G., Hardcastle, M. J., Smith, D. J. B., et al. 2018, MNRAS, 475, 3010
Gürkan, G., Hardcastle, M. J., Best, P. N., et al. 2019, A&A, 622, A11 (LOFAR SI)
Hardcastle, M. J. 2018, MNRAS, 475, 2768
Hardcastle, M. J., & Croston, J. H. 2010, MNRAS, 404, 2018
Hardcastle, M. J., & Krause, M. G. H. 2013, MNRAS, 430, 174
Hardcastle, M. J., & Krause, M. G. H. 2014, MNRAS, 443, 1482
Hardcastle, M. J., & Worrall, D. M. 1999, MNRAS, 309, 969
Hardcastle, M. J., Birkinshaw, M., Cameron, R., et al. 2002, ApJ, 581, 948
Hardcastle, M. J., Evans, D. A., & Croston, J. H. 2007, MNRAS, 376, 1849
Hardcastle, M. J., Evans, D. A., & Croston, J. H. 2009, MNRAS, 396, 1929
Hardcastle, M. J., Massaro, F., Harris, D. E., et al. 2012, MNRAS, 424, 1774
Hardcastle, M. J., Gürkan, G., van Weeren, R. J., et al. 2016, MNRAS, 462, 1910
Harvanek, M., Ellingson, E., Stocke, J. T., & Rhee, G. 2001, AJ, 122, 2874
Heckman, T. M., & Best, P. N. 2014, ARA&A, 52, 589
Heinz, S., Brüggen, M., Young, A., & Levesque, E. 2006, MNRAS, 373, L65
Herpich, F., Mateus, A., Stasińska, G., Cid Fernandes, R., & Vale Asari, N. 2016, MNRAS, 462, 1826
Hill, G. J., & Lilly, S. J. 1991, ApJ, 367, 1
Hill, G. J., Gebhardt, K., Komatsu, E., et al. 2008, ASP Conf. Ser., 399, 115
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Ineson, J., Croston, J. H., Hardcastle, M. J., et al. 2015, MNRAS, 453, 2682
Ineson, J., Croston, J. H., Hardcastle, M. J., & Mingo, B. 2017, MNRAS, 467, 1586
Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, ApJ, 735, 112
Jarvis, M., Taylor, R., Agudo, I., et al. 2016, Proc. MeerKAT Science: On the Pathway to the SKA (MeerKAT2016), https://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=277, 6
Kaiser, C. R., & Alexander, P. 1997, MNRAS, 286, 215
Kaiser, C. R., Dennett-Thorpe, J., & Alexander, P. 1997, MNRAS, 292, 723
Kataoka, J., & Stawarz, Ł. 2005, ApJ, 622, 797
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961
Krause, M. 2005, A&A, 431, 45
Laing, R. A., Riley, J. M., & Longair, M. S. 1983, MNRAS, 204, 151
Lilly, S. J., & Longair, M. S. 1984, MNRAS, 211, 833
Lilly, S. J., McLean, I. S., & Longair, M. S. 1984, MNRAS, 209, 401
Luo, Q., & Sadler, E. M. 2010, ApJ, 713, 398
Mainzer, A., Bauer, J., Grav, T., et al. 2011, ApJ, 731, 53
Mateos, S., Alonso-Herrero, A., Carrera, F. J., et al. 2012, MNRAS, 426, 3271
McNamara, B. R., & Nulsen, P. E. J. 2012, New J. Phys., 14, 055023
Mendygral, P. J., Jones, T. W., & Dolag, K. 2012, ApJ, 750, 166
Mingo, B., Watson, M. G., Rosen, S. R., et al. 2016, MNRAS, 462, 2631
Mocz, P., Fabian, A. C., & Blundell, K. M. 2011, MNRAS, 413, 1107
Mohan, N., & Rafferty, D. 2015, Astrophysics Source Code Library [record ascl:1502.007]
Morganti, R., Oosterloo, T. A., Tadhunter, C. N., van Moorsel, G., & Emonts, B. 2005, A&A, 439, 521
Nesvadba, N. P. H., Lehnert, M. D., De Breuck, C., Gilbert, A. M., & van Breugel, W. 2008, A&A, 491, 407
Norris, R. P., Hopkins, A. M., Afonso, J., et al. 2011, PASA, 28, 215
O'Dea, C. P., Daly, R. A., Kharb, P., Freeman, K. A., & Baum, S. A. 2009, A&A, 494, 471
O'Sullivan, S. P., Machalski, J., Van Eck, C. L., et al. 2019, A&A, 622, A16 (LOFAR SI)
Pizzolato, F., & Soker, N. 2005, ApJ, 632, 821
Prestage, R. M., & Peacock, J. A. 1988, MNRAS, 230, 131
Reiprich, T. H., & Böhringer, H. 2002, ApJ, 567, 716
Reynolds, C. S., Heinz, S., & Begelman, M. C. 2002, MNRAS, 332, 271
Rovilos, E., Georgantopoulos, I., Akylas, A., et al. 2014, MNRAS, 438, 494
Russell, H. R., McNamara, B. R., Fabian, A. C., et al. 2017, MNRAS, 472, 4024
Sabater, J., Best, P. N., Hardcastle, M. J., et al. 2019, A&A, 622, A17 (LOFAR SI)
Sadler, E. M., Ekers, R. D., Mahony, E. K., Mauch, T., & Murphy, T. 2014, MNRAS, 438, 796
Sakelliou, I., Peterson, J. R., Tamura, T., et al. 2002, A&A, 391, 903
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Scheuer, P. A. G. 1974, MNRAS, 166, 513
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schmidt, M. 1968, ApJ, 151, 393
Secrest, N. J., Dudik, R. P., Dorland, B. N., et al. 2015, ApJS, 221, 12
Shabala, S. S., Deller, A., Kaviraj, S., et al. 2017, MNRAS, 464, 4706
Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017, A&A, 598, A104
Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, A&A, 622, A1 (LOFAR SI)
Smith, D. J. B., Dunne, L., da Cunha, E., et al. 2012, MNRAS, 427, 703
Smith, D. J. B., Best, P. N., Duncan, K. J., et al. 2016, in SF2A-2016: Proc. Annual meeting of the French Society of Astronomy and Astrophysics, eds. C. Reylé, J. Richard, L. Cambrésy, et al., 271
Smolčić, V., Delvecchio, I., Zamorani, G., et al. 2017, A&A, 602, A2
Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30
Taylor, M. B. 2005, ASP Conf. Ser., 347, 29
Turner, R. J., & Shabala, S. S. 2015, ApJ, 806, 59
Turner, R. J., Rogers, J. G., Shabala, S. S., & Krause, M. G. H. 2018, MNRAS, 473, 4179
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518
Voit, G. M., & Donahue, M. 2015, ApJ, 799, L1
Whittam, I. H., Jarvis, M. J., Green, D. A., Heywood, I., & Riley, J. M. 2017, MNRAS, 471, 908
Williams, W. L., Calistro Rivera, G., Best, P. N., et al. 2018, MNRAS, 475, 3429
Williams, W. L., Hardcastle, M. J., Best, P. N., et al. 2019, A&A, 622, A2 (LOFAR SI)
Willott, C. J., Rawlings, S., Blundell, K. M., & Lacy, M. 1999, MNRAS, 309, 1017
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Zanni, C., Bodo, G., Rossi, P., et al. 2003, A&A, 402, 949

Appendix A: Additional figures

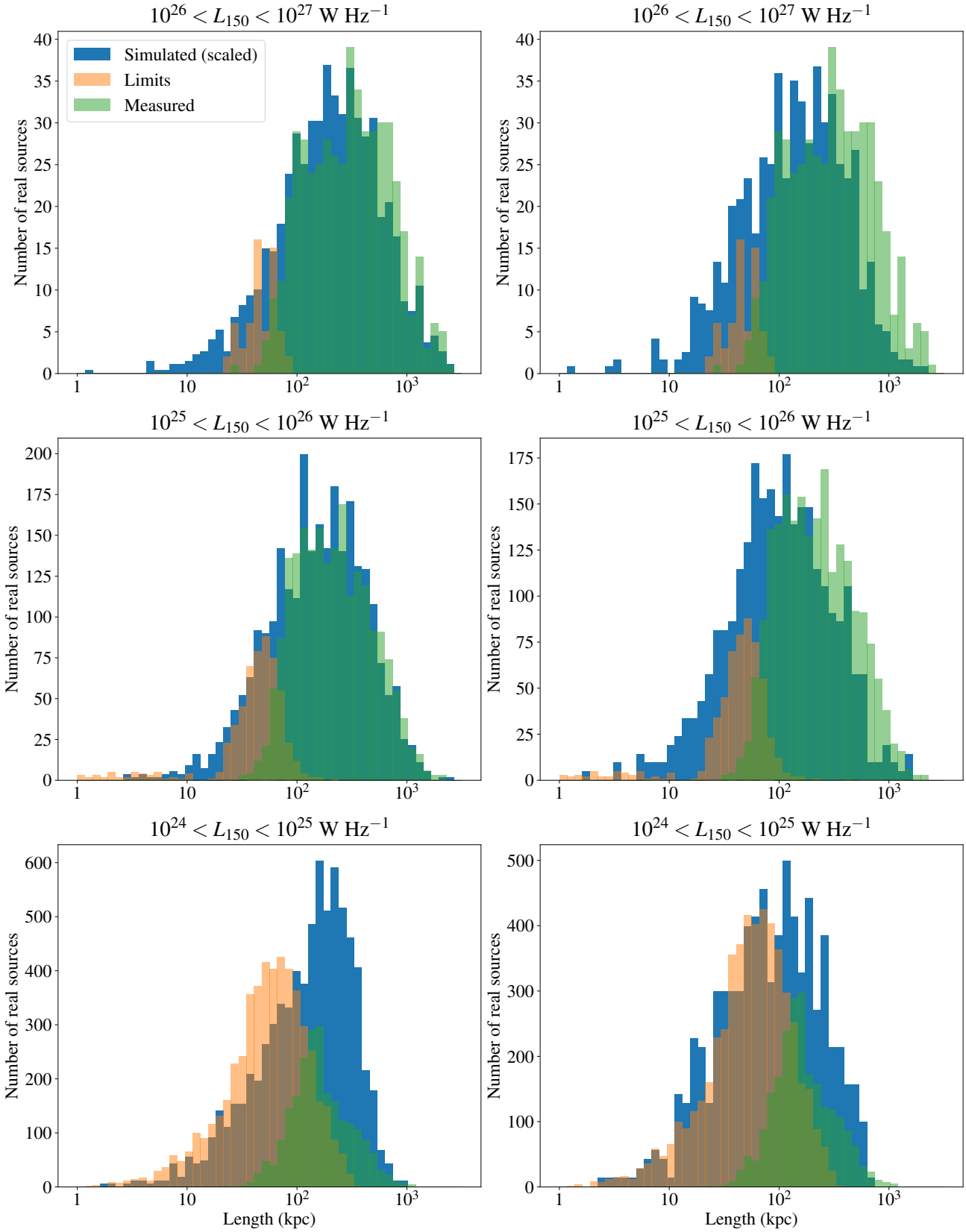

Fig. A.1. Distributions of real and simulated projected sizes for three luminosity ranges (*from top to bottom panels*, three slices of the $P-D$ diagram of Fig. 8) and two lifetime distributions (*left panels*: model (i), uniform, *right panels*: model (ii), log-uniform).

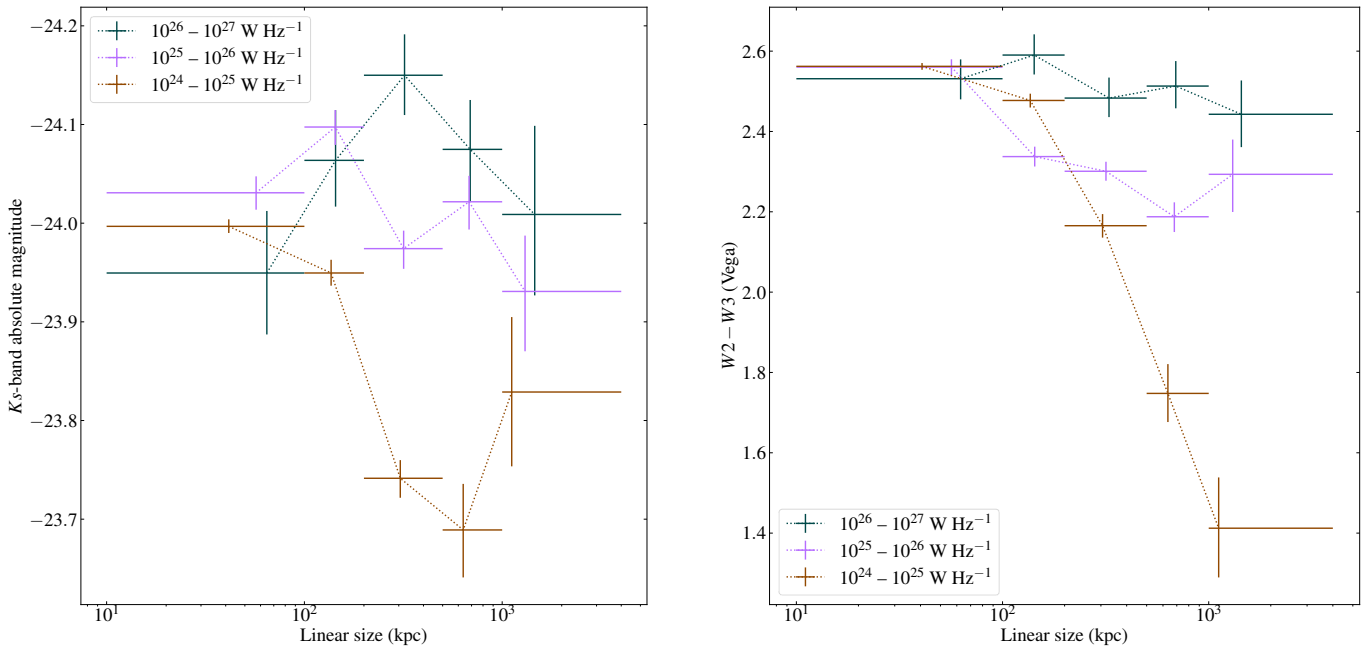

Fig. A.2. Mean properties of the host galaxies of RLAGN as a function of projected linear size. *Left panel*: K_s-band rest-frame magnitude. *Right panel*: WISE band 2 minus band 3 colours (Vega magnitudes). In both plots the vertical error bars represent 1σ errors derived from bootstrap, the horizontal bars represent the bin range, and the position of the central point represents the mean projected linear size within the bin. Dotted lines joining points in a particular sample are to guide the eye only.

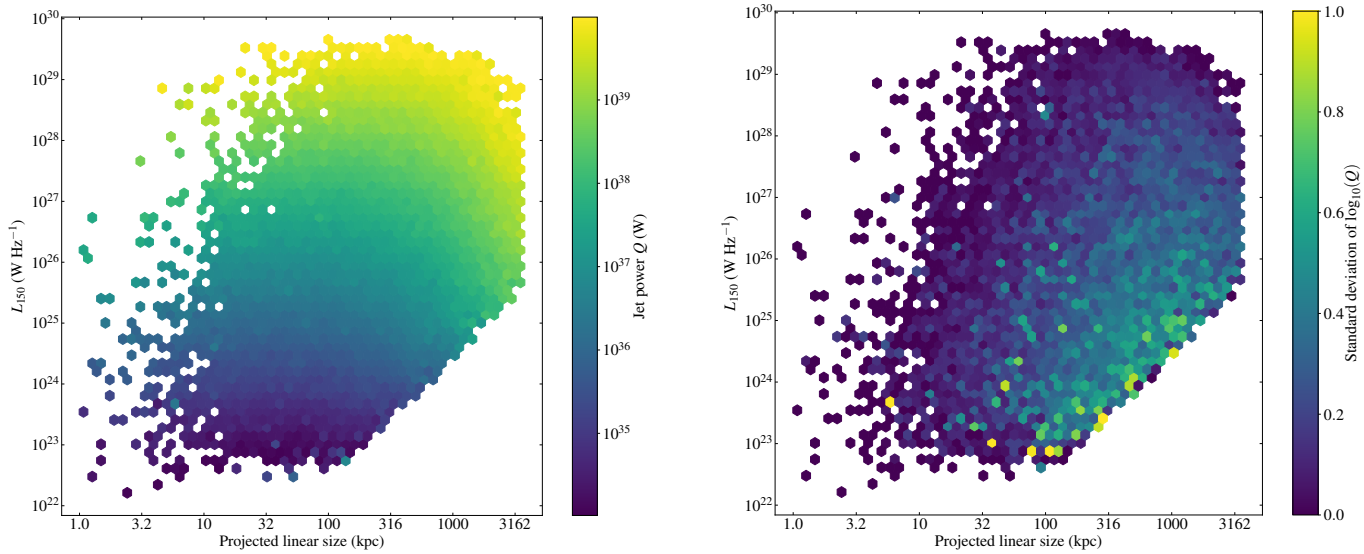

Fig. A.3. *Left panel*: mean jet power Q as a function of radio luminosity and projected linear size for the full simulated sample over all redshifts. *Right panel*: standard deviation of $\log_{10}(Q)$ for each bin in the left-hand panel, showing the scatter in the inference introduced by different environments, projection angles, ages, and redshifts. See the text for details of the modelling.

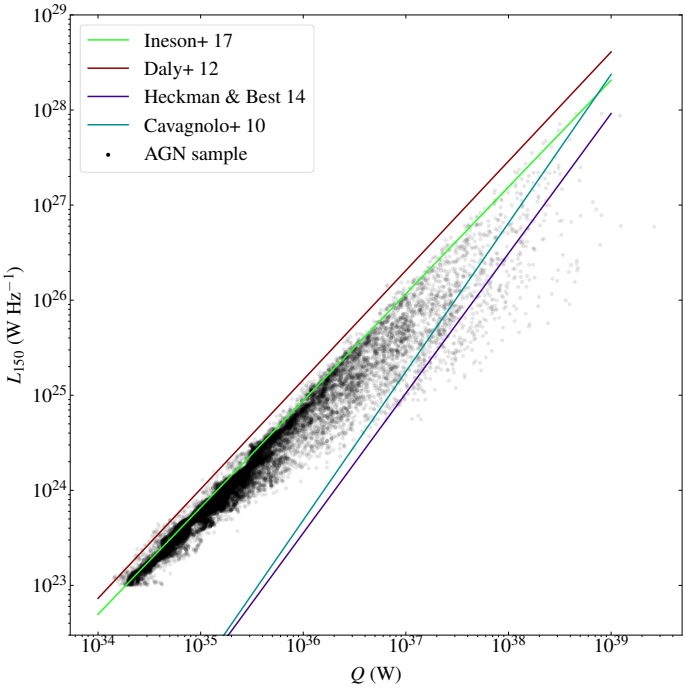

Fig. A.4. Radio luminosity plotted against the inferred jet power. The relations obtained by Cavagnolo et al. (2010), Daly et al. (2012), Heckman & Best (2014), Ineson et al. (2017) are also shown.

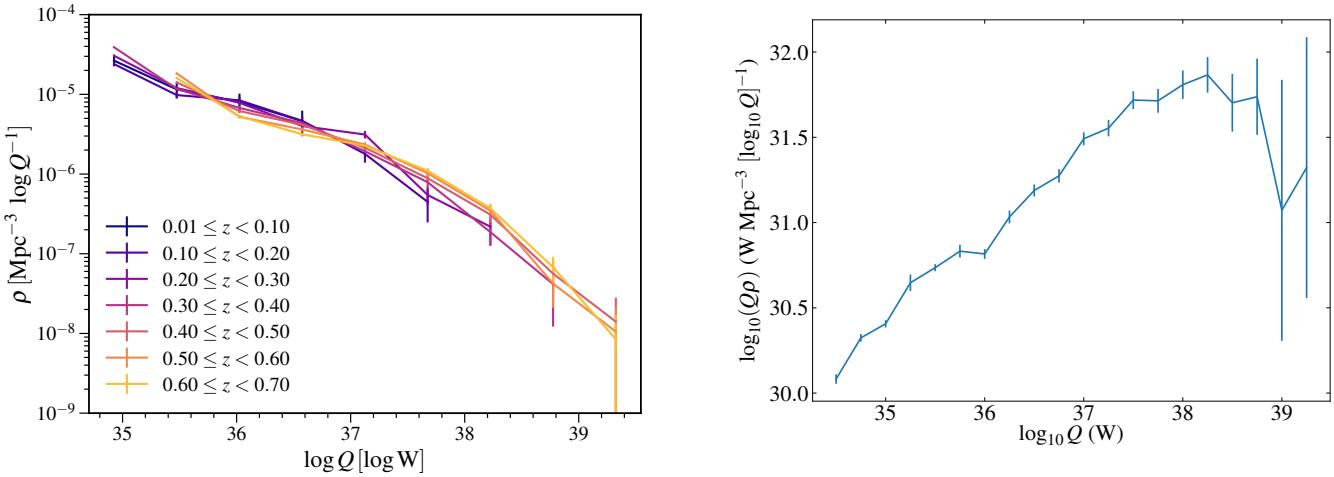

Fig. A.5. Jet kinetic luminosity function for the $z < 0.7$ LOFAR RLAGN sample. *Left panel*: kinetic luminosity function divided in bins of redshift. *Right panel*: full $z < 0.7$ luminosity function multiplied by Q to show the peak at $\sim 10^{38}$ W.