

# Measuring the giant radio galaxy length distribution with the LoTSS<sup>★</sup>

Martijn S. S. L. Oei<sup>1</sup>, Reinout J. van Weeren<sup>1</sup>, Aivin R. D. J. G. I. B. Gast<sup>2</sup>, Andrea Botteon<sup>1</sup>,  
Martin J. Hardcastle<sup>3</sup>, Pratik Dabhade<sup>4</sup>, Tim W. Shimwell<sup>1</sup>, Huub J. A. Röttgering<sup>1</sup>, and Alexander Drabent<sup>5</sup>

<sup>1</sup> Leiden Observatory, Leiden University, Niels Bohrweg 2, 2300 RA Leiden, The Netherlands  
e-mail: [oei@strw.leidenuniv.nl](mailto:oei@strw.leidenuniv.nl)

<sup>2</sup> Somerville College, University of Oxford, Woodstock Road, Oxford OX2 6HD, UK

<sup>3</sup> Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

<sup>4</sup> Observatoire de Paris, LERMA, Collège de France, CNRS, PSL University, Sorbonne University, 75014 Paris, France

<sup>5</sup> Thüringer Landessternwarte, Sternwarte 5, 07778 Tautenburg, Germany

Received 16 March 2022 / Accepted 24 October 2022

## ABSTRACT

**Context.** Many massive galaxies launch jets from the accretion disk of their central black hole, but only  $\sim 10^3$  instances are known in which the associated outflows form giant radio galaxies (GRGs, or giants): luminous structures of megaparsec extent that consist of atomic nuclei, relativistic electrons, and magnetic fields. Large samples are imperative to understanding the enigmatic growth of giants, and recent systematic searches in homogeneous surveys constitute a promising development. For the first time, it is possible to perform meaningful precision statistics with GRG lengths, but a framework to do so is missing.

**Aims.** We measured the intrinsic GRG length distribution by combining a novel statistical framework with a LOFAR Two-metre Sky Survey (LoTSS) sample of freshly discovered giants. In turn, this allowed us to answer an array of questions on giants. For example, we can now assess how rare a 5 Mpc giant is compared with one of 1 Mpc, and how much larger – given a projected length – the corresponding intrinsic length is expected to be. Notably, we can now also infer the GRG number density in the Local Universe.

**Methods.** We assumed the intrinsic GRG length distribution to be Paretian (i.e. of power-law form) with tail index  $\xi$ , and predicted the observed distribution by modelling projection and selection effects. To infer  $\xi$ , we also systematically searched the LoTSS for hitherto unknown giants and compiled the largest catalogue of giants to date.

**Results.** We show that if intrinsic GRG lengths are Pareto distributed with index  $\xi$ , then projected GRG lengths are also Pareto distributed with index  $\xi$ . Selection effects induce curvature in the observed projected GRG length distribution: angular length selection flattens it towards the lower end, while surface brightness selection steepens it towards the higher end. We explicitly derived a GRG's posterior over intrinsic lengths given its projected length, laying bare the  $\xi$  dependence. We also discovered 2060 giants within LoTSS DR2 pipeline products; our sample more than doubles the known population. Spectacular discoveries include the largest, second-largest, and fourth-largest GRG known ( $l_p = 5.1$  Mpc,  $l_p = 5.0$  Mpc, and  $l_p = 4.8$  Mpc), the largest GRG known hosted by a spiral galaxy ( $l_p = 2.5$  Mpc), and the largest secure GRG known beyond redshift 1 ( $l_p = 3.9$  Mpc). We increase the number of known giants whose angular length exceeds that of the Moon from 10 to 23; among the discoveries is the angularly largest known radio galaxy in the Northern Sky, which is also the angularly largest known GRG ( $\phi = 2^\circ$ ). Combining theory and data, we determined that intrinsic GRG lengths are well described by a Pareto distribution, and measured the index  $\xi = -3.5 \pm 0.5$ . This implies that, given its projected length, a GRG's intrinsic length is expected to be just 15% larger. Finally, we determined the comoving number density of giants in the Local Universe to be  $n_{\text{GRG}} = 5 \pm 2 (100 \text{ Mpc})^{-3}$ .

**Conclusions.** We developed a practical mathematical framework that elucidates the statistics of giant radio galaxy lengths. Through a LoTSS search, we also discovered 2060 new giants. By combining both advances, we determined that intrinsic GRG lengths are well described by a Pareto distribution with index  $\xi = -3.5 \pm 0.5$ , and that giants are truly rare in a cosmological sense: most clusters and filaments of the Cosmic Web are not currently home to a giant. Thus, our work yields new observational constraints for analytical models and simulations featuring radio galaxy growth.

**Key words.** galaxies: active – galaxies: jets – galaxies: kinematics and dynamics – radio continuum: galaxies – intergalactic medium

## 1. Introduction

When gas, dust, and stars accrete onto a supermassive black hole (SMBH) in the centre of a galaxy, collimated jets arise along the Kerr rotation axis that blast some of the infalling material into the intergalactic medium (IGM) (e.g., [Blandford & Rees 1974](#)). In this process, the ejecta dissolve into a relativistic plasma that

drags along a magnetic field and glows in synchrotron light. The resulting luminous structure is called a radio galaxy (RG); the central black hole that has generated it an active galactic nucleus (AGN).

It is increasingly clear that RGs and their AGN play an important role in galaxy evolution and cosmology. By heating the interstellar medium (ISM) or even expelling it from their host galaxies through galactic superwinds, AGN quench star formation (e.g., [Di Matteo et al. 2005](#)). [Beckmann et al. \(2017\)](#) have shown that AGN-induced star formation quenching is most

<sup>★</sup> Table F.1 is only available at the CDS via anonymous ftp to [cdsarc.cds.unistra.fr](https://cdsarc.cds.unistra.fr) (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/672/A163>

pronounced in massive galaxies. There is also compelling evidence that the accompanying RGs provide the energy necessary to stop (e.g., McNamara & Nulsen 2012; Yang et al. 2019) bremsstrahlung-mediated cooling flows (Fabian et al. 1984) in clusters of galaxies. In the absence of cooling flows, the intra-cluster medium (ICM) remains dilute and hot, and galaxies in the centres of clusters are denied infalling gas that could otherwise reignite star formation. Cosmological simulations that include this RG feedback indeed resolve (e.g., Croton et al. 2006) the overprediction of baryonic masses and luminosities of central cluster galaxies that early simulations found. Finally, RGs may be responsible for magnetising the IGM that pervades the filaments of the Cosmic Web (e.g., Vazza et al. 2017).

Despite the emerging picture that RGs trace quenched star formation, inhibit cooling flows, and magnetise filaments, our knowledge of them is far from complete. Concerning geometry, a major unknown is the exact connection between the morphology of RGs and the pressure field of the ambient IGM, especially in filaments and cluster outskirts. Another question is whether small and large RGs come from the same initial population, or whether their growth is driven by distinct physical processes. Finally, we do not know how large can RGs become, and, more generally, how many RGs there are of each length.

To test RG growth models that answer these and other questions, it is imperative to study the subpopulation of most spatially extreme RGs: the giant radio galaxies (GRGs). The defining feature of giants is that their proper lengths – when projected onto the celestial sphere – exceed some threshold  $l_{p,GRG}$ , which is canonically chosen as 0.7 Mpc or 1 Mpc. If  $l_{p,GRG} = 0.7$  Mpc, then the preceding literature describes a total of 1281 giants.

In recent years, several studies have successfully searched for giants in systematic, wide-area surveys such as the NRAO VLA Sky Survey (NVSS; Condon et al. 1998) and the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al. 2017). A combination of manual (i.e. visual) and automated searches (Solovoyov & Verkhodanov 2011, 2014; Amirkhanyan 2016; Proctor 2016; Dabhade et al. 2017, 2020a) in the NVSS yielded 313 new giants (24% of the aforementioned literature population). Meanwhile, Dabhade et al. (2020b) discovered 225 new giants (17% of this same population) in the LoTSS DR1 (Shimwell et al. 2019), whose survey footprint is 80 times smaller than NVSS’s. Such searches have the advantage of introducing almost homogeneous selection effects throughout the survey footprint, which can potentially be modelled and thus corrected for during any subsequent statistical inference.

In this work this idea comes to fruition, by conducting a precision analysis of the intrinsic giant radio galaxy length distribution. To do so, we require two ingredients. First, in Sect. 2, we develop a statistical framework that allows one to answer probabilistic questions regarding both large samples of giants and individual specimens. Then, in Sect. 3, we describe our LoTSS DR2 (Shimwell et al. 2022) GRG search campaign and the trove of previously unknown giants it has yielded; moreover, we describe the assemblage of the most complete GRG catalogue to date. In Sect. 4, combining theory and data, we infer the tail index parameter that describes the intrinsic GRG length distribution, which constrains future models and simulations aimed at understanding RG growth. In Sect. 5, we discuss caveats of the current work and give recommendations for future extensions, before we present conclusions in Sect. 6.

We assume a concordance inflationary  $\Lambda$  cold dark matter cosmology with parameters  $\mathfrak{M}$  from Planck Collaboration VI (2020); that is to say  $\mathfrak{M} = (h = 0.6766, \Omega_{BM,0} = 0.0490, \Omega_{M,0} = 0.3111, \Omega_{\Lambda,0} = 0.6889)$ , where  $H_0 := h \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . We

define the spectral index  $\alpha$  such that it relates to flux density  $F_\nu$  at frequency  $\nu$  as  $F_\nu \propto \nu^\alpha$ , and define giants using threshold  $l_{p,GRG} := 0.7 \text{ Mpc}$ . Regarding the terminology, we use ‘angular length’ where others use ‘largest angular size’ (LAS), and ‘projected proper length’ where others use ‘largest linear size’ (LLS)<sup>1</sup>.

## 2. Theory

To measure the intrinsic GRG length distribution, we must first establish a suitable statistical framework. In this section, we provide a summary of the theory developed in Appendix A. Following Occam’s razor, we construct a model with minimal assumptions that provides new insight into the GRG phenomenon and the detection biases inherent to systematic search campaigns.

### 2.1. Intrinsic proper length

Firstly, we assume that giants and non-giant RGs share a common length distribution<sup>2</sup>. In particular, because power laws are abundant in Nature, we assume that the intrinsic proper length random variable (RV)  $L$  has a Pareto distribution with tail index  $\xi < -1$  and support from  $l_{\min} > 0$  onwards. If an RV is Pareto distributed, then the relative occurrence of two possible outcomes equals their ratio raised to a power: the tail index  $\xi$ . In astrophysics, Pareto distributions describe the kinetic energies of freshly accelerated electrons in large-scale structure and supernova shocks (e.g., Kirk & Schneider 1987), the initial masses of main-sequence stars (e.g., Kroupa 2001), and the luminosities of gamma-ray bursts (e.g., Bloom et al. 2001), to name a few examples. Previous works (e.g., Andernach et al. 2021) have already hinted at the approximate validity of a Pareto distribution description for GRG lengths. By comparing our final model to observations, as discussed in Sect. 4 and visualised in Fig. 14, we demonstrate that this assumption is indeed a powerful approximation in the current case.

The probability density function (PDF)  $f_L : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  thus becomes

$$f_L(l) = \begin{cases} 0 & l \leq l_{\min}, \\ -\frac{(\xi+1)}{l_{\min}} \left(\frac{l}{l_{\min}}\right)^\xi & l > l_{\min}. \end{cases} \quad (1)$$

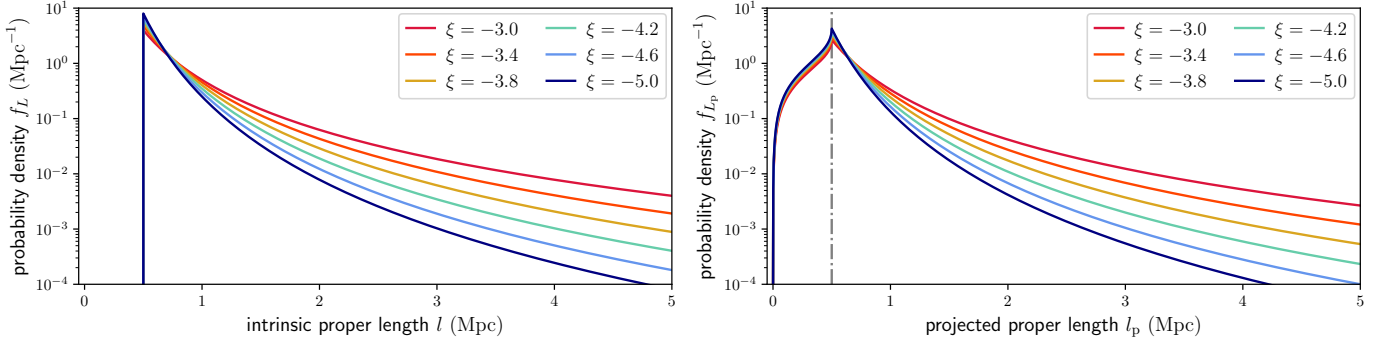
We refer the reader to Appendix A.1 for a derivation of this expression and a demonstration of its connection to the literature’s most common parametrisation.

### 2.2. Projected proper length

From the distribution of intrinsic lengths and the assumption of random radio galaxy orientations, we now derive the distribution of projected lengths. This distribution is more easily compared to observations, which usually lack inclination angle information.

<sup>1</sup> An object’s size can refer to either its one-, two-, or three-dimensional extent. We therefore consider ‘angular size’ to be ambiguous terminology; we propose that ‘angular length’ better captures one-dimensionality. Naturally, an object’s ‘length’ is understood to be its total one-dimensional extent, so that the qualifier ‘largest’ seems superfluous. We further remark that ‘length’ is synonymous with, but more succinct than, ‘linear size’. In cosmology, one must distinguish between proper and comoving lengths, especially when objects are not gravitationally bound – like GRG lobes. In this work, we consider two types of proper lengths: intrinsic proper lengths and projected proper lengths.

<sup>2</sup> Luckily, this assumption turns out to be irrelevant in the forthcoming GRG-only expressions, which are this work’s focus.



**Fig. 1.** PDFs of radio galaxy intrinsic proper lengths  $L$  and projected proper lengths  $L_p$ . If the intrinsic lengths  $L$  are Pareto distributed above some cut-off  $l_{\min}$ , then their projections on the sky  $L_p$  are also Pareto distributed above this cut-off. The tail indices are the same. We show the PDFs  $f_L$  (left) and  $f_{L_p}$  (right) for  $l_{\min} = 0.5$  Mpc,  $l_{\max} = \infty$  (see Appendix A) and various tail indices  $\xi$ . The support of  $f_L$  starts at  $l_{\min}$ , which is marked by the vertical grey line in the right panel.

### 2.2.1. Distribution for RGs

To model length and orientation, we consider a vector  $\mathbf{L} \in \mathbb{R}^3$  (of length  $L := \|\mathbf{L}\|_2$ ) for each RG. In accordance with the IAU Solar System convention for positive poles, the unit vector  $\hat{\mathbf{L}} \in \mathbb{S}^2$  marks the direction from which the central Kerr black hole is seen rotating in anticlockwise direction<sup>3</sup>. We define the inclination angle  $\Theta$  as the angle between  $\mathbf{L}$  and a vector parallel to the line of sight pointing towards the observer<sup>4</sup>. Observations that allow one to measure the orientation of the RG axis in 3D are time-intensive, and so usually only the RG length projected onto the plane of the sky is known.

Geometrically, we model RGs as line segments – as if they were ‘thin sticks’, with vanishing volumes – so that the projected proper length  $L_p$  relates to  $L$  and  $\Theta$  through

$$L_p = L \sin \Theta. \quad (2)$$

We assume that  $\hat{\mathbf{L}}$  is drawn from a uniform distribution on  $\mathbb{S}^2$ , so that the PDF  $f_{\Theta} : [0, \pi] \rightarrow \mathbb{R}_{\geq 0}$  becomes  $f_{\Theta}(\theta) = \frac{1}{2} \sin \theta$ . Since  $L$  and  $\Theta$  are independent, we find in Appendix A.2.1 that the PDF of  $L_p$ ,  $f_{L_p} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ , is

$$f_{L_p}(l_p) = \begin{cases} 0 & \text{if } l_p \leq 0, \\ \frac{-(\xi+1)}{l_{\min}} \frac{l_p}{l_{\min}} I\left(\xi-1, \frac{l_p}{l_{\min}}\right) & \text{if } 0 < l_p < l_{\min}, \\ \frac{(\xi+1)^2}{l_{\min}} \frac{\sqrt{\pi}}{4} \left(\frac{l_p}{l_{\min}}\right)^{\xi} \frac{\Gamma(-\frac{\xi}{2}-\frac{1}{2})}{\Gamma(-\frac{\xi}{2}+1)} & \text{if } l_p \geq l_{\min}, \end{cases} \quad (3)$$

where

$$I(a, b) := \int_1^{\infty} \frac{\eta^a d\eta}{\sqrt{\eta^2 - b^2}} \text{ for } a < 0, |b| < 1. \quad (4)$$

We note that for  $l_p > l_{\min}$ , the projected proper length has a Pareto distribution with the same tail index as the intrinsic proper length distribution. We compare  $f_L$  and  $f_{L_p}$  in Fig. 1.

### 2.2.2. Distribution for giants

For giants specifically (i.e. RGs such that  $L_p > l_{p,\text{GRG}}$ , where  $l_{p,\text{GRG}}$  is some constant threshold; in this work,  $l_{p,\text{GRG}} :=$

0.7 Mpc), the projected proper length distribution becomes a Pareto distribution with tail index  $\xi$  again:

$$f_{L_p | L_p > l_{p,\text{GRG}}}(l_p) = \begin{cases} 0 & \text{if } l_p \leq l_{p,\text{GRG}} \\ \frac{-(\xi+1)}{l_{p,\text{GRG}}} \left(\frac{l_p}{l_{p,\text{GRG}}}\right)^{\xi} & \text{if } l_p > l_{p,\text{GRG}}. \end{cases} \quad (5)$$

In other words, for giants, projection retains the Paretianity of lengths. A measurement of the tail index of the projected length distribution is immediately also a measurement of the tail index of the intrinsic length distribution.

The survival function, which gives the probability that a GRG has a projected proper length exceeding  $l_p$ , takes on a particularly simple form:

$$\mathbb{P}(L_p > l_p | L_p > l_{p,\text{GRG}}) = \left(\frac{l_p}{l_{p,\text{GRG}}}\right)^{\xi+1}. \quad (6)$$

The mean projected proper length of giants is the expectation value of  $L_p | L_p > l_{p,\text{GRG}}$ :

$$\mathbb{E}[L_p | L_p > l_{p,\text{GRG}}] = l_{p,\text{GRG}} \frac{\xi+1}{\xi+2}, \quad (7)$$

which is only defined when  $\xi < -2$ . For example, when  $\xi = -4$ ,  $\mathbb{E}[L_p | L_p > l_{p,\text{GRG}}] = \frac{3}{5} l_{p,\text{GRG}}$ , which becomes 1.05 Mpc for  $l_{p,\text{GRG}} := 0.7$  Mpc and 1.5 Mpc for  $l_{p,\text{GRG}} := 1$  Mpc.

Appendix A.2.2 provides a derivation for all three expressions.

### 2.3. Deprojection factor

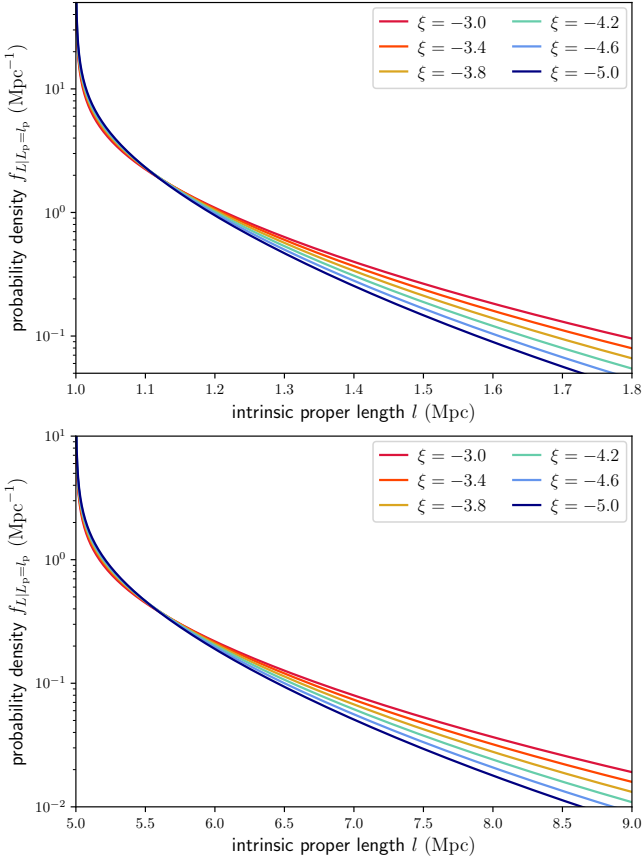
The deprojection factor,  $D := \frac{L}{L_p} = \sin^{-1} \Theta$ , quantifies how much larger intrinsic lengths are compared with projected lengths. The PDF of  $D$ ,  $f_D : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ , is

$$f_D(d) = \begin{cases} 0 & \text{if } d \leq 1; \\ \frac{1}{d^2 \sqrt{d^2 - 1}} & \text{if } d > 1. \end{cases} \quad (8)$$

The mean deprojection factor  $\mathbb{E}[D] = \frac{\pi}{2}$ . Deprojection factors can become arbitrarily large under the current model, because projected lengths can become arbitrarily small. As discussed in Sect. 5.2, this is not a very realistic set-up. In reality, an RG’s projected length is bounded from below by its lobes, which have a non-vanishing volume and thus extend along all three spatial dimensions. Upon projection, the projected length therefore cannot shrink beyond some lower limit that depends on the lobe geometry. In Appendix A.3, we show that by enriching the conventional stick-like geometry with spherical lobes, deprojection factors indeed become bounded.

<sup>3</sup>  $\mathbb{S}^2$  is the unit two-sphere.

<sup>4</sup> When  $\Theta = 0$ , the positive pole’s jet points towards us, and the black hole is seen rotating in anticlockwise direction; when  $\Theta = 90^\circ$ , the jets lie in the plane of the sky; when  $\Theta = 180^\circ$ , the positive pole’s jet points away from us, and the black hole rotates in clockwise direction.



**Fig. 2.** Posterior PDFs of intrinsic lengths for a given projected length  $L | L_p = l_p$ . If tail index  $\xi$  is known, then an RG's  $l_p$  fixes the probability distribution over its possible  $l$ . This distribution is strongly skewed, and the same for all  $l_p$  – save for horizontal translation and vertical scaling. We illustrate this point by showing posterior PDFs for giants with two different  $l_p$ . *Top:*  $l_p = 1$  Mpc. *Bottom:*  $l_p = 5$  Mpc. For  $\xi = -4$ , the posterior mean  $\mathbb{E}[L | L_p = l_p] = 1.13 l_p$  and the posterior standard deviation  $\sqrt{\mathbb{V}[L | L_p = l_p]} = 0.23 l_p$  (see Table 1).

#### 2.4. Intrinsic proper length posterior and its moments

Because an RG's intrinsic length is more physically informative than its projected length, we ideally obtain the former. In this subsection, we quantify what a measurement  $L_p = l_p$  already reveals about  $L$ .

We first note that the projected length bounds the intrinsic length from below. The intrinsic length can be much larger, however, but this is improbable for two reasons: large lengths are rarer than small lengths, although how drastic this effect is depends on  $\xi$ ; in addition, viewing directions with large inclination angles are uncommon. The best we can do is to construct a posterior distribution for  $L$  given  $L_p = l_p$ . This posterior has a concise analytic form. If  $l_p > l_{\min}$ , which is the relevant case for giants, the distribution of  $L | L_p = l_p$  is

$$f_{L|L_p=l_p}(l) = \begin{cases} 0 & \text{if } l \leq l_p \\ \frac{-\xi}{2^{1+\xi}\pi} \frac{\Gamma^2(-\frac{\xi}{2})}{\Gamma(-\xi)} \frac{1}{l_p} \frac{1}{\sqrt{\left(\frac{l}{l_p}\right)^2 - 1}} \left(\frac{l}{l_p}\right)^{\xi-1} & \text{if } l > l_p. \end{cases} \quad (9)$$

For  $l \gg l_p$ ,  $f_{L|L_p=l_p}(l) \propto \left(\frac{l}{l_p}\right)^{\xi-2}$ : the posterior PDF tends to a power law in  $l$  with index  $\xi - 2$ . In Fig. 2, we visualise the posterior PDF for several values of  $\xi$ .

**Table 1.** Intrinsic proper length posterior mean and standard deviation in multiples of projected proper length  $l_p$ , given for various tail indices  $\xi$ .

	$\mathbb{E}[L   L_p = l_p](\xi) (l_p)$	$\sqrt{\mathbb{V}[L   L_p = l_p](\xi) (l_p)}$
$\xi = -2$	$\frac{4}{\pi} \approx 1.27$	$\sqrt{2 - \frac{16}{\pi^2}} \approx 0.62$
$\xi = -3$	$\frac{3\pi}{8} \approx 1.18$	$\sqrt{\frac{3}{2} - \frac{9\pi^2}{64}} \approx 0.33$
$\xi = -4$	$\frac{32}{9\pi} \approx 1.13$	$\sqrt{\frac{4}{3} - \frac{1024}{81\pi^2}} \approx 0.23$
$\xi = -5$	$\frac{45\pi}{128} \approx 1.10$	$\sqrt{\frac{5}{4} - \frac{2025\pi^2}{16384}} \approx 0.17$

Clearly, to evaluate Eq. (9), one must choose  $l_p$  – however, the shape of the distribution is the same for all choices. We illustrate this by comparing the PDF for a comparatively small GRG ( $l_p = 1.0$  Mpc; top panel) to the PDF for Alcyoneus<sup>5</sup> ( $l_p = 5.0$  Mpc; bottom panel).

The posterior mean is

$$\mathbb{E}[L | L_p = l_p] = l_p \cdot \frac{-\xi}{2^{2\xi+3}\pi} \frac{\Gamma^4(-\frac{\xi}{2})}{\Gamma^2(-\xi)}; \quad (10)$$

the posterior variance is

$$\mathbb{V}[L | L_p = l_p] = l_p^2 \left( \frac{\xi}{\xi + 1} - \frac{\xi^2}{2^{4\xi+6}\pi^2} \frac{\Gamma^8(-\frac{\xi}{2})}{\Gamma^4(-\xi)} \right). \quad (11)$$

Both mean and standard deviation scale linearly in  $l_p$ : the projection effect is a multiplicative noise source. In Table 1, we provide explicit values for various  $\xi$ .

Higher moments exist up to order  $[-\xi]$ ; because the PDF  $f_{L | L_p=l_p}(l)$  is strongly skewed, such moments do further specify the distribution.

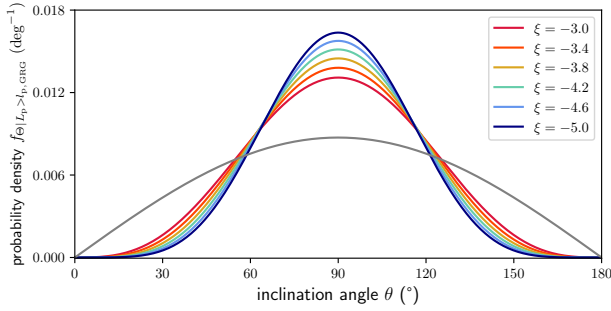
It is important to note that it is formally incorrect to statistically deproject RGs by drawing samples from deprojection factor  $D$  and multiplying them with some measurement  $L_p = l_p$ , or even more crudely, by multiplying the latter with  $\mathbb{E}[D]$ . The reason that renders such approaches invalid is that  $L_p = L \sin \Theta$  and  $D = \sin^{-1} \Theta$  are not independent RVs. We refer to Appendix A.4 for an explicit proof of this fact, and for derivations of this subsection's expressions.

#### 2.5. GRG inclination angle

Radio galaxies with jets that make a small angle with the plane of the sky are more likely to have a projected length exceeding  $l_{p,\text{GRG}}$  than those with jets that make a large angle with the plane of the sky. For this reason, the inclination angle distribution of giants is different from that of RGs: it is more peaked around  $\theta = 90^\circ$ . More precisely, the PDF  $f_{\Theta | L_p > l_{p,\text{GRG}}}: [0, \pi] \rightarrow \mathbb{R}_{\geq 0}$  of the GRG inclination angle  $\Theta | L_p > l_{p,\text{GRG}}$  has the general form

$$f_{\Theta | L_p > l_{p,\text{GRG}}}(\theta) = \frac{(1 - F_L(\frac{l_{p,\text{GRG}}}{\sin \theta})) f_{\Theta}(\theta)}{1 - F_{L_p}(l_{p,\text{GRG}})}. \quad (12)$$

<sup>5</sup> Alcyoneus is the projectively longest giant known to date (Oei et al. 2022a).



**Fig. 3.** PDFs of GRG inclination angles  $\Theta | L_p > l_{p,\text{GRG}}$  (colours) and RG inclination angles  $\Theta$  (grey). Giants more often have orientations close to the sky plane. The strength of this tendency is governed by  $\xi$ , with larger  $\xi$  meaning more dispersion.

Under our Pareto distribution assumption for  $L$ , this concretises to

$$f_{\Theta | L_p > l_{p,\text{GRG}}}(\theta) = \frac{2}{-(\xi + 1)\sqrt{\pi}} \frac{\Gamma(-\frac{\xi}{2} + 1)}{\Gamma(-\frac{\xi}{2} - \frac{1}{2})} \sin^{-\xi} \theta. \quad (13)$$

We note that  $f_{\Theta | L_p > l_{p,\text{GRG}}}(\theta) \propto \sin^{-\xi} \theta$ ; the factor in front serves only as a normalisation constant. The distribution is independent of the choice of  $l_{p,\text{GRG}}$  and depends on a single parameter:  $\xi$ . We visualise  $f_{\Theta | L_p > l_{p,\text{GRG}}}(\theta)$  in Fig. 3.

Appendix A.5 contains a brief derivation.

## 2.6. GRG angular length

The model predicts the distribution of GRG angular lengths in the Local Universe up to comoving distance  $r_{\text{max}}$ . The GRG angular length RV  $\Phi | L_p > l_{p,\text{GRG}}$  relates to the GRG projected proper length RV  $L_p | L_p > l_{p,\text{GRG}}$  and the comoving distance RV  $R$  as

$$\Phi | L_p > l_{p,\text{GRG}} = L_p | L_p > l_{p,\text{GRG}} \cdot \frac{1 + z(R)}{R}. \quad (14)$$

(We note that this relation is valid only in a flat Friedmann–Lemaître–Robertson–Walker (FLRW) universe.) We also assume that the GRG number density is constant in the Local Universe. The PDF of  $\Phi | L_p > l_{p,\text{GRG}}$  has useful analytic forms under two different idealisations.

In a Euclidean universe,  $z(R) = 0$ , and the minimal GRG angular length  $\phi_{\text{GRG}} = \frac{l_{p,\text{GRG}}}{r_{\text{max}}}$ . Then

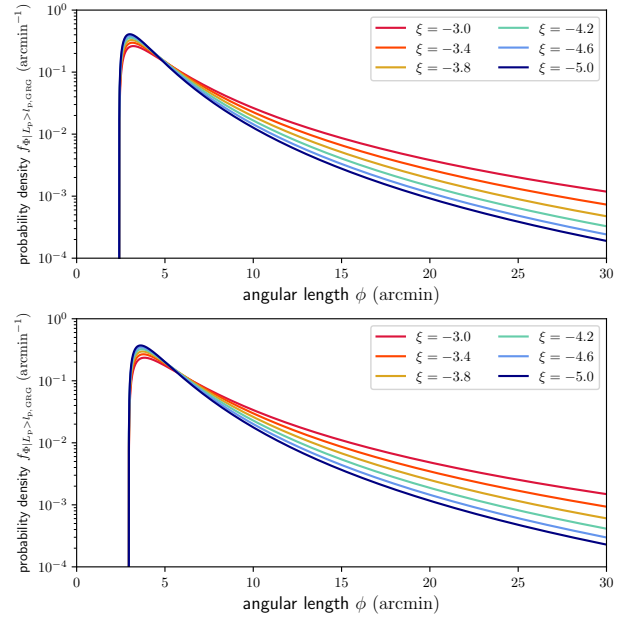
$$f_{\Phi | L_p > l_{p,\text{GRG}}}(\phi) = \begin{cases} 0 & \text{if } \phi \leq \phi_{\text{GRG}} \\ -3 \frac{\xi+1}{\xi+4} \cdot \frac{1}{\phi_{\text{GRG}}} \left( \left( \frac{\phi}{\phi_{\text{GRG}}} \right)^\xi - \left( \frac{\phi}{\phi_{\text{GRG}}} \right)^{-4} \right) & \text{if } \phi > \phi_{\text{GRG}}, \end{cases} \quad (15)$$

which is valid as long as  $\xi \neq -4$ .

In an expanding universe at low redshifts, the Hubble–Lemaître law  $z(R) \approx \frac{R}{d_H}$  holds; the Hubble distance  $d_H := \frac{c}{H_0}$ . In this case,

$$f_{\Phi | L_p > l_{p,\text{GRG}}}(\phi) = \begin{cases} 0 & \text{if } \phi \leq \phi_{\text{GRG}} + \frac{l_{p,\text{GRG}}}{d_H} \\ -\frac{3(\xi+1)\phi^\xi}{r_{\text{max}}^3 \frac{\xi+1}{p_{\text{GRG}}} \int_{\frac{1}{r_{\text{max}} + \frac{1}{d_H}}^{\frac{\phi}{r_{\text{max}} + \frac{1}{d_H}}} \frac{dk}{k^{\xi+1} \left( k - \frac{1}{d_H} \right)^4}} & \text{if } \phi > \phi_{\text{GRG}} + \frac{l_{p,\text{GRG}}}{d_H}. \end{cases} \quad (16)$$

Figure 4 shows GRG angular length PDFs under both idealisations.



**Fig. 4.** PDFs of GRG angular lengths  $\Phi | L_p > l_{p,\text{GRG}}$ . We fix  $r_{\text{max}} = 1$  Gpc (and  $l_{p,\text{GRG}} = 0.7$  Mpc), and vary  $\xi$ . *Top*: Euclidean universe. *Bottom*: expanding universe at low redshifts.

The PDFs undergo a minor shift upon changing universe type but are otherwise similar. For most current-day applications, it will therefore be unnecessary to calculate an even more refined version of  $f_{\Phi | L_p > l_{p,\text{GRG}}}(\phi)$ . Appendix A.6 contains derivations and details.

## 2.7. Maximum likelihood estimation of the tail index

The GRG projected proper length distribution features just one parameter of physical interest: the tail index  $\xi$ . If observational selection effects are negligible, one can directly use maximum likelihood estimation (MLE) on GRG data to infer  $\xi$ . In particular, we consider a set of projected lengths  $\{L_{p,1}, \dots, L_{p,N}\} \sim L_p | L_p > l_{p,\text{GRG}}$  from  $N$  giants. Appendix A.7 shows that the maximum likelihood estimate of  $\xi$  is the RV  $\xi_{\text{MLE}}$ , given by

$$\xi_{\text{MLE}} = -\frac{N}{\sum_{i=1}^N \ln \frac{L_{p,i}}{l_{p,\text{GRG}}}} - 1. \quad (17)$$

## 2.8. Observed projected proper length

### 2.8.1. General considerations

In the preceding theory, we have ignored observational selection effects that favour some projected proper lengths over others. In practice, several such effects occur; the importance of each varies per survey and (G)RG search campaign within it. One of them is the bias against physically long RGs that the interferometer’s largest detectable angular scale can induce<sup>6</sup>. As a result, the projected proper length of an observed RG might not be adequately

<sup>6</sup> For example, the Faint Images of the Radio Sky at Twenty Centimeters (FIRST) survey used the Very Large Array (VLA) in B-configuration, leading it to detect angular scales of at most two arcminutes. By contrast, the largest angular scale of the LoTSS – the survey relevant to this work – is about a degree. (For the 6'' and 20'' resolutions, the shortest baseline is 100 metres; for the 60'' and 90'' resolutions, the shortest baseline is 68 metres.) As virtually all giants are of subdegree angular length, we need not consider this bias in our case.

modelled through RV  $L_p$ . Instead, we must introduce a new RV  $L_{p,\text{obs}}$ .

We define the completeness  $C : \mathbb{R}_{>0} \times \mathbb{R}_{>0} \rightarrow [0, 1]$  at  $(l_p, z_{\text{max}})$  to be the fraction of all RGs with projected proper length  $l_p$  in the cosmological volume up to  $z_{\text{max}}$  that is detected in a particular RG search campaign. Then, assuming that the distribution of  $L_p$  does not evolve with redshift between  $z = z_{\text{max}}$  and  $z = 0$  (i.e.  $\xi$  remains constant),

$$C(l_p, z_{\text{max}}) = \frac{\int_0^{z_{\text{max}}} p_{\text{obs}}(l_p, z) r^2(z) E^{-1}(z) dz}{\int_0^{z_{\text{max}}} r^2(z) E^{-1}(z) dz}, \quad (18)$$

where  $p_{\text{obs}}(l_p, z)$  is the probability that an RG of projected proper length  $l_p$  at cosmological redshift  $z$  is detected through the campaign, and  $r(z)$  is the comoving radial distance at cosmological redshift  $z$ . In a flat FLRW universe, the dimensionless Hubble parameter  $E$  is

$$E(z) := \frac{H(z)}{H_0} = \sqrt{\Omega_{R,0}(1+z)^4 + \Omega_{M,0}(1+z)^3 + \Omega_{\Lambda,0}}. \quad (19)$$

The PDF of the observed projected proper length RV  $L_{p,\text{obs}}$  becomes

$$f_{L_{p,\text{obs}}}(l_p) = \frac{C(l_p) f_{L_p}(l_p)}{\int_0^{\infty} C(l'_p) f_{L_p}(l'_p) dl'_p}, \quad (20)$$

where we suppress the  $z_{\text{max}}$ -dependence for succinctness. We note that multiplying  $p_{\text{obs}}(l_p, z)$  with an  $l_p$ - and  $z$ -independent factor affects the completeness  $C(l_p, z_{\text{max}})$ , but cancels in Eq. (20);  $f_{L_{p,\text{obs}}}$  will be independent of it. Finally, the PDF of the GRG observed projected proper length RV  $L_{p,\text{obs}} | L_{p,\text{obs}} > l_{p,\text{GRG}}$  is

$$f_{L_{p,\text{obs}} | L_{p,\text{obs}} > l_{p,\text{GRG}}}(l_p) = \begin{cases} 0 & \text{if } l_p \leq l_{p,\text{GRG}} \\ \frac{C(l_p) f_{L_p}(l_p)}{\int_{l_{p,\text{GRG}}}^{\infty} C(l'_p) f_{L_p}(l'_p) dl'_p} & \text{if } l_p > l_{p,\text{GRG}}. \end{cases} \quad (21)$$

We derive these expressions in Appendix A.8.1.

### 2.8.2. Fuzzy angular length threshold

We provide a concrete example of an important observational selection effect in visual searches for GRG candidates in survey images. To cope with the sheer number of detectable RGs in modern surveys like the LoTSS, a natural criterion is to only add sources to a candidate list if they appear – by eye – to have an angular length larger than some threshold. However, it is hard to precisely assess the angular length of a candidate before actually measuring it; sometimes, a candidate with a smaller angular length than the threshold will feature in the list, while some candidates with a larger angular length than the threshold will not. This leads to the notion of a ‘fuzzy angular length threshold’, where the probability that an RG with angular length  $\phi$  is observed through the visual search increases (e.g., linearly) from 0 to 1 between  $\phi_{\text{min}}$  and  $\phi_{\text{max}}$ :

$$p_{\text{obs,AL}}(l_p, z) = \min \left\{ \max \left\{ \frac{\phi(l_p, z) - \phi_{\text{min}}}{\phi_{\text{max}} - \phi_{\text{min}}}, 0 \right\}, 1 \right\}, \quad (22)$$

$$\phi(l_p, z) = \frac{l_p(1+z)}{r(z)}. \quad (23)$$

See the top row of Fig. 5 for several examples of associated completeness curves  $C(l_p)$  and observed projected proper length PDFs. See Appendix A.8.2 for additional information.

### 2.8.3. Surface brightness limitations

Another important observational selection effect is due to a survey’s finite noise level. The noise determines the surface brightness threshold  $b_{v,\text{th}}$  (typically comparable to the noise level itself) below which radio galaxy features remain visually undetected.

*Fanaroff–Riley class II.* We model the lobe surface brightness RV  $B_v$  at the central observing frequency  $\nu_{\text{obs}}$  as

$$B_v = \frac{b_{v,\text{ref}} \cdot S}{(1+Z)^{3-\alpha}} \left( \frac{L}{l_{\text{ref}}} \right)^\zeta, \quad (24)$$

where  $b_{v,\text{ref}}$  is the median surface brightness of RGs of intrinsic proper length  $l_{\text{ref}}$  at cosmological redshift  $z = 0$  and frequency  $\nu_{\text{obs}}$ , and  $S$  is a lognormally distributed RV with median 1 and dispersion parameter  $\sigma_{\text{ref}}$  that captures the variability in surface brightness among this population of RGs. The denominator models the fact that surface brightness is not conserved with distance in an FLRW universe;  $Z$  is the cosmological redshift RV up to  $z = z_{\text{max}}$  and  $\alpha$  is the typical lobe spectral index. The exponent  $\zeta < 0$  characterises the scaling between intrinsic proper length and surface brightness. (If RGs are self-similar, so that morphology does not predict length, one finds  $\zeta = -2$ .) For this selection effect, the observing probability is

$$p_{\text{obs,SB}}(l_p, z) = \int_{s_{\text{min}}}^{\infty} \sqrt{1 - \left( \frac{s_{\text{min}}}{s} \right)^{-\frac{2}{\zeta}}} f_S(s) ds, \quad (25)$$

$$s_{\text{min}} = \frac{b_{v,\text{th}}}{b_{v,\text{ref}}} \left( \frac{l_p}{l_{\text{ref}}} \right)^{-\zeta} (1+z)^{3-\alpha}, \quad (26)$$

$$f_S(s) = \frac{1}{\sqrt{2\pi}\sigma_{\text{ref}}s} \exp\left(-\frac{\ln^2 s}{2\sigma_{\text{ref}}^2}\right). \quad (27)$$

We note that  $p_{\text{obs,SB}}$  does not depend on  $b_{v,\text{th}}$  or  $b_{v,\text{ref}}$  separately, but on their ratio only. See the bottom row of Fig. 5 for several examples of associated completeness curves  $C(l_p)$  and observed projected proper length PDFs.

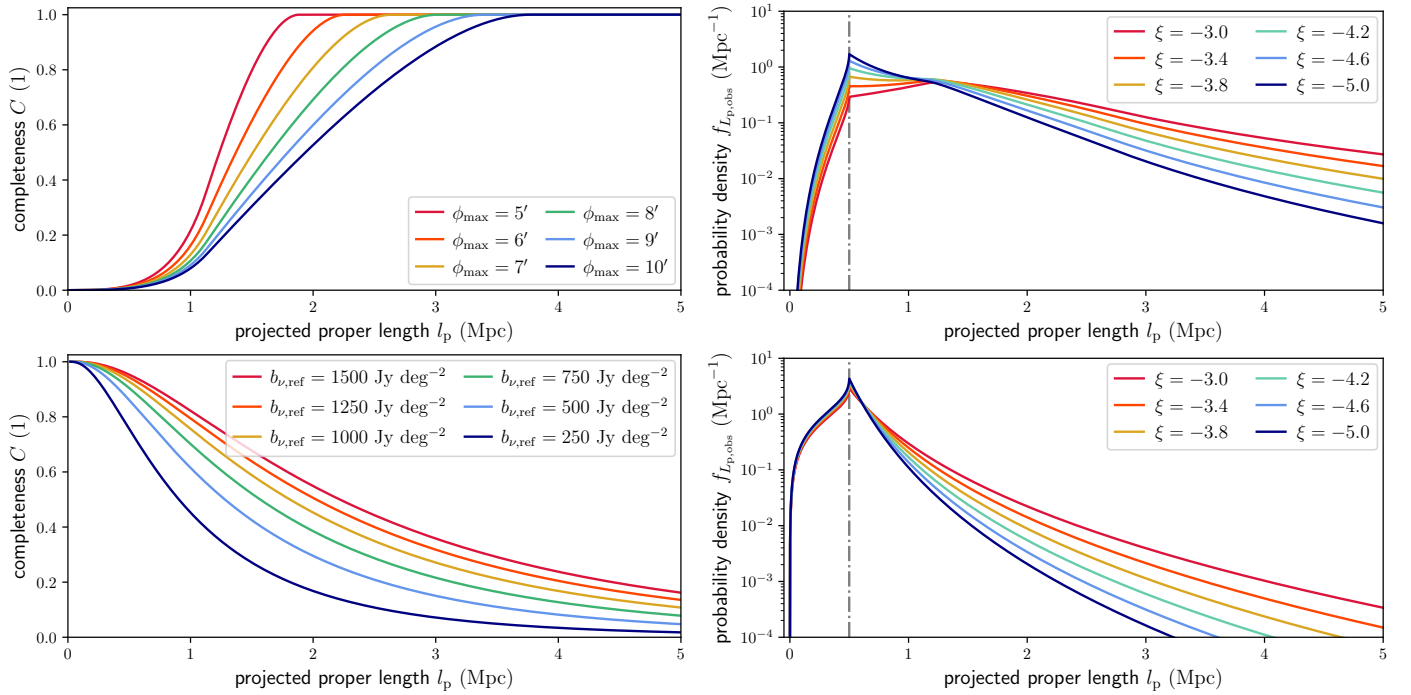
*Fanaroff–Riley class I.* For FRI RGs, the assumption of a constant surface brightness beyond the core is inaccurate. The simplest correction in which FRI RGs retain a well-defined notion of length assumes a linearly decreasing surface brightness profile, which peaks at the core and goes to zero at the RG’s two endpoints. (A power-law profile does not work: in such case, the surface brightness only goes asymptotically to zero – but never actually reaches it.) In this case, we define RV  $B_v$  to be the mean surface brightness along an RG’s jets, which can be regarded as a typical value for that RG. As  $B_v$  again obeys Eq. (24), we find that the formulaic structure of  $p_{\text{obs,SB}}(l_p, z)$  is identical for FRI and FRII giants, except that FRI giants require a change

$$b_{v,\text{th}} \rightarrow \frac{b_{v,\text{th}}}{2 \left( 1 - \frac{l_{p,\text{GRG}}}{l_p} \right)}, \quad (28)$$

which affects  $p_{\text{obs,SB}}(l_p, z)$  through  $s_{\text{min}}$ . There is no change for  $l_p = 2l_{p,\text{GRG}}$ . Although the formulaic structure might be the same, the best-fit parameters can differ. For example, it is possible that  $\sigma_{\text{ref,FRI}} \neq \sigma_{\text{ref,FRII}}$  or  $\zeta_{\text{FRI}} \neq \zeta_{\text{FRII}}$ . See Appendix A.8.3 for derivations and numerical implementation considerations.

To include both aforementioned selection effects at the same time, a natural approximation is to assume that the observing probability simply factorises:

$$p_{\text{obs}}(l_p, z) \approx p_{\text{obs,AL}}(l_p, z) \cdot p_{\text{obs,SB}}(l_p, z). \quad (29)$$



**Fig. 5.** Completeness functions (*left column*) and PDFs of observed projected proper lengths  $L_{p,obs}$  (*right column*). Selection effects leave imprints on the distribution of radio galaxies'  $L_{p,obs}$ . In the *top row*, we show how imposing an angular length threshold in a GRG search campaign leads to incompleteness (*left*), which causes the PDF  $f_{L_{p,obs}}$  (*right*) to differ from  $f_{L_p}$ . We assume RGs with angular length  $\phi < \phi_{\min}$  have probability 0 to be included in a sample, whilst RGs with angular length  $\phi > \phi_{\max}$  have probability 1. The inclusion probability is assumed to increase linearly between  $\phi_{\min}$  and  $\phi_{\max}$ . In the left panel, we fix  $\phi_{\min} = 3'$  and vary  $\phi_{\max}$ ; in the right panel, we also fix  $\phi_{\max} = 7'$ . In the *bottom row*, we show how a survey's surface brightness limitations lead to incompleteness (*left*), which causes the PDF  $f_{L_{p,obs}}$  (*right*) to differ from  $f_{L_p}$ . We assume that lobe surface brightnesses are lognormally distributed; we parametrise the distribution for RGs of intrinsic length  $l_{ref} = 0.7$  Mpc at  $z = 0$  observed at  $\nu_{obs} = 144$  MHz with a median  $b_{\nu,ref}$  and dispersion parameter  $\sigma_{ref}$ . In the *left panel*, we fix  $\sigma_{ref} = 1.5$  and vary  $b_{\nu,ref}$ ; in the *right panel*, we also fix  $b_{\nu,ref} = 1000$  Jy deg $^{-2}$ . We assume a lobe spectral index  $\alpha = -1$ , a surface brightness detection threshold  $b_{\nu,th} = 25$  Jy deg $^{-2}$ , and self-similar growth:  $\zeta = -2$ . For both selection effects, we consider RGs up to cosmological redshift  $z_{\max} = 0.5$  only.

## 2.9. GRG number density

A central question in the field of radio galaxies is: how intrinsically rare are giants? By counting giants in a search campaign with well-understood selection effects, we can give an answer. More precisely, one can estimate the comoving number density of giants in the Local Universe,  $n_{GRG}$ , through the number of observed giants up to cosmological redshift  $z_{\max}$  in a uniformly searched region of sky of solid angle  $\Omega$ ,  $N_{GRG,obs}(z_{\max}, \Omega)$ . Then, under the standard assumption  $l_{p,GRG} > l_{\min}$ , Appendix A.9 shows that

$$n_{GRG} = \frac{I_{p,GRG}^{\xi+1}}{-(\xi+1)} \cdot \frac{H_0}{c} \cdot \frac{4\pi N_{GRG,obs}(z_{\max}, \Omega)}{\int_{l_{p,GRG}}^{\infty} I_p^{\xi} \int_0^{z_{\max}} p_{obs}(l_p, z) 4\pi r^2(z) E^{-1}(z) dz dl_p}. \quad (30)$$

Although the appropriate  $p_{obs}(l_p, z)$  varies per search campaign, it is always possible to bound  $p_{obs}(l_p, z)$  from above – for example by 1. In such case, Eq. (30) bounds  $n_{GRG}$  from below.

## 2.10. GRG lobe volume-filling fraction

Because giants attain cosmological lengths, they might contribute to the energisation and magnetisation of Cosmic Web filaments in regions that smaller radio galaxies cannot reach. A key statistic that measures the enrichment of the Cosmic Web by giants is the volume-filling fraction (VFF) of their lobes. Assum-

ing that lobes do not grow along with the expansion of the Universe, the proper VFF changes over cosmic time:  $VFF_{GRG}(z) = VFF_{GRG}(z=0) \cdot (1+z)^3$ , where

$$\begin{aligned} VFF_{GRG}(z=0) &:= n_{GRG} \cdot \mathbb{E}[V | L_p > l_{p,GRG}] \\ &= n_{GRG} \cdot \mathbb{E}[\Upsilon \cdot L^3 | L_p > l_{p,GRG}], \end{aligned} \quad (31)$$

where  $V$  is the combined volume of the lobes and  $\Upsilon := \frac{V}{L^3}$  is a dimensionless RV that captures the diversity in radio galaxy lobe shapes. We find under self-similar growth

$$VFF_{GRG}(z=0) = n_{GRG} \cdot \mathbb{E}[\Upsilon] \cdot \mathbb{E}[L^3 | L_p > l_{p,GRG}]. \quad (32)$$

$\mathbb{E}[\Upsilon]$  can be estimated from observations, but one must be wary of selection effects. Unfortunately,  $\mathbb{E}[L^3 | L_p > l_{p,GRG}]$  does not exist for  $\xi \geq -4$ , which is the regime supported by observations. A useful lower bound then is

$$VFF_{GRG}(z=0) > n_{GRG} \cdot \mathbb{E}[\Upsilon] \cdot \mathbb{E}^3[L | L_p > l_{p,GRG}], \quad (33)$$

where

$$\mathbb{E}[L | L_p > l_{p,GRG}] = l_{p,GRG} \frac{\Gamma(-\frac{\xi}{2} - 1) \Gamma(-\frac{\xi}{2} + 1)}{\Gamma(-\frac{\xi}{2} - \frac{1}{2}) \Gamma(-\frac{\xi}{2} + \frac{1}{2})}. \quad (34)$$

This is the mean intrinsic proper length of giants, which is only defined when  $\xi < -2$ .<sup>7</sup>

<sup>7</sup> For example, when  $\xi = -3$ ,  $\mathbb{E}[L | L_p > l_{p,GRG}] = \frac{3\pi}{4} l_{p,GRG}$ , which becomes 1.65 Mpc for  $l_{p,GRG} := 0.7$  Mpc and 2.36 Mpc for  $l_{p,GRG} := 1$  Mpc. When  $\xi = -4$ ,  $\mathbb{E}[L | L_p > l_{p,GRG}] = \frac{16}{3\pi} l_{p,GRG}$ , which becomes 1.19 Mpc for  $l_{p,GRG} := 0.7$  Mpc and 1.70 Mpc for  $l_{p,GRG} := 1$  Mpc.

An alternative is to deviate slightly from our Pareto ansatz and truncate the GRG projected proper length distribution at some  $l_{p,\max}$ ; then

$$\text{VFF}_{\text{GRG}}(z=0) = n_{\text{GRG}} \cdot \mathbb{E}[\Upsilon_p] \cdot \frac{\xi + 1}{\xi + 4} \cdot \frac{l_{p,\max}^{\xi+4} - l_{p,\text{GRG}}^{\xi+4}}{l_{p,\max}^{\xi+1} - l_{p,\text{GRG}}^{\xi+1}}, \quad (35)$$

where  $\Upsilon_p := \frac{V}{l_p^3}$ . See Appendix A.10 for a derivation and further details.

### 2.11. Unification model constraints from quasar and non-quasar giants

The unification model and its extensions (e.g., [Hardcastle & Croston 2020](#)) form an elegant family of hypotheses that aim to explain the observational diversity of active galaxies. It posits that active galaxies with quasars differ from those without quasars primarily because of differences in orientation of the dusty tori surrounding SMBHs. In particular, the central idea is that a quasar appears brighter to the observer than a non-quasar AGN because the axis of its dusty torus happens to be virtually parallel to the line of sight. As such, only quasars would offer an unobscured view of the luminous accretion disk surrounding the SMBH, whilst also beaming relativistic jet emission towards the observer. Using our statistical framework, we predict the general ramifications of the basic unification model on a GRG sample.

The basic unification model suggests to divide the radio galaxy population in two, distinguishing between RGs generated by AGN with quasar appearance (quasar RGs) and RGs generated by AGN without quasar appearance (non-quasar RGs). We assume that quasar RGs have inclination angles  $\theta \leq \theta_{\max}$  or  $\theta \geq 180^\circ - \theta_{\max}$  whilst non-quasar RGs have  $\theta_{\max} < \theta < 180^\circ - \theta_{\max}$ <sup>8</sup>. If quasar RGs are more closely aligned with the line of sight than non-quasar RGs but are otherwise similar, fewer of them will satisfy  $l_p \geq l_{p,\text{GRG}}$  and thus be classified as giants. Therefore, the quasar GRG fraction  $f_Q$  – the fraction of quasar giants in an actual GRG sample – constrains  $\theta_{\max}$ . We model  $f_Q$  as an RV:

$$f_Q := \frac{N_Q}{N}; \quad N_Q \sim \text{Binom}(N, p_Q), \quad (36)$$

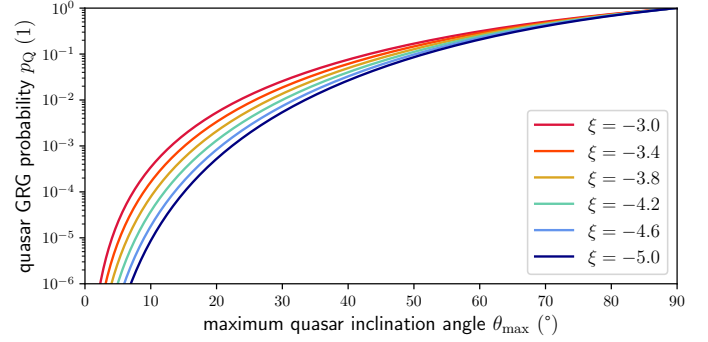
where the RV  $N_Q$  is the number of quasar giants in the sample, the constant  $N$  is the total number of giants in the sample and the parameter  $p_Q$  is the quasar GRG probability. Our framework predicts

$$p_Q := \frac{\mathbb{P}(L_{p,\text{obs}} \geq l_{p,\text{GRG}}, \sin \Theta \leq \sin \theta_{\max})}{\mathbb{P}(L_{p,\text{obs}} \geq l_{p,\text{GRG}})} = \frac{4\Gamma(-\frac{\xi}{2} + 1)}{-(\xi + 1)\sqrt{\pi}\Gamma(-\frac{\xi}{2} - \frac{1}{2})} \int_0^{\sin \theta_{\max}} \frac{x^{-\xi} dx}{\sqrt{1-x^2}}. \quad (37)$$

See Appendix A.11 for a derivation. Interestingly, as long as quasar giants and non-quasar giants are subject to the same selection effects, these selection effects do not affect  $p_Q$ .

Figure 6 shows that, for all relevant  $\xi$ ,  $p_Q$  is a steeply and monotonically increasing function of  $\theta_{\max}$ . Thus, knowing  $\xi$ , one can use an empirical  $f_Q$  to determine  $p_Q$  and in turn  $\theta_{\max}$ .

<sup>8</sup> Geometrically,  $\theta_{\max}$  represents the opening angle of the two coaxial conical gaps in the dusty torus of the AGN.



**Fig. 6.** Probability  $p_Q$  that an observed giant is a quasar giant under the unification model. Under this model, giants generated by AGN with quasar appearance (quasar giants) have inclination angles  $\theta \leq \theta_{\max}$  or  $\theta \geq 180^\circ - \theta_{\max}$  and giants generated by AGN without quasar appearance (non-quasar giants) have intermediate  $\theta$ . As long as quasar giants and non-quasar giants are subject to the same selection effects, these selection effects do not affect  $p_Q$ . Instead, in such case,  $p_Q$  only depends on the maximum quasar inclination angle  $\theta_{\max}$  and the tail index  $\xi$ .

Does one expect quasar giants to have a different distribution for  $L_{p,\text{obs}} | L_{p,\text{obs}} \geq l_{p,\text{GRG}}$  than non-quasar giants? Interestingly, our framework allows us to prove that the inclination angle differences between the two classes affect their relative rarity, but not their observed projected proper length distribution. Under the basic unification model, quasar giants and non-quasar giants obey the same  $L_{p,\text{obs}} | L_{p,\text{obs}} \geq l_{p,\text{GRG}}$  if they are subject to the same selection effects.

### 2.12. Extreme giants in a sample

The model can predict the occurrence of giants with extreme projected proper lengths in a sample of cardinality  $N$ . The probability that an observed GRG has a projected proper length exceeding  $l_p$  is

$$p_{>l_p} := \mathbb{P}(L_{p,\text{obs}} > l_p | L_{p,\text{obs}} > l_{p,\text{GRG}}) = \frac{1 - F_{L_{p,\text{obs}}}(l_p)}{1 - F_{L_{p,\text{obs}}}(l_{p,\text{GRG}})}, \quad (38)$$

so that the number of observed giants with a projected proper length exceeding  $l_p$  is  $N_{>l_p} \sim \text{Binom}(N, p_{>l_p})$ . Its expectation is  $\mathbb{E}[N_{>l_p}] = N \cdot p_{>l_p}$ . Furthermore, the probability that the sample contains at least one giant with projected proper length  $l_p$  or higher is

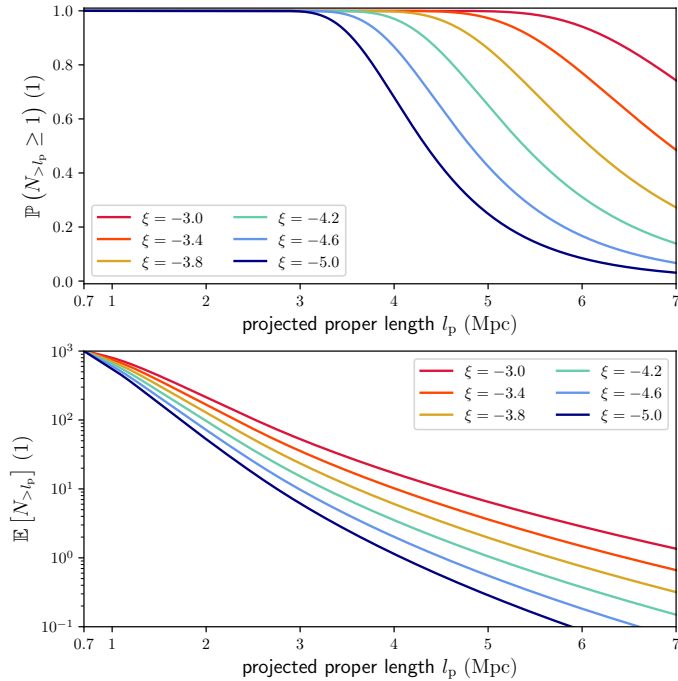
$$\mathbb{P}(N_{>l_p} \geq 1) = 1 - (1 - p_{>l_p})^N. \quad (39)$$

See Appendix A.12 for details. Figure 7 shows  $\mathbb{E}[N_{>l_p}]$  and  $\mathbb{P}(N_{>l_p} \geq 1)$  for various  $\xi$ . As an example, the case  $\xi = -3.0$  predicts that a sample of  $N = 1000$  giants with redshifts below  $z_{\max} = 0.5$  should contain almost ten giants of  $l_p > 5$  Mpc, and still several of  $l_p > 6$  Mpc. Such predictions are useful as they can be directly compared to elementary sample statistics.

## 3. Sample compilation and properties

To measure the intrinsic GRG length distribution, we also require a large sample of giants collected from a single survey through a systematic approach. This ensures approximately homogeneous selection effects, which we can correct for in subsequent analysis using the statistical framework of Sect. 2.





**Fig. 7.** Predictions of the existence and expected number of giants that exceed projected length  $l_p$  in a sample of cardinality  $N$ , as functions of  $l_p$ . Both tail index  $\xi$  and selection effect parameters affect these predictions. We consider a sample of  $N = 1000$  giants with redshifts below  $z_{\max} = 0.5$ , use  $\phi_{\min} = 3'$ ,  $\phi_{\max} = 7'$ ,  $b_{\nu, \text{ref}} = 1000 \text{ Jy deg}^{-2}$ , and  $\sigma_{\text{ref}} = 1.5$ , and keep other parameters as in Fig. 5. *Top*: the probability that at least one observed giant has a projected length of at least  $l_p$ . *Bottom*: the expected number of observed giants with a projected length of at least  $l_p$ .

### 3.1. LoTSS DR2

The Low-Frequency Array (LOFAR; van Haarlem et al. 2013) is a powerful, Pan-European radio interferometer that features both (sub)arcsecond-scale resolution and sensitivity to degree-scale structures. Dabhade et al. (2020b) have already demonstrated that this combination is ideal for detecting giants: these authors found a record 225 new specimens in the LoTSS DR1, the first data release of the LOFAR’s Northern Sky survey at central frequency  $\nu_{\text{obs}} = 144 \text{ MHz}$ .<sup>9</sup> Excitingly, the recent LoTSS DR2 (Shimwell et al. 2022) improves the data calibration and increases the survey footprint from  $424 \text{ deg}^2$  to  $5635 \text{ deg}^2$  – that is by more than a factor 13. By default, the LoTSS features imagery at  $6''$  and  $20''$  resolutions. To further facilitate the discovery of giants (among other goals), we reprocessed the LoTSS by subtracting compact sources and imaging at  $60''$  and  $90''$  resolution; more details are given in Oei et al. (2022a, and in prep.). This  $60''$  and  $90''$  imagery has turned out to be effective in highlighting jets and lobes of RGs of large angular and physical extent, whose surface brightnesses are low and which therefore have remained undetected in shallower surveys, and even in the LoTSS DR2 at higher resolutions. We demonstrate this fact in Fig. 8 by comparing the LoTSS DR2 at  $6''$  and  $60''$  for three giants whose discovery has relied on the lower-resolution images. After the serendipitous discovery of several such hitherto unknown giants in the  $60''$  and  $90''$  images, we decided to initiate a systematic, multi-resolution, visual GRG search

<sup>9</sup> As in the current work, the authors defined giants using the projected proper length threshold  $l_{p, \text{GRG}} = 0.7 \text{ Mpc}$ .

through the area covered by LoTSS DR2 pipeline products as of September 2022<sup>10</sup>. This search comprised of a hundreds-of-hours-long inspection of the LoTSS maps at  $6''$  and  $60''$ , alongside Pan-STARRS DR1 (Chambers et al. 2016) and SDSS DR9 (Ahn et al. 2012) maps, in Aladin Desktop 11.0 (Bonnarel et al. 2000).

Reliable automated search strategies do not yet exist for several reasons. Giants showcase a rich morphological variety (see Sect. 3.6) and are easily confused with other types of astrophysical sources (see Sects. 3.3 and 3.4). Moreover, the known population is too small to effectively apply supervised learning techniques. However, it appears possible to find giants in morphological outlier lists of unsupervised learning techniques such as self-organising maps (SOMs; Mostert et al. 2021). The efficacy of such techniques in GRG searches has not yet been quantified.

### 3.2. Angular length threshold

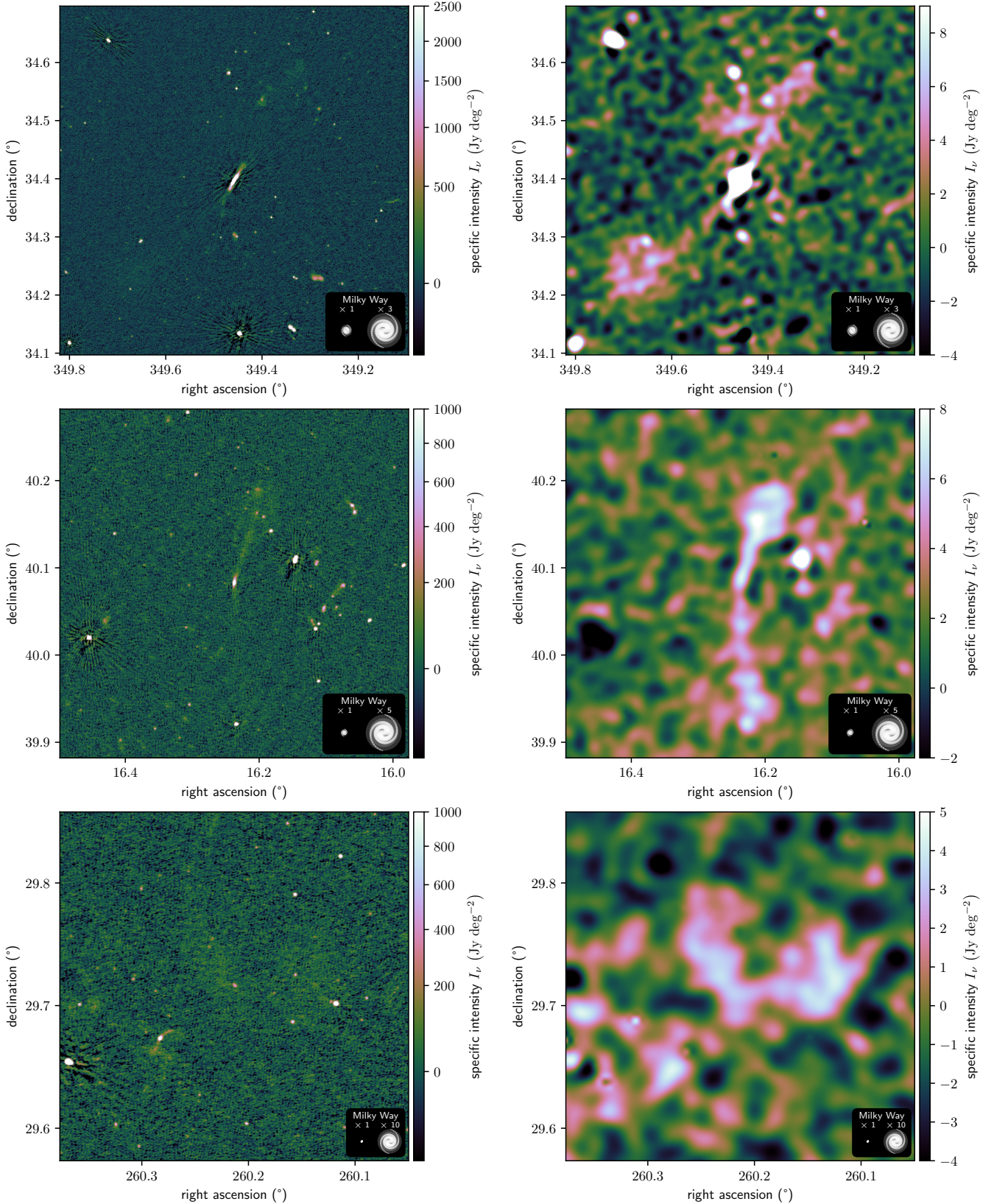
To limit the amount of manual work, we decided to search only for GRG candidates whose angular length exceeds some threshold. In Fig. 9, we show how the angular length  $\phi$  of giants with six different projected proper lengths  $l_p$  varies as a function of cosmological redshift  $z$ . Because of the expansion of the Universe, giants have an arcminute-scale minimum angular length: if  $l_{p, \text{GRG}} = 0.7 \text{ Mpc}$ , all giants obey  $\phi > 1.3'$ <sup>11</sup>. For the purpose of finding a GRG, it is therefore never useful to inspect a source with an angular length less than  $1.3'$ . Our highest priority has been to find giants with  $z < 0.2$ , which lie in a volume for which the total matter density field is known or will be known in the coming years through the combined power of deep spectroscopic surveys and Bayesian inference frameworks, such as the Bayesian Origin Reconstruction from Galaxies (BORG; Jasche & Wandelt 2013; Jasche et al. 2015; Jasche & Lavaux 2019). In an upcoming publication, we combine a sample of low-redshift giants with the BORG to analyse the large-scale environments of giants (Oei et al., in prep.). Figure 9 shows that all giants with  $l_p > 1 \text{ Mpc}$  at  $z < 0.2$  have an angular length  $\phi > 5'$ . For this reason, we have chosen  $5'$  as the angular length threshold of our search campaign. This choice has kept the visual inspection duration to order  $\sim 10^2 \text{ h}$ , while still enabling us to target all Mpc-exceeding giants in the Local Universe ( $z < 0.2$ ). In practice, this threshold is ‘fuzzy’: it is hard to accurately estimate angular lengths by eye before performing an actual measurement, so that our list of GRG candidates does contain some with a smaller angular length than the specified threshold. Inversely, it will presumably lack some GRG candidates with an angular length exceeding the threshold.

### 3.3. GRG candidates in the radio

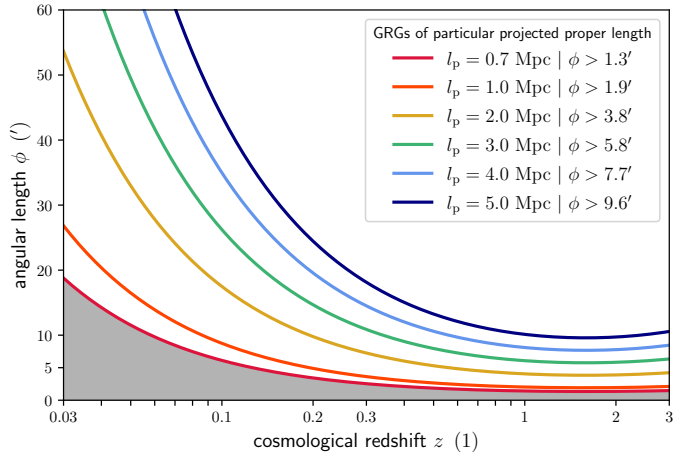
We first identified GRG candidates in the LoTSS at  $6''$  and  $60''$ . These maps serve complementary roles. The  $6''$  images reveal the precise morphology of radio galaxy cores and jets, which are necessary to pinpoint the host galaxy. Figure 10 provides a representative sense of the host galaxy identification accuracy these data allow for when combined with modern optical surveys such as the DESI Legacy Imaging Surveys (Dey et al. 2019) DR9. In contrast, the  $60''$  images have such compact sources removed or highly suppressed, but better highlight diffuse structures, such

<sup>10</sup> In this process, we have skipped the enclosed LoTSS DR1 area, which has already been analysed by Dabhade et al. (2020b).

<sup>11</sup> If  $l_{p, \text{GRG}} = 1 \text{ Mpc}$ , all giants obey  $\phi > 1.9'$ .



**Fig. 8.** LoTSS DR2 cutouts of three newly discovered giants at 6'' (left column) and 60'' (right column). By subtracting compact sources from calibrated 144 MHz visibility data and imaging at low resolution (60'' and 90''), we reveal otherwise speculative giant radio galaxies at the unexplored  $\sim 1 \text{ Jy deg}^{-2}$  surface brightness level. The claimed host galaxy is in the image centre. *Top:* a GRG of projected proper length  $l_p = 1.4 \pm 0.3 \text{ Mpc}$ , whose angular length  $\phi = 32.3 \pm 0.2'$  is larger than that of the full Moon. *Middle:* a GRG of  $l_p = 1.6 \pm 0.6 \text{ Mpc}$  and  $\phi = 16.4 \pm 0.2'$ . *Bottom:* a GRG of  $l_p = 3.6 \pm 0.1 \text{ Mpc}$  and  $\phi = 8.5 \pm 0.2'$ . For scale, we show the stellar Milky Way disk (with a diameter of 50 kpc) generated using the Ringermacher & Mead (2009) formula, alongside a 3, 5, or 10 times inflated version.



**Fig. 9.** Relations between cosmological redshift  $z$  and angular length  $\phi$  for six giants of different projected lengths  $l_p$ . Due to the expansion of the Universe, there is a minimum angular length for each  $l_p$ . If one defines giants as RGs with  $l_p \geq l_{p,\text{GRG}} = 0.7 \text{ Mpc}$ , all giants have an angular length of  $1.3'$  or above. If one instead defines giants as RGs with  $l_p \geq l_{p,\text{GRG}} = 1 \text{ Mpc}$ , all giants have an angular length of  $1.9'$  or above.

as RG lobes. Being similar in morphology, we made sure not to interpret diffuse emission from low-redshift spiral galaxies and their circumgalactic media, or radio halos and relics in galaxy clusters, as RG lobe emission. We required that all new RGs feature a detection of at least two<sup>12</sup> lobes, or of at least one lobe and one jet oriented towards the lobe(s), at least one of the resolutions used in this work ( $6''$ ,  $20''$ ,  $60''$ , and  $90''$ ).

### 3.4. GRG candidates in the optical

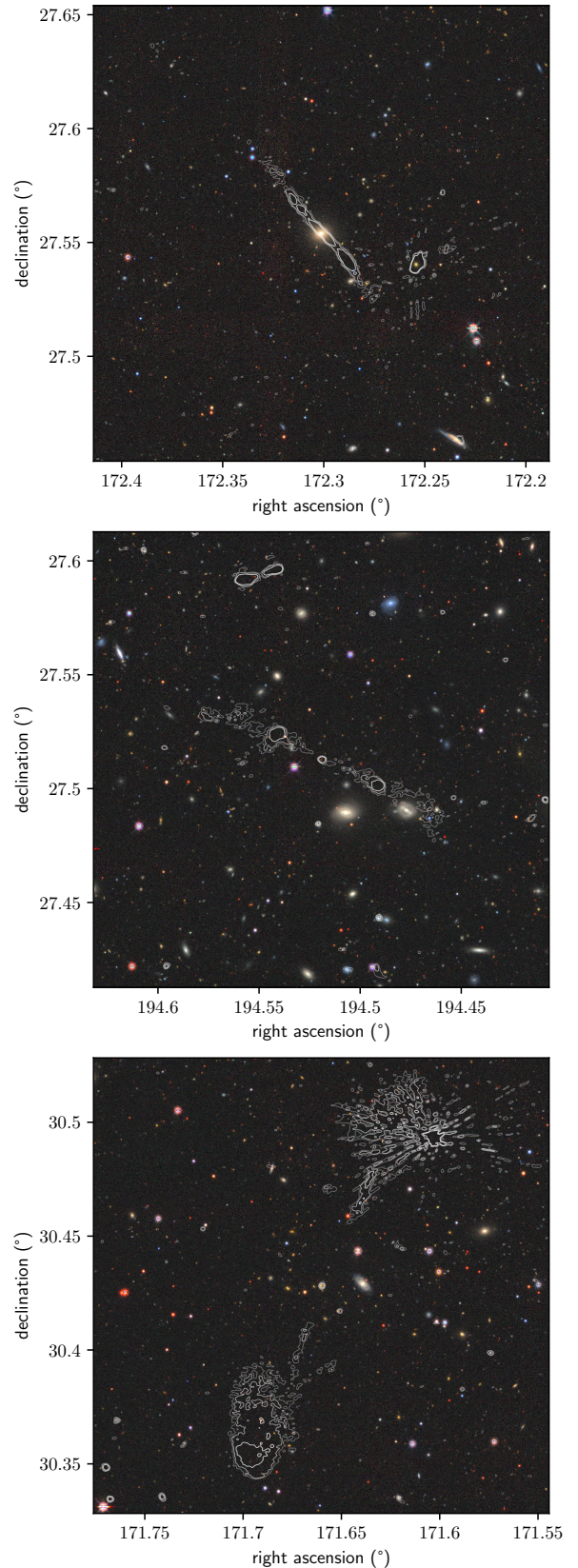
To confirm that a radio structure really is a radio galaxy, we compared the radio images with optical images of the same sky region. If a patch of radio emission is indeed due to RG jets or lobes, the patch itself must have no codirectional galactic counterpart in the optical. If the radio emission is due to a low-redshift spiral galaxy or a galaxy cluster instead, a corresponding easily recognisable counterpart will exist. We also took care not to erroneously associate the lobes of two distinct RGs. For this reason, we were more cautious to associate a pair of lobes to a suspected host galaxy when, in the optical, one could discern other galaxies in the angular vicinity of the lobes that could have generated them instead.

The Pan-STARRS and SDSS images used for these purposes complement each other, as they differ in quality throughout the sky – and in particular around sources of high optical flux density. Neither consistently outperforms the other. Only Pan-STARRS covers the full Northern Sky and could thus always be relied upon.

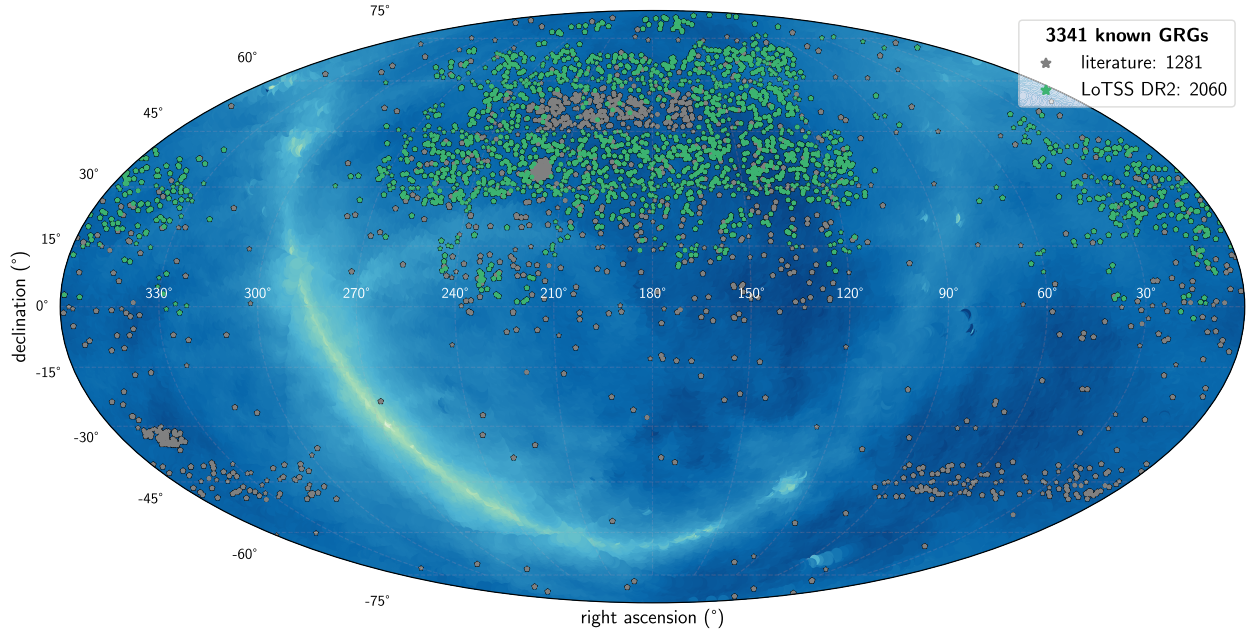
### 3.5. Host galaxy identification

We also used the Pan-STARRS and SDSS maps for the identification of host galaxies. We collected host redshifts from the SDSS DR12 (Alam et al. 2015) and *Gaia* (Gaia Collaboration 2016) DR3 (Gaia Collaboration 2021) through automated VizieR queries, from the Galaxy List for the Advanced Detector Era (GLADE) 2.4 (Dály et al. 2018), and from the DESI

<sup>12</sup> A small fraction of observed RGs are double-double radio galaxies, which show four lobes.



**Fig. 10.**  $12' \times 12'$  DESI Legacy Imaging Surveys DR9 ( $g, r, z$ )-details with LoTSS DR2  $6''$  contours ( $3\sigma$ ,  $5\sigma$ ,  $10\sigma$ ) overlaid. At  $6''$  resolution, LoTSS images allow for more accurate host galaxy identification in SDSS, Pan-STARRS, and Legacy Survey images than was possible before. *Top*: the jet of the giant at rank 33 of Table 2 and shown in the middle-left panel of Fig. B.1. *Middle*: the giant at rank 37 of Table 2. *Bottom*: the giant at rank 43 of Table 2.



**Fig. 11.** Mollweide view of the sky showing locations of all known giants, of which 62% are discoveries presented in this work. In the background, we show the specific intensity function of the Milky Way at  $\nu_{\text{obs}} = 150$  MHz (Zheng et al. 2017). The LoTSS DR2 has avoided the Galactic Plane, where extended emission complicates calibration and deconvolution. Our search footprint encloses a grey spherical rectangle, which represents the LoTSS DR1 search by Dabhade et al. (2020b), and a grey spherical cap, which represents the Boötes LOFAR Deep Field search by Simonte et al. (2022).

DR9 photometric redshift catalogue (Zou et al. 2022). If redshifts from multiple sources were available, we favoured SDSS over GLADE data, GLADE over *Gaia* data, and *Gaia* over DESI data. Similarly, we only adopted photometric redshifts if spectroscopic ones were not available.

For a small subset of RGs, a definite host galaxy could not be established beyond reasonable doubt, but a set of candidates containing the host galaxy could. In these cases, the lowest-redshift candidate provides a lower bound to the RG’s projected proper length<sup>13</sup>. If this lower bound exceeds  $l_{p,\text{GRG}} = 0.7$  Mpc, then the actual projected proper length certainly does so too, and the RG can be classified as a GRG – despite persisting uncertainty concerning the identity of the host galaxy.

### 3.6. LoTSS DR2 GRG sample

Our search campaign has led to the identification of 2060 hitherto unknown giants. To establish novelty, we assembled a literature catalogue with all known giants as of September 2022, combining the catalogue of Dabhade et al. (2020a)<sup>14</sup> with the giants discovered in Galvin et al. (2020), Ishwara-Chandra et al. (2020), Tang et al. (2020), Bassani et al. (2021), Brüggem et al. (2021), Delhaize et al. (2021), Masini et al. (2021), Kuźmicz & Jamroz (2021), Andernach et al. (2021), Mahato et al. (2022), Gürkan et al. (2022), and Simonte et al. (2022). Fusing our sample with this literature catalogue, we obtain a final catalogue with  $N = 3341$  giants.

Figure 11 shows the locations of all known giants in the sky. We list basic properties of the 50 projectively largest new

discoveries in Table 2, and refer to Appendix F for access to these data for our entire sample. In Fig. 12 and Figs. B.1–B.3, we present images for discoveries with projected proper lengths  $l_p$  in the ranges 5.1–4 Mpc, 4–3 Mpc, 3–2 Mpc, and 2–0.7 Mpc, respectively. For each giant, we use the LoTSS resolution  $\theta_{\text{FWHM}} \in \{6'', 20'', 60'', 90''\}$  that most clearly conveys the morphology through a single image. The selection reflects our sample’s diversity in shapes and sizes and provides a sense of the data quality.

#### 3.6.1. Angular lengths

The angular length distribution of the newly found giants is as follows: the smallest  $\phi = 1.5'$ , the median  $\phi = 5.2'$ , the largest  $\phi = 2.2^\circ$ , and 80% of angular lengths fall within  $[3.4', 9.8']$ . Thirteen of our discoveries – listed in Table 3 – are larger than the Moon in the sky (whose angular diameter varies over time, but here taken to be  $\phi = 30'$ ). Our search more than doubles the known number of such spectacular giants – from 10 to 23.

The GRG associated with NGC 2300 (see the middle-left panel of Fig. 13) is the giant with the largest angular length ever found, and the radio galaxy with the largest angular length in the Northern Sky<sup>15</sup>. It remains possible that the GRG has been generated by spiral galaxy NGC 2276 instead, with which elliptical galaxy NGC 2300 is interacting. However, this scenario seems unlikely, as only a fraction  $\sim 10^{-3}$  of known giants are hosted by spirals. Its discovery emphasises that low-frequency interferometers like the LOFAR and the MWA, which are sensitive to degree-scale angular scales, are important to complete a low-redshift census of giant radio galaxies. Sky-wide, a simple extrapolation of our findings suggests that several ( $\sim 10^1$ )

<sup>13</sup> This is true as long as all candidates have a redshift below that of the angular diameter distance maximum:  $z = 1.59$  for the cosmology adopted.

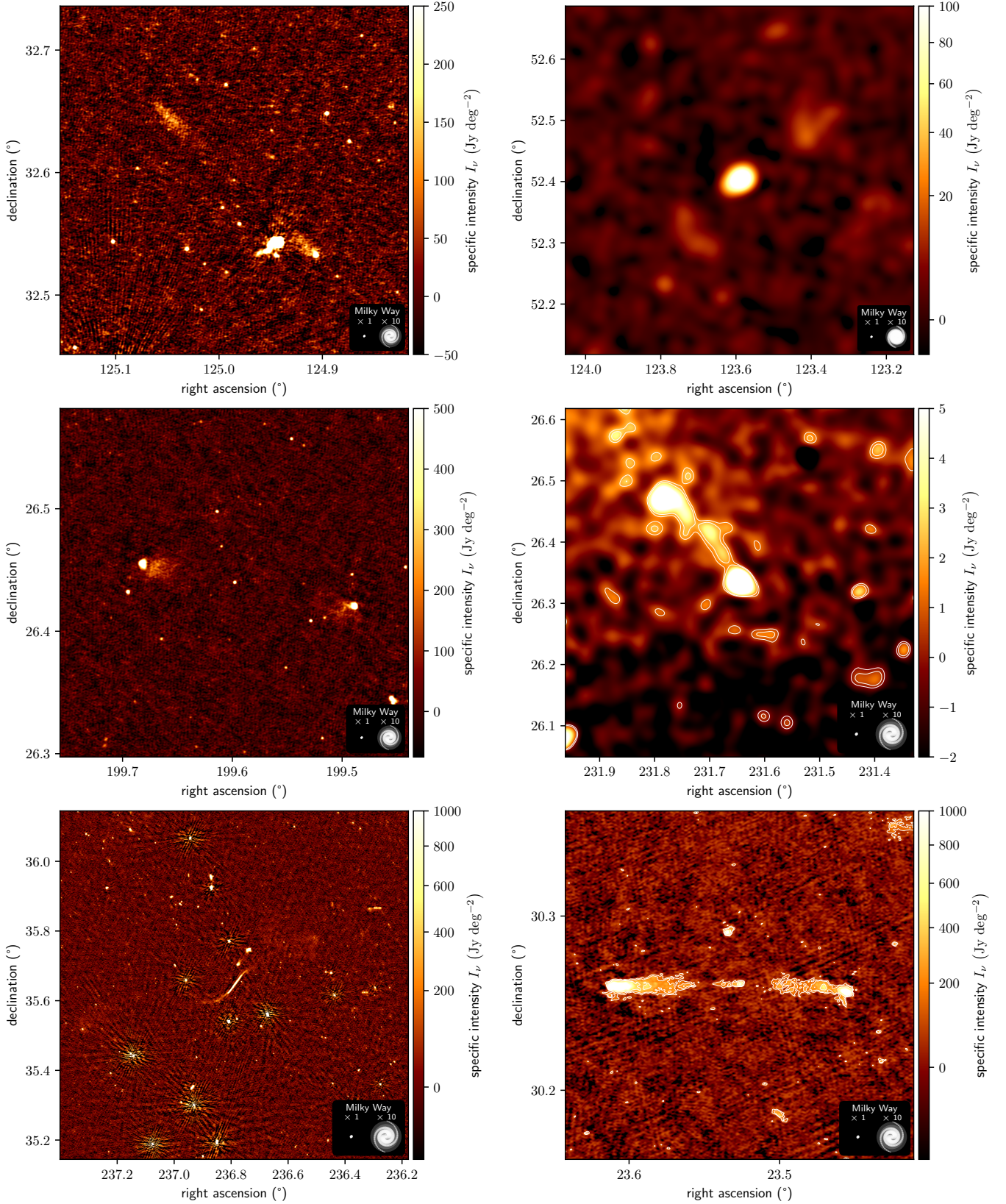
<sup>14</sup> This catalogue, complete up to and including April 2020, contains the giants found in 40 prior publications (for a list, see Dabhade et al. 2020a’s Sect. 1: Introduction), alongside their own discoveries.

<sup>15</sup> At  $\phi = 8^\circ$ , the Southern Sky’s Centaurus A (Cooper et al. 1965) is the radio galaxy with the largest angular length overall (e.g., McKinley et al. 2022); despite this, at  $l_p = 0.48$  Mpc, it is not a giant.

**Table 2.** Properties of the 50 projectively longest giants out of a total of 2060 discovered during our LoTSS DR2 search campaign.

Rank	Host name SDSS DR12	Host coordinates J2000 (°)	Redshift (1)	Redshift type	Angular length (′)	Projected proper length (Mpc)	Host stellar mass ( $10^{11} M_{\odot}$ )	Host SMBH mass ( $10^9 M_{\odot}$ )	Host quasar
1	J081956.41+323537.6	124.9851, 32.5938	0.749 ± 0.073	<i>p</i>	11.2	5.07 ± 0.20	...	...	...
2	J081421.68+522410.0	123.5904, 52.4028	0.2467 ± 6 × 10 <sup>-5</sup>	<i>s</i>	20.8	4.99 ± 0.04	2.4 ± 0.4	0.4 ± 0.2	<i>n</i>
3	J142910.70+311245.0	217.2946, 31.2125	0.5921 ± 0.0001	<i>s</i>	11.7	4.80 ± 0.06	...	2.3 ± 2.0	<i>n</i>
4	J131823.42+262622.8	199.5976, 26.4397	0.6230 ± 5 × 10 <sup>-5</sup>	<i>s</i>	11.0	4.62 ± 0.06	...	...	<i>y</i>
5	J152634.77+262003.2	231.6449, 26.3342	0.1507 ± 2 × 10 <sup>-5</sup>	<i>s</i>	28.0	4.56 ± 0.03	3.7 ± 0.6	1.4 ± 0.3	<i>n</i>
6	J121815.66+382407.5	184.5653, 38.4021	0.634 ± 0.064	<i>p</i>	10.6	4.49 ± 0.21	...	...	...
7	J175735.88+405154.2	269.3995, 40.8651	0.585 ± 0.036	<i>p</i>	10.5	4.29 ± 0.14	...	...	...
8	J161622.52+111135.7	244.0939, 11.1933	0.3574 ± 7 × 10 <sup>-5</sup>	<i>s</i>	13.4	4.15 ± 0.05	9.5 ± 1.8	5.7 ± 3.1	<i>n</i>
9	J154709.22+353846.1	236.7884, 35.6462	0.0794 ± 1 × 10 <sup>-5</sup>	<i>s</i>	43.8	4.08 ± 0.01	4.6 ± 0.1	3.9 ± 0.9	<i>n</i>
10	J013406.32+301537.2	23.5264, 30.2604	0.884 ± 0.138	<i>p</i>	8.5	4.06 ± 0.22	...	...	...
11	J082747.88+662813.6	126.9495, 66.4705	0.968 ± 0.160	<i>p</i>	8.2	4.02 ± 0.20	...	...	...
12	J012440.54+194003.9	21.1689, 19.6678	0.578 ± 0.162	<i>p</i>	9.6	3.90 ± 0.58	...	...	...
13	...	238.4466, 28.4763	1.094 ± 0.122	<i>p</i>	7.7	3.88 ± 0.12	...	...	...
14	...	275.3624, 26.6599	0.0850 ± 0.0001	<i>s</i>	39.0	3.86 ± 0.02	...	...	...
15	J162656.58+543421.3	246.7358, 54.5726	0.4887 ± 0.0001	<i>s</i>	10.3	3.84 ± 0.06	...	1.0 ± 1.6	<i>n</i>
16	J220605.67+275100.3	331.5237, 27.8501	0.317 ± 0.116	<i>p</i>	13.2	3.78 ± 0.99	...	...	...
17	J023544.96+310447.5	38.9373, 31.0799	0.541 ± 0.063	<i>p</i>	9.6	3.77 ± 0.23	...	...	...
18	...	136.9661, 67.1071	0.754 ± 0.081	<i>p</i>	8.3	3.77 ± 0.17	...	...	...
19	J180117.72+510722.4	270.3239, 51.1229	0.448 ± 0.086	<i>p</i>	10.3	3.66 ± 0.42	...	...	...
20	J123900.69+360924.5	189.7529, 36.1568	0.5935 ± 6 × 10 <sup>-5</sup>	<i>s</i>	8.8	3.62 ± 0.06	...	...	<i>y</i>
21	J102430.93+381842.8	156.1289, 38.3119	0.411 ± 0.028	<i>p</i>	10.7	3.62 ± 0.16	...	...	...
22	J172051.08+294256.8	260.2129, 29.7158	≥ 0.620 ± 0.044	<i>p</i>	8.5	≥ 3.56 ± 0.13	...	...	...
23	J090534.54+563052.0	136.3939, 56.5145	0.898 ± 0.056	<i>p</i>	7.4	3.55 ± 0.10	...	...	...
24	J133105.80+293435.7	202.7742, 29.5766	0.734 ± 0.034	<i>p</i>	7.9	3.55 ± 0.09	...	...	...
25	J223649.76+251242.5	339.2074, 25.2118	0.749 ± 0.075	<i>p</i>	7.8	3.53 ± 0.15	...	...	...
26	J230125.38+240148.2	345.3558, 24.0301	0.450 ± 0.089	<i>p</i>	9.9	3.53 ± 0.41	...	...	...
27	J004848.01+021003.1	12.2001, 2.1675	≥ 0.3604 ± 9 × 10 <sup>-5</sup>	<i>s</i>	11.2	≥ 3.49 ± 0.05	...	3.0 ± 1.5	<i>n</i>
28	J220239.13+070656.7	330.6631, 7.1158	0.4649 ± 7 × 10 <sup>-5</sup>	<i>s</i>	9.6	3.48 ± 0.05	...	0.9 ± 0.6	<i>n</i>
29	J165113.78+320943.4	252.8074, 32.1621	0.744 ± 0.045	<i>p</i>	7.7	3.48 ± 0.10	...	...	...
30	...	102.0173, 70.8276	0.714 ± 0.049	<i>p</i>	7.8	3.47 ± 0.12	...	...	...
31	...	11.2869, 28.7951	0.668 ± 0.042	<i>p</i>	8.0	3.46 ± 0.11	...	...	...
32	...	127.9215, 67.1934	0.451 ± 0.027	<i>p</i>	9.5	3.39 ± 0.12	...	...	...
33	J112912.14+273313.9	172.3006, 27.5539	0.0732 ± 1 × 10 <sup>-5</sup>	<i>s</i>	38.6	3.34 ± 0.01	3.7 ± 0.1	3.6 ± 0.8	<i>n</i>
34	J163659.07+541725.4	249.2461, 54.2904	0.5027 ± 5 × 10 <sup>-5</sup>	<i>s</i>	8.8	3.33 ± 0.06	...	53.4 ± 59.1	<i>n</i>
35	J092826.93+230448.0	142.1122, 23.0800	0.491 ± 0.045	<i>p</i>	8.9	3.33 ± 0.17	...	...	...
36	J001152.65+310024.3	2.9694, 31.0068	0.757 ± 0.083	<i>p</i>	7.3	3.32 ± 0.15	...	...	...
37	J125804.46+273046.0	194.5186, 27.5128	0.741 ± 0.083	<i>p</i>	7.3	3.29 ± 0.16	...	...	...
38	...	182.5080, 44.0903	1.031 ± 0.163	<i>p</i>	6.6	3.28 ± 0.15	...	...	...
39	J084127.02+554627.1	130.3626, 55.7742	0.7912 ± 3 × 10 <sup>-5</sup>	<i>s</i>	7.1	3.28 ± 0.07	...	25.2 ± 27.0	<i>n</i>
40	J143011.92+410404.2	217.5497, 41.0678	0.5868 ± 0.0001	<i>s</i>	8.0	3.27 ± 0.06	...	2.1 ± 1.4	<i>n</i>
41	J135119.31+340844.1	207.8305, 34.1456	0.923 ± 0.121	<i>p</i>	6.7	3.24 ± 0.14	...	...	...
42	J134436.33+291239.6	206.1514, 29.2110	0.766 ± 0.134	<i>p</i>	7.1	3.24 ± 0.23	...	...	...
43	J112638.34+302541.9	171.6598, 30.4283	0.3049 ± 4 × 10 <sup>-5</sup>	<i>s</i>	11.6	3.23 ± 0.04	...	1.5 ± 0.6	<i>n</i>
44	J012342.20+293633.1	20.9259, 29.6092	2.525 ± 0.300	<i>p</i>	6.5	3.22 ± 0.11	...	...	...
45	J223224.15+285753.3	338.1007, 28.9648	0.566 ± 0.038	<i>p</i>	8.0	3.21 ± 0.12	...	...	...
46	J103731.47+312948.9	159.3811, 31.4969	0.5228 ± 0.0001	<i>s</i>	8.3	3.21 ± 0.06	...	2.2 ± 1.4	<i>n</i>
47	J154742.69+384119.4	236.9279, 38.6887	0.280 ± 0.024	<i>p</i>	12.2	3.21 ± 0.20	...	...	...
48	J114333.93+425800.5	175.8914, 42.9668	0.802 ± 0.075	<i>p</i>	6.9	3.20 ± 0.12	...	...	...
49	J122329.86+313116.0	185.8744, 31.5211	0.666 ± 0.042	<i>p</i>	7.4	3.20 ± 0.11	...	...	...
50	...	212.1410, 67.8036	1.103 ± 0.181	<i>p</i>	6.3	3.17 ± 0.14	...	...	...

**Notes.** The giants are ranked by projected proper length. The column ‘rank’ thus denotes each giant’s projected proper length rank within this new sample, not within the total known population. Lying outside of the coverage, some GRG host galaxies have no SDSS DR12 name. The column ‘host coordinates’ contains the central right ascension and declination of the host galaxy. The columns ‘redshift’ and ‘redshift type’ provide cosmological redshift estimates, derived from spectroscopy *s* or photometry *p*. The column ‘angular length’ denotes the largest great-circle distance between two LoTSS DR2–detectable GRG (end)points. These angular lengths may increase in future deeper surveys. We have not measured angular length errors on a case by case basis, but estimate them to be 0.15′. The column ‘projected proper length’ propagates both redshift and angular length uncertainty and assumes the [Planck Collaboration VI \(2020\)](#) cosmology. The columns ‘host stellar mass’ and ‘host SMBH mass’, further discussed in Sect. 3.6.4 and Appendix C, provide SDSS-derived estimates of host stellar and supermassive black hole masses. Finally, the column ‘host quasar’ indicates whether the host’s AGN has a quasar appearance: *y* (yes) or *n* (no). In cases where only a set of candidates containing the host galaxy could be established beyond reasonable doubt, we list the properties of the lowest-redshift candidate. In this way, the provided projected proper length bounds the actual projected proper length from below. In such cases, to signify uncertainty, we mark the host name, host coordinates, and physical host properties in grey. We note that this lowest-redshift candidate is often, but not always, also the most probable host. For access to these data for all 2060 newly discovered giants, see Appendix F.



**Fig. 12.** Details of the LoTSS DR2–estimated specific intensity function  $I_\nu(\hat{f})$  at central observing frequency  $\nu_{\text{obs}} = 144$  MHz and resolutions  $\theta_{\text{FWHM}} \in \{6'', 90''\}$ , centred around the hosts of newly discovered giants. Row-wise *from left to right, from top to bottom*, the projected proper length  $l_p$  is 5.1 Mpc, 5.0 Mpc (Oei et al. 2022a), 4.6 Mpc, 4.6 Mpc, 4.1 Mpc, and 4.1 Mpc; in the same order,  $\theta_{\text{FWHM}}$  is  $6''$ ,  $90''$ ,  $6''$ ,  $90''$ ,  $6''$ , and  $6''$ . The GRG in the bottom-left panel appears larger in the sky than the Moon. In the *middle-right panel*, contours signify 2.5 and 3.5 sigma-clipped standard deviations (SDs) above the sigma-clipped median; in the *bottom-right panel*, they signify 3, 5, and 10 such SDs. For scale, we show the stellar Milky Way disk (with a diameter of 50 kpc) generated using the Ringermacher & Mead (2009) formula, alongside a 10 times inflated version.

**Table 3.** (Extended) Table 2 ranks, host names, angular lengths, and LoTSS DR2 image references for all 13 newly discovered giants that appear larger in the sky than the Moon ( $\phi > 30'$ ).

Rank	Host name	Angular length ( $'$ )	LoTSS DR2 image reference
Table 2	NGC or LEDA		
1441	NGC 2300	2.2°	Fig. 13, middle left
141	NGC 6185	1.0°	Fig. 13, top left
9	LEDA 56028	43.8'	Fig. 12, bottom left
144	LEDA 54794	40.0'	–
14	LEDA 5060619	39.0'	Fig. B.1, top left
33	LEDA 1811497	38.6'	Fig. B.1, middle left
1692	NGC 2789	34.5'	Fig. 13, bottom right
1643	LEDA 38523	33.8'	Fig. 13, middle right
1770	NGC 1044	32.7'	Fig. 13, bottom left
178	LEDA 2048533	32.3'	Fig. 8, top row
1486	NGC 7385	32.0'	–
326	LEDA 3090801	31.0'	Fig. B.2, bottom left
465	LEDA 37801	30.1'	Fig. 13, top right

degree-scale angular length giants similar to the GRG of NGC 2300 still await discovery at the LoTSS DR2 depth.

### 3.6.2. Redshifts

The redshift distribution of the newly found giants is as follows: the lowest  $z = 0.00635 \pm 6 \times 10^{-5}$ , the median  $z = 0.29$ , the highest  $z = 2.6394 \pm 6 \times 10^{-4}$ , and 80% of redshifts fall within [0.12, 0.68]. Because of our focus on giants of large angular length (see Sect. 3.2), we have found only 36 giants beyond  $z > 1$ . One of these, the GRG at rank 13 of Table 2, is the largest secure giant found beyond redshift 1. It lies at  $z = 1.1 \pm 0.1$  and spans  $l_p = 3.9 \pm 0.1$  Mpc. Its host galaxy does not appear to contain a quasar.

### 3.6.3. Projected proper lengths

With  $l_p = 5.1 \pm 0.2$  Mpc and  $l_p = 4.99 \pm 0.04$  Mpc, our LoTSS DR2 sample contains the first two 5 Mpc-scale giants. We have presented a dedicated analysis of the latter GRG in Oei et al. (2022a). 11 discoveries have  $l_p \geq 4$  Mpc, 53 have  $3 \leq l_p < 4$  Mpc, 291 have  $2 \leq l_p < 3$  Mpc, 1215 have  $1 \leq l_p < 2$  Mpc, and 490 have  $0.7 \leq l_p < 1$  Mpc. The median  $l_p = 1.35$  Mpc, and 80% of projected proper lengths fall within [0.82 Mpc, 2.29 Mpc].

### 3.6.4. Stellar and supermassive black hole masses

Following Oei et al. (2022a), we collected host stellar masses  $M_\star$  from Chang et al. (2015) and Salim et al. (2018), and estimated host SMBH masses  $M_\bullet$  via SDSS DR12 stellar velocity dispersions (Alam et al. 2015) and the M-sigma relation of Kormendy & Ho (2013)'s Eq. (7). From all 3341 giants in our final catalogue, only 732 (22%) could be assigned a stellar mass in this way, and only 1115 (33%) an SMBH mass; for both quantities, our LoTSS DR2 sample accounts for four-fifths of the resulting subpopulation. Figure C.1 shows both  $M_\star$  and  $M_\bullet$  in relation to projected proper length  $l_p$ .

The median  $M_\star = 3.4 \times 10^{11} M_\odot$ , and 80% of stellar masses fall within [ $1.8 \times 10^{11} M_\odot$ ,  $5.3 \times 10^{11} M_\odot$ ]. We discover two giants whose hosts, J150329.07+374850.3 and J073505.24+415827.5, are the least massive known: both have a stellar mass  $M_\star =$

$4.8 \times 10^{10} M_\odot$ . These are small giants, with  $l_p = 0.8$  Mpc and  $l_p = 0.7$  Mpc, respectively. The top panel of Fig. C.1 hints at a weak positive correlation between  $M_\star$  and  $l_p$ , which future work should confirm or reject.

The median  $M_\bullet = 1.0 \times 10^9 M_\odot$ , and 80% of SMBH masses fall within [ $0.4 \times 10^9 M_\odot$ ,  $2.2 \times 10^9 M_\odot$ ]. The SMBH masses of J123703.24+275819.5, J163659.07+541725.4, and J103129.54+502959.1 are the highest estimated yet, with  $M_\bullet = 2 \times 10^{11} M_\odot$ ,  $M_\bullet = 5 \times 10^{10} M_\odot$ , and  $M_\bullet = 5 \times 10^{10} M_\odot$ , respectively. The latter masses equal the theoretical maximum mass of accreting black holes of typical spin (King 2016). Curiously, although J103129.54+502959.1's  $M_\bullet$  is among the highest estimated SMBH mass of any GRG host, the GRG itself is relatively small:  $l_p = 0.81 \pm 0.06$  Mpc. Conversely, the bottom panel of Fig. C.1 shows that multi-Mpc giants can have hosts with SMBH masses that are two orders of magnitude lower than J103129.54+502959.1's.

### 3.6.5. Spiral or lenticular host galaxies

Remarkably, although NGC 6185 is a spiral galaxy of Hubble–de Vaucouleurs class SAa (Jansen et al. 2000), it appears to have generated the giant shown in the top-left panel of Fig. 13. Such systems are exceedingly rare: not only are few RGs giants, but also virtually all giants have an elliptical galaxy as their host. With  $l_p = 2.54 \pm 0.01$  Mpc, this is the largest known spiral galaxy-hosted radio galaxy. Hitherto, the largest spiral galaxy-hosted giant in the literature has been J2345-0449 (Bagchi et al. 2014), with  $l_p = 1.6$  Mpc. Given the favourably low redshift of  $z = 0.03430 \pm 7 \times 10^{-5}$ , NGC 6185 and its enigmatic giant solicit a dedicated analysis (Oei et al. 2023). Besides NGC 6185, spiral or lenticular galaxies J080403.40+404809.3 and J091459.66+294348.8 (known alternatively as NGC 2789; see the bottom-right panel of Fig. 13) have also generated giants; these have projected proper lengths  $l_p = 1.1$  Mpc and  $l_p = 0.9$  Mpc, respectively. Morphological host classification through SDSS, Pan-STARRS, and DESI imagery is reliable only up to  $z \sim 0.1$ –0.2, depending on viewing angle and various other factors. Our LoTSS DR2 sample contains 342 giants with definite hosts at  $z = 0.15$  or below, among which are all 3 giants with spiral or lenticular hosts discussed here. It thus appears that the fraction of GRG hosts that is of such non-elliptical nature is  $\sim 1\%$ . A more detailed morphological characterisation of the sample appears possible, for example using data from Hart et al. (2016), but this is beyond the scope of the current work.

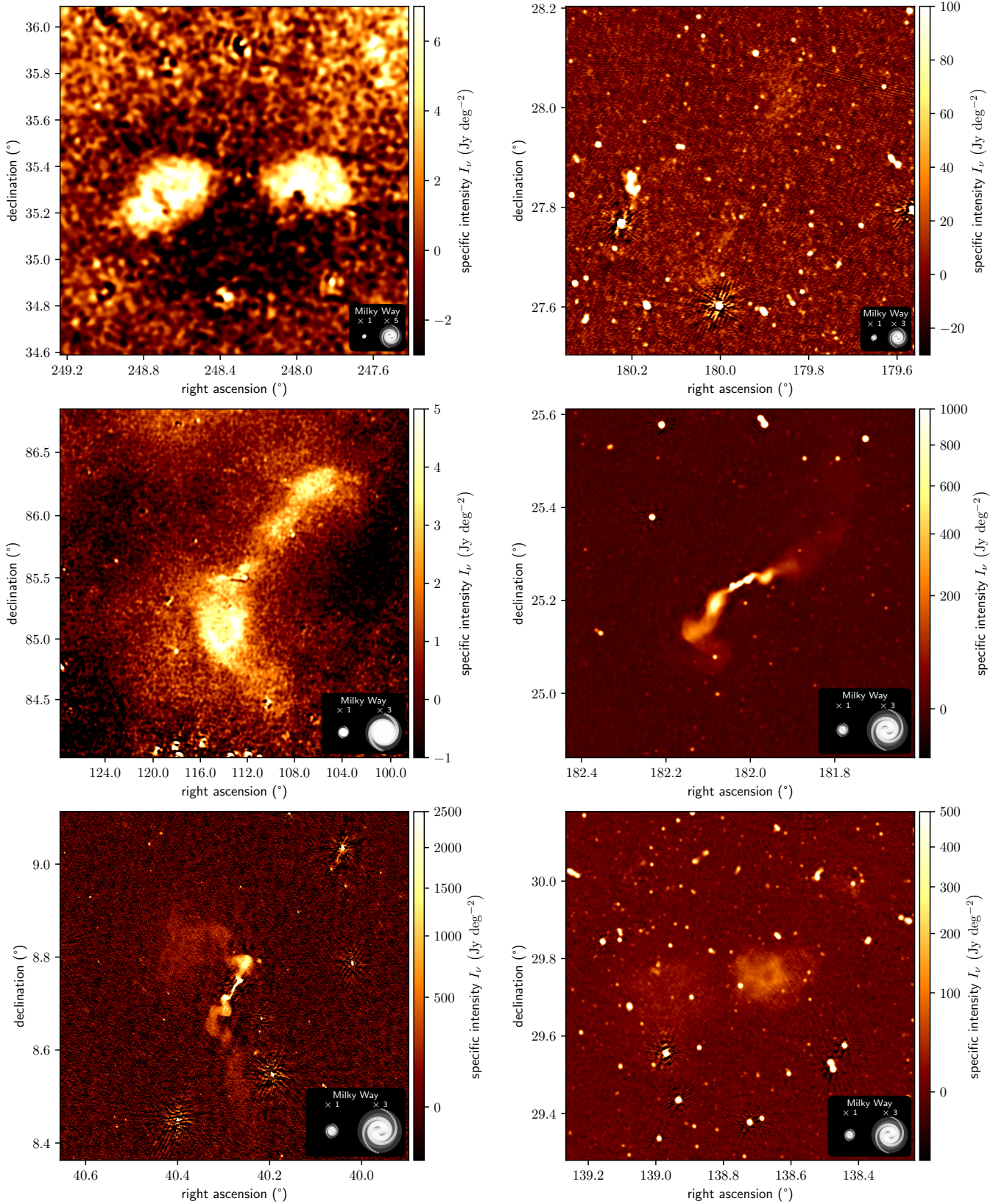
## 4. Results

After first building a statistical framework and then collecting a large sample of giants from a single survey through a systematic approach, we were ready to infer the intrinsic GRG length distribution. In particular, we aimed to establish whether the RG intrinsic proper length RV  $L$  is well described by a Pareto distribution, and if so, what its tail index  $\xi$  is. Subsequently, we inferred derived quantities, such as the comoving GRG number density in the Local Universe.

### 4.1. Giant radio galaxy length distribution

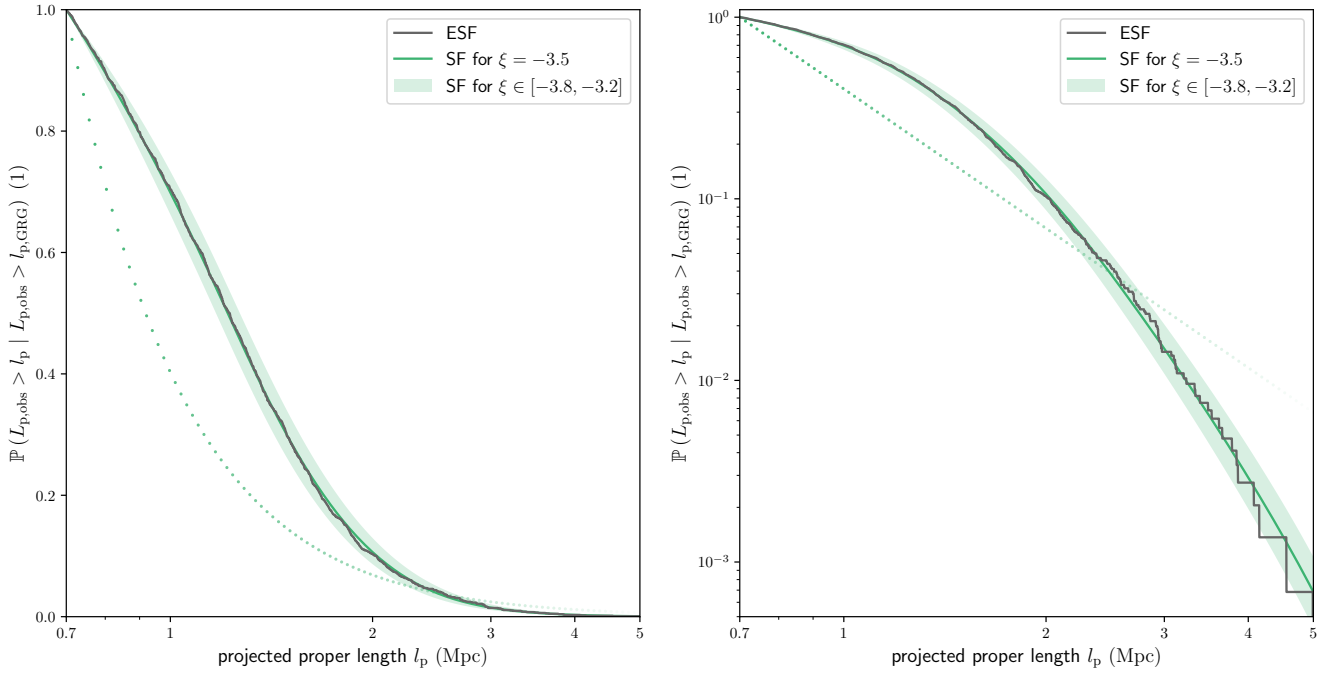
#### 4.1.1. Empirical survival function

From our LoTSS DR2 GRG sample, we computed the empirical cumulative distribution function (ECDF) of the GRG observed projected proper length RV  $L_{p,\text{obs}} \mid L_{p,\text{obs}} > l_{p,\text{GRG}}$  (see Eq. (21)).



**Fig. 13.** Details of the LoTSS DR2–estimated specific intensity function  $I_\nu(\hat{r})$  at central observing frequency  $\nu_{\text{obs}} = 144$  MHz and resolutions  $\theta_{\text{FWHM}} \in \{6'', 20'', 90''\}$ , centred around the hosts of newly discovered giants. Row-wise from left to right, from top to bottom, the projected proper length  $l_p$  is 2.5 Mpc, 1.9 Mpc, 1.1 Mpc, 1.0 Mpc, 0.9 Mpc, and 0.9 Mpc; in the same order,  $\theta_{\text{FWHM}}$  is  $90''$ ,  $20''$ ,  $90''$ ,  $20''$ ,  $6''$ , and  $20''$ . All appear larger in the sky than the Moon. The top-left panel shows the giant of NGC 6185, a spiral galaxy. This is the first spiral galaxy–hosted giant known with  $l_p > 2$  Mpc. The middle-left panel shows a structure we interpret as a radio galaxy belonging to the elliptical galaxy NGC 2300. At  $\phi = 2.2^\circ$ , this giant has the largest angular length of all uncovered thus far.





**Fig. 14.** Empirical survival function of the observed giant radio galaxy projected proper length RV (ESF; dark grey) and the corresponding survival function  $1 - F_{L_{p,obs} | L_{p,obs} > l_{p,GRG}}$  (SF; green curve) using the maximum a posteriori probability parameters (MAP; see Table 4). Observed GRG projected lengths are well described by a Pareto distribution modified to include selection effects. Keeping the selection effect parameters fixed, we show how models vary with tail index  $\xi$  (green range). We also show the selection effect-free SF  $1 - F_{L_{p} | L_{p} > l_{p,GRG}}$  using the MAP  $\xi$  (green dots). We included all LoTSS DR2 search campaign giants up to  $z_{max} = 0.5$ . *Left:* logarithmic horizontal axis and linear vertical axis. *Right:* logarithmic horizontal axis and logarithmic vertical axis.

We only included giants for which  $l_p$  itself – rather than a lower bound – is known, and set either  $z_{max} = 0.5$  or  $z_{max} = 0.25$ ; this retained 1473 or 811 giants for analysis, respectively<sup>16</sup>. The empirical survival function (ESF), which equals one minus the ECDF, is shown in dark grey in both panels of Fig. 14. Just as the ECDF approximates the CDF, the ESF approximates the survival function (SF). In this case, for any  $l_p$ , the ESF provides the probability that a randomly drawn LoTSS DR2 GRG will have a projected proper length exceeding  $l_p$ . If  $L_{p,obs} | L_{p,obs} > l_{p,GRG}$  were Paretian, its ESF (hereafter: ‘the’ ESF) would resemble a straight line in Fig. 14’s right panel – both axes have logarithmic scaling. However, the ESF clearly displays curvature.

#### 4.1.2. Selection effects

Under the ansatz that  $L$  is Paretian, as proposed in Sect. 2.1, the aforementioned ESF’s curvature implies a significant role for observational selection effects. The reason is the following. The ansatz implies that the GRG projected proper length RV  $L_p | L_p > l_{p,GRG}$  is also Paretian (see Eq. (5)), as is  $L_{p,obs} | L_{p,obs} > l_{p,GRG}$  when selection effects are negligible (see Sect. 2.8.1 and set  $C(l_p) = 1$ ). But in our case  $L_{p,obs} | L_{p,obs} > l_{p,GRG}$  is not Paretian: its ESF is curved. To avoid contradiction, we must relax at least one assumption. If the Pareto ansatz is held, then selection effects must be at play.

When selection effects are non-negligible, we must devise a procedure to disentangle them from the data if our  $\xi$  estimate is to be uncontaminated. To this end, we performed joint Bayesian

inference with a model that includes both  $\xi$  and parameters that describe the selection effects.

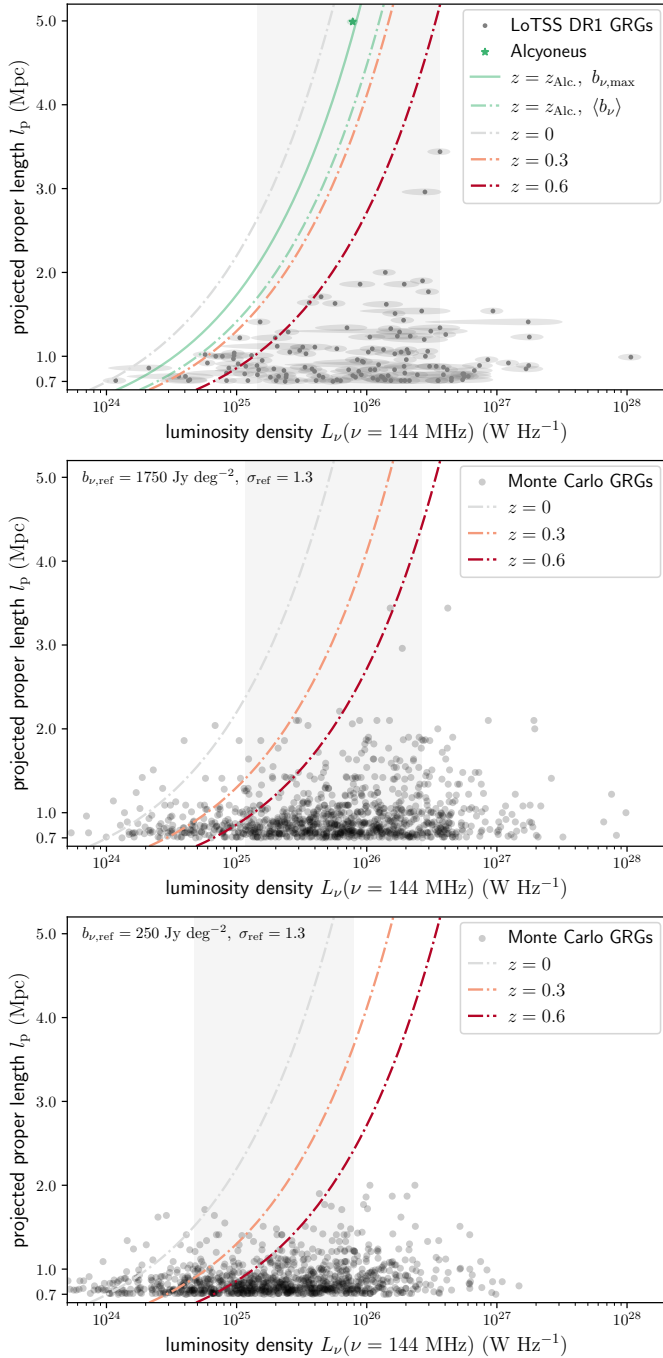
In particular, we considered the roles of the fuzzy angular length threshold and surface brightness selection effects, introduced in Sects. 2.8.2 and 2.8.3, respectively. In Sect. 3.2, we explain that we have attempted to maintain a  $5'$  angular length threshold during our LoTSS DR2 GRG search. If we want to use Sect. 2.8.2 to correct for the bias against faraway and physically small giants that this threshold has imprinted onto our sample, we must estimate parameters  $\phi_{min}$  and  $\phi_{max}$ . A natural choice is to assume that they lie symmetrically around the intended angular length threshold of  $5'$  – but at what distance from it? We propose to consider this distance,  $\frac{1}{2}(\phi_{max} - \phi_{min})$ , as a yet unknown model parameter that we fitted to the data. Similarly, if we want to use Sect. 2.8.3 to correct for the bias against physically large giants that the limited depth of the LoTSS DR2 imprints onto our sample, we must estimate parameters  $b_{v,ref}$  and  $\sigma_{ref}$ . Our approach was to, again, fit these parameters – possibly making use of any available prior knowledge.

The meaning of  $b_{v,ref}$  depends on the choice of  $l_{ref}$  and  $\nu_{obs}$ ; we defined  $l_{ref} := 0.7$  Mpc and used  $\nu_{obs} = 144$  MHz. We furthermore assumed  $b_{v,th} = 1 \times \sigma_{I_v}$ , with  $\sigma_{I_v} = 25$  Jy deg<sup>-2</sup> being the typical LoTSS DR2  $6''$  RMS noise (Shimwell et al. 2022).

#### 4.1.3. Prior

We exploited two sources of prior knowledge. Firstly, we attempted to directly estimate  $b_{v,ref}$  and  $\sigma_{ref}$  by selecting from all LoTSS DR2 giants with  $l_p \leq 1$  Mpc a random subset of size 50 (10%). For these giants, we estimated the mean surface brightnesses of both lobes from the LoTSS DR2  $6''$  imagery, differentiating between the brighter and the fainter lobe. Because our goal was to estimate  $b_{v,ref}$ , we attempted to undo cosmological and

<sup>16</sup> We explored two choices for  $z_{max}$  as our model assumes that  $\xi$  remains constant between  $z \in [0, z_{max}]$ . We further discuss this assumption in Sect. 5.5.



**Fig. 15.** Comparison between luminosity density–projected proper length relations for observed and simulated giants. Each dash-dotted curve denotes a family of giants at a given redshift, assuming  $f_{L_\nu} = 0.3$  and  $f_l = 0.3$ , whose mean lobe surface brightnesses equal the LoTSS DR2 noise level and who are thus borderline-detectable. The grey band denotes the median-centred luminosity density range that contains 68% of giants. *Top:* 139 LoTSS DR1 giants (Dabhade et al. 2020b), alongside Alcyoneus, with redshift  $z_{\text{Alc.}} = 0.25$  (Oei et al. 2022a). The solid green curve is similar to the dash-dotted green curve, but represents maximum instead of mean lobe surface brightness. *Middle:* 1000 simulated giants, assuming  $b_{\nu,\text{ref}} = 1750 \text{ Jy deg}^{-2}$  and  $\sigma_{\text{ref}} = 1.3$ . *Bottom:* 1000 simulated giants, assuming  $b_{\nu,\text{ref}} = 250 \text{ Jy deg}^{-2}$  and  $\sigma_{\text{ref}} = 1.3$ .

growth-induced surface brightness dimming assuming a universal lobe spectral index  $\alpha = -1$  and self-similar growth:  $\zeta = -2$ . The resulting surface brightnesses correspond to  $z = 0$ , and to the epoch in each giant’s life when  $l = l_{\text{ref}}$ . The median of

the corrected bright lobe mean surface brightnesses is  $b_{\nu,\text{ref}} = 1.3 \times 10^3 \text{ Jy deg}^{-2}$ , whilst the median of the corrected faint lobe mean surface brightnesses is  $b_{\nu,\text{ref}} = 0.7 \times 10^3 \text{ Jy deg}^{-2}$ . All lobes taken together, the median becomes  $b_{\nu,\text{ref}} = 1.0 \times 10^3 \text{ Jy deg}^{-2}$ . We performed maximum likelihood estimation assuming the surface brightness distribution is lognormal and found  $\sigma_{\text{ref}} = 1.3$ , again using all lobes. Note, however, that we have only used observed giants here, whilst  $b_{\nu,\text{ref}}$  and  $\sigma_{\text{ref}}$  should correspond to the entire population of giants. As fainter giants will have preferentially fallen out, we might have overestimated  $b_{\nu,\text{ref}}$  and underestimated  $\sigma_{\text{ref}}$ .

To probe whether we had overestimated  $b_{\nu,\text{ref}}$  and underestimated  $\sigma_{\text{ref}}$ , we used data from Dabhade et al. (2020b) and a Monte Carlo approach. First, for their sample of 239 LoTSS DR1 giants, we computed total luminosity densities  $L_\nu$  at rest-frame frequency  $\nu = 144 \text{ MHz}$ . (Given that LoTSS DR1 and DR2 noise levels are similar, this population is also representative of the LoTSS DR2.) In the top panel of Fig. 15, we show  $L_\nu$  versus  $l_p$  for all 139 for which  $z < 0.6$ . To increase the range of lengths covered, we additionally show data on Alcyoneus (Oei et al. 2022a). Next, under the same assumptions of a constant spectral index and self-similar growth, we derived a simple luminosity density–surface brightness relationship that allows for back-and-forth conversion between the two – at least, given projected lengths and redshifts. The mean lobe surface brightness  $\langle b_\nu \rangle$  is proportional to the total luminosity density  $L_\nu$ , and assuming a pair of spherical lobes

$$\langle b_\nu \rangle = \frac{2f_{L_\nu} \cdot L_\nu}{\pi^2 \cdot \mathbb{E}[D](\eta(f_l)) \cdot f_l^2 \cdot l_p^2 \cdot (1+z)^{3-\alpha}}. \quad (40)$$

Here  $f_{L_\nu}$  is the fraction of the total luminosity density that belongs to the lobes,  $f_l$  is the fraction of the RG’s axis length that lies inside the lobes, and  $\mathbb{E}[D](\eta(f_l))$  is the mean deprojection factor as given by Eq. (A.29). The peak surface brightness  $b_{\nu,\text{max}}$  relates to  $\langle b_\nu \rangle$  as  $b_{\nu,\text{max}} = \frac{3}{2}\langle b_\nu \rangle$ . Appendix D contains derivations for both these results. For Alcyoneus, at  $z_{\text{Alc.}} = 0.25$ ,  $f_{L_\nu} = 0.3$  and  $f_l = 0.3$  (Oei et al. 2022a). Assuming these parameter values, again in the top panel of Fig. 15, we show luminosity density–projected length pairs of RGs at  $z = z_{\text{Alc.}}$  whose peak (solid green curve) or mean (dash-dotted green curve) surface brightness equals the LoTSS DR2 noise level. Thus, each curve represents a family of borderline-detectable giants at Alcyoneus’s redshift. The other dash-dotted curves indicate similar barely detectable families, but for other redshifts. Without optimising any free parameters, the curves correctly predict that Alcyoneus’s lobes have surface brightnesses comparable to the 6’’ LoTSS DR2 noise level and explain the absence of observations in the top-left corner of the figure. We conclude that Eq. (40) appears reasonable, but note that RGs may significantly differ in their values of  $f_{L_\nu}$  and  $f_l$ .<sup>17</sup>

Bolstered, we made use of Eq. (40) to Monte Carlo simulate – for particular values of  $b_{\nu,\text{ref}}$  and  $\sigma_{\text{ref}}$  – luminosity density–projected length relationships as they appear to observers. The simulated giants have projected lengths adopted from the observed giants, randomly sampled redshifts up to  $z = 0.6$  assuming a spatially constant GRG number density, and randomly sampled reference surface brightnesses (i.e. those for RGs at  $z = 0$  that have intrinsic length  $l = l_{\text{ref}}$ ) whose distribution is determined by  $b_{\nu,\text{ref}}$  and  $\sigma_{\text{ref}}$ . We then used Eq. (24) to compute surface brightnesses as they would be observed, and retained only those giants whose surface brightness exceeds the

<sup>17</sup> This is also the reason that some giants in Fig. 15 cross their redshift’s dash-dotted curve, which represents  $f_{L_\nu} = 0.3$  and  $f_l = 0.3$  only.

LoTSS DR2 noise level. For these simulated detectable giants, we finally generated luminosity densities using Eq. (40), assuming wide uniform distributions  $f_{L_\nu} \sim f_l \sim \text{Uniform}(0.1, 0.9)$ . The middle and bottom panels of Fig. 15 show results for  $b_{\nu,\text{ref}} = 1750 \text{ Jy deg}^{-2}$  and  $b_{\nu,\text{ref}} = 250 \text{ Jy deg}^{-2}$ , respectively; we adopted  $\sigma_{\text{ref}} = 1.3$  from our LoTSS DR2 GRG surface brightness measurements. The median luminosity density of the  $z < 0.6$  LoTSS DR1 giants is  $L_\nu = 1.1 \times 10^{26} \text{ W Hz}^{-1}$ , that of the  $b_{\nu,\text{ref}} = 1750 \text{ Jy deg}^{-2}$  simulated giants is  $L_\nu = 0.6 \times 10^{26} \text{ W Hz}^{-1}$ , and that of the  $b_{\nu,\text{ref}} = 250 \text{ Jy deg}^{-2}$  simulated giants is  $L_\nu = 0.2 \times 10^{26} \text{ W Hz}^{-1}$ .<sup>18</sup> Interestingly, the higher reference surface brightness median provides a better fit to the data. Even in case our Monte Carlo approach would predict luminosity densities that are biased low by a factor two, the higher median remains favoured.

In conclusion, it seems reasonable to suppose that our measurement  $b_{\nu,\text{ref}} = 1.0 \times 10^3 \text{ Jy deg}^{-2}$  is not, or only mildly, biased high by selection effects. Still, we take a conservative approach and in setting priors we assume a 75% error on our measurement of  $b_{\nu,\text{ref}}$  and a 50% error on our measurement of  $\sigma_{\text{ref}}$ . Thus, the priors for  $b_{\nu,\text{ref}}$  and  $\sigma_{\text{ref}}$  – which we choose to be Gaussian – have 68% credible intervals [ $250 \text{ Jy deg}^{-2}, 1750 \text{ Jy deg}^{-2}$ ] and [0.65, 1.95]. We retain flat priors for  $\xi$  and  $\frac{1}{2}(\phi_{\text{max}} - \phi_{\text{min}})$ .

#### 4.1.4. Inference

To compute the posterior distribution for  $\xi$ ,  $\frac{1}{2}(\phi_{\text{max}} - \phi_{\text{min}})$ ,  $b_{\nu,\text{ref}}$  and  $\sigma_{\text{ref}}$ , we first brute-force evaluated the likelihood function over a regular grid that covers a total of 3.3 million parameter combinations<sup>19</sup>. For each proposed parameter quartet, we computed the PDF of  $L_{\text{p,obs}} | L_{\text{p,obs}} > l_{\text{p,GRG}}$ , and obtained the likelihood assuming that the LoTSS DR2 GRG projected proper lengths are IID draws from it. To obtain the PDF, we successively evaluated Eqs. (22), (25), (29), (18), (3), and (21), alongside their direct dependencies. This required the numerical evaluation of five integrals. Compared with using Riemann sums, we achieved substantial accuracy improvements at virtually no added numerical cost by approximating these integrals with the trapezoid rule and the composite Simpson's rule.

We summarise the likelihood function in Table E.1 and Fig. E.1. To obtain the posterior, we simply multiplied the likelihood function by the prior and normalised the result.

In Table 4, for each parameter, we list the maximum a posteriori probability (MAP) estimate, alongside estimates for the posterior mean and standard deviation. In Fig. 16, we visualise all one- and two-dimensional posterior marginals, in which we mark the MAP (white circle) and the posterior mean (white cross). The joint marginal for  $\xi$  and  $b_{\nu,\text{ref}}$  shows that these parameters have a strong negative correlation, indicating that with current data, the steep slope of the ESF at high  $l_p$  can equally be described with a steep intrinsic slope and mild surface brightness selection (i.e.  $\xi$  low and  $b_{\nu,\text{ref}}$  high), or by a shallow intrinsic slope and strong surface brightness selection (i.e.  $\xi$  high and  $b_{\nu,\text{ref}}$  low). We leave it up to future studies to break this degeneracy, either by using larger samples, by measuring  $b_{\nu,\text{ref}}$  directly,

<sup>18</sup> For another interesting case,  $b_{\nu,\text{ref}} = 500 \text{ Jy deg}^{-2}$  (not shown), the median  $L_\nu = 0.3 \times 10^{26} \text{ W Hz}^{-1}$ .

<sup>19</sup> The computation took a few thousand CPU hours to complete, but can be trivially distributed among nodes, and within a node among CPUs. Model extensions that introduce additional parameters shall necessitate more efficient inference techniques, such as Markov chain Monte Carlo.

**Table 4.** Maximum a posteriori probability (MAP) and posterior mean and standard deviation (SD) estimates of the free parameters in intrinsic GRG length distribution inference.

$z_{\text{max}} = 0.5$		
Parameter	MAP	Posterior mean and SD
$\xi$	-3.55	$-3.5 \pm 0.5$
$\frac{1}{2}(\phi_{\text{max}} - \phi_{\text{min}})$	1.9'	$1.9 \pm 0.2'$
$b_{\nu,\text{ref}}$	$600 \text{ Jy deg}^{-2}$	$720 \pm 380 \text{ Jy deg}^{-2}$
$\sigma_{\text{ref}}$	1.2	$1.2 \pm 0.2$
$z_{\text{max}} = 0.25$		
Parameter	MAP	Posterior mean and SD
$\xi$	-3.5	$-3.5 \pm 0.4$
$\frac{1}{2}(\phi_{\text{max}} - \phi_{\text{min}})$	1.85'	$1.7 \pm 0.3'$
$b_{\nu,\text{ref}}$	$900 \text{ Jy deg}^{-2}$	$1020 \pm 490 \text{ Jy deg}^{-2}$
$\sigma_{\text{ref}}$	1.15	$1.3 \pm 0.4$

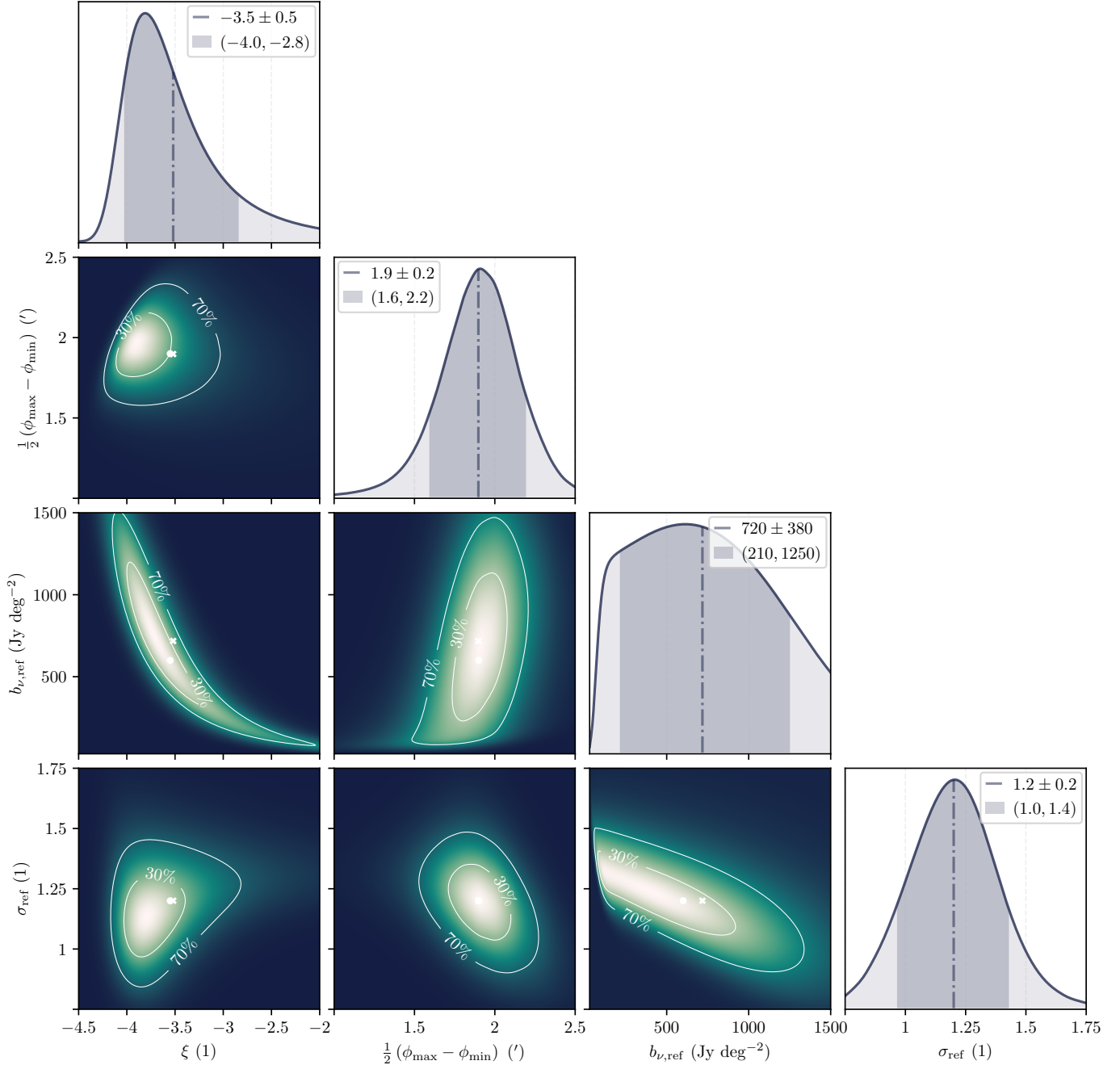
**Notes.** The model assumes  $\xi$  is constant for  $z \in [0, z_{\text{max}}]$ . We determined the posterior twice: for  $z_{\text{max}} = 0.5$ , using 1473 giants, and for  $z_{\text{max}} = 0.25$ , using 811 giants.

or by improving survey sensitivities so that surface brightness selection effect modelling becomes superfluous altogether.

#### 4.1.5. Goodness of fit

In both panels of Fig. 14, we compare the ESF and SF of  $L_{\text{p,obs}} | L_{\text{p,obs}} > l_{\text{p,GRG}}$  for  $z_{\text{max}} = 0.5$ , using the MAP parameters for the latter. The model appears able to produce a tight fit to the data. The mean and standard deviation of the ESF–SF residuals are 0.01% and 0.3%, whilst the mean and standard deviation of the absolute ESF–SF residuals are both 0.2%. Using a Kolmogorov–Smirnov test, we formally verified that our best parameters are indeed good parameters – in the sense that they represent a plausible model underlying the data. The Kolmogorov–Smirnov statistic is the maximum deviation between the ESF and SF, and equals 1% in our case. The p-value – the probability that an ESF–SF discrepancy of at least this magnitude would occur if the SF represents the true underlying distribution – is  $p = 99\%$ . For any reasonable significance level, we do not reject the null hypothesis. The model, given our best parameters, indeed represents a possible description of the data. We conclude that the distribution of GRG intrinsic proper lengths, after correcting for selection effects, is consistent with a single Pareto distribution with tail index  $\xi = -3.5$ . We show the SF of this distribution in both panels of Fig. 14 (fading green dots). For low  $l_p$ , the observed slope is shallower (due to angular length selection), whilst for high  $l_p$ , the observed slope is steeper (due to surface brightness selection).

A quasi-Pareto distribution can arise naturally as the tail of a lognormal distribution (e.g., Malevergne et al. 2011), and there are reasons to believe that the entire radio galaxy length distribution is indeed approximately lognormal (Oei et al., in prep.). This provides an explanation of the approximately Paretian nature of the giant radio galaxy length distribution found in this section. The specific value of the tail index  $\xi$  is set by both the physics of radio galaxy growth and the distribution of radio galaxies over large-scale environments, the latter of which we measure in Oei et al. (in prep.). Our result  $\xi = -3.5 \pm 0.5$  is a new constraint for dynamical models such as those of Turner & Shabala (2015) and Hardcastle (2018).



**Fig. 16.** Joint posterior distribution over  $\xi$  – the parameter of interest – and  $\frac{1}{2}(\phi_{\max} - \phi_{\min})$ ,  $b_{\nu,\text{ref}}$  and  $\sigma_{\text{ref}}$  – the selection effect parameters, based on 1473 projected lengths of LoTSS DR2 giants up to  $z_{\max} = 0.5$ . We show all two-parameter marginals of the posterior, with contours enclosing 30% and 70% of total probability. We mark the maximum a posteriori probability parameters (white circle) and the posterior mean parameters (white cross). The single-parameter marginals again show the estimated posterior mean, now marked by a vertical line, alongside shaded median-centred 80% credible intervals. To compare the posterior to the likelihood function, which is also the posterior for a uniform prior, see Fig. E.1.

#### 4.2. Giant radio galaxy number density

If in addition to our discoveries, we know how many giants our search campaign has missed, then we can infer the true comoving GRG number density in the Local Universe. The posterior distribution over selection effect parameters  $\frac{1}{2}(\phi_{\max} - \phi_{\min})$ ,  $b_{\nu,\text{ref}}$  and  $\sigma_{\text{ref}}$  induces a probability distribution over the search completeness function  $C(l_p)$ .  $C(l_p)$  denotes the probability that a giant of projected proper length  $l_p$  in comoving space up to  $z = z_{\max}$  is detected through the search. We first generated parameter samples from our posterior using rejection sampling,

and then used each to calculate a  $C(l_p)$  sample. We show the distribution over  $C(l_p)$  for  $z_{\max} = 0.5$  in Fig. 17. For small  $l_p$ ,  $C$  is low as many giants drop out due to angular length selection; for large  $l_p$ ,  $C$  is low as many giants drop out due to surface brightness selection. The completeness peaks around  $l_p \sim 2$  Mpc; however, even there the majority of giants remains undetected.

We inferred a probability distribution over the true comoving GRG number density  $n_{\text{GRG}}$  by combining Eqs. (18) and (30) with the LoTSS DR2 GRG catalogue and samples from Sect. 4.1.4’s posterior. The resulting skewed distribution, with mean and

SD  $n_{\text{GRG}} = 4.6 \pm 2.4(100 \text{ Mpc})^{-3}$  and 80% credible interval 3.1–6.7(100 Mpc) $^{-3}$ , is shown in Fig. 18. We note that, although the uncertainty in  $b_{\text{v,ref}}$  induces a large uncertainty in  $C$  from  $l_p \sim 1.5$  Mpc onwards, the completeness uncertainty at large projected lengths does not substantially contribute to the uncertainty in  $n_{\text{GRG}}$ . This is because the GRG population is dominated by smaller giants, for which the completeness appears better constrained.

What picture arises regarding the abundance of giant radio galaxies in the Local Universe’s large-scale structure? If we model the Cosmic Web through comoving cubic unit cells (Oei et al. 2022b) with 50 Mpc sides, and each cubic unit cell contributes one cluster and three filaments, then a cube with 100 Mpc sides features a total of eight clusters and 24 filaments. For comparison, in a (100 Mpc) $^3$  volume up to  $z_{\text{max}} = 0.5$ , the SDSS-III cluster catalogue of Wen et al. (2012) contains on average 11.2 clusters of any mass, and 4.5 clusters of mass  $M_{200} > 10^{14} M_{\odot}$ . Since clusters contain  $\sim 20\%$  of giants (Oei et al., in prep.), we find the average number of giants per cluster to be  $\sim 10^{-1}$ . If one assumes that filaments contain the remaining  $\sim 80\%$  of giants, and uses the fact that the average number of filaments per cluster is of order unity, it follows that the average number of giants per filament is also  $\sim 10^{-1}$ . In all likelihood, most clusters and filaments do not currently contain a giant.

#### 4.3. Giant radio galaxy lobe volume-filling fraction

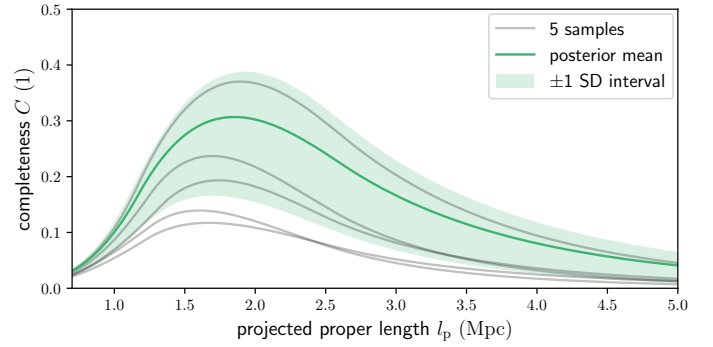
Because giant radio galaxies enrich the IGM with hot plasma and magnetic fields far beyond the circumgalactic media of their hosts, they may provide a meaningful contribution to the heating and magnetisation of – in particular – the most rarefied parts of the filament IGM. By combining the GRG number density and the GRG jet power distribution (e.g., Dabhade et al. 2020a), one could estimate the instantaneous heating and magnetisation contributions directly. We recommend such analysis for future research.

We evaluated Eq. (35) to obtain an estimate of the fraction of the Local Universe’s proper volume that GRG lobes occupy. We used Alcyoneus as a reference giant, for which  $V = 2.5 \pm 0.3 \text{ Mpc}^3$  and  $l_p = 4.99 \pm 0.04 \text{ Mpc}$  (Oei et al. 2022a); this suggests  $\mathbb{E}[\Upsilon_p] \approx 2\%$ . Future work should determine whether Alcyoneus’s case is typical, as observations, such as those shown in Figs. 12–13, suggest that giants exhibit a large variety of shapes – and thus total lobe volume–cubed length ratios. Interestingly, simulations by Krause et al. (2012) have found that these shapes also depend on environmental parameters such as ambient pressure and density. Truncating the GRG projected length distribution at  $l_{p,\text{max}} = 7 \text{ Mpc}$ , so that its support is exactly an order of magnitude, Eq. (35) predicts  $\text{VFF}_{\text{GRG}}(z = 0) = 3_{-1}^{+4} \times 10^{-7}$ .<sup>20</sup>

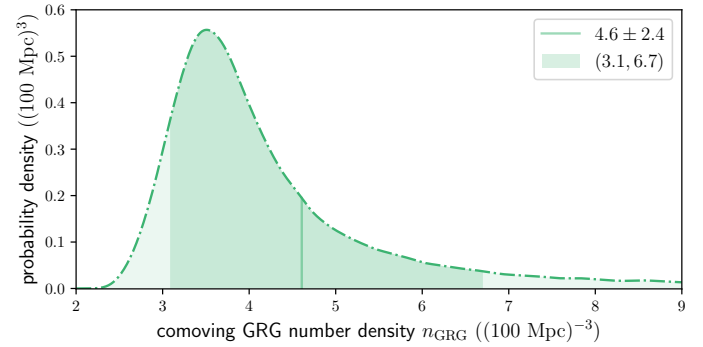
Whether this result is sensitive to changes in  $l_{p,\text{max}}$  depends on  $\xi$ , with  $\xi = -3$  being a special value under self-similar growth. In that case, small and large giants contribute equally to  $\text{VFF}_{\text{GRG}}$ : although large giants are rarer ( $f_{l_p} \propto l_p^{-3}$ ), their larger lobe volumes ( $V \propto l_p^3$ ) exactly compensate. For  $\xi < -3$ , small giants provide the dominant contribution to  $\text{VFF}_{\text{GRG}}$  and the choice of  $l_{p,\text{max}}$  can be irrelevant; for  $\xi > -3$ , large giants dominate and the choice of  $l_{p,\text{max}}$  always matters.

If we assume that giants occur in clusters and filaments only and use the fact that clusters and filaments comprise about 5% of the Local Universe’s volume (Forero-Romero et al. 2009), then

<sup>20</sup> We found the weaker constraints  $\text{VFF}_{\text{GRG}}(z = 0) > 13_{-4}^{+21} \times 10^{-8}$ ,  $\text{VFF}_{\text{GRG}}(z = 0) > 11_{-3}^{+16} \times 10^{-8}$ , and  $\text{VFF}_{\text{GRG}}(z = 0) > 6_{-1}^{+4} \times 10^{-8}$  using Eqs. (33), (A.100), and (A.101), respectively.



**Fig. 17.** Completeness  $C$  of a sample of giant radio galaxies up to cosmological redshift  $z_{\text{max}}$  as a function of projected proper length  $l_p$ . From samples of the posterior distribution, we infer the LoTSS DR2 GRG search campaign completeness up to  $z_{\text{max}} = 0.5$ . We show completeness curves for five randomly selected samples (grey) and for the posterior mean (dark green). We also show an interval around the completeness mean with the completeness standard deviation (SD) as the half-width (light green). The completeness peaks around  $l_p = 2 \text{ Mpc}$ .



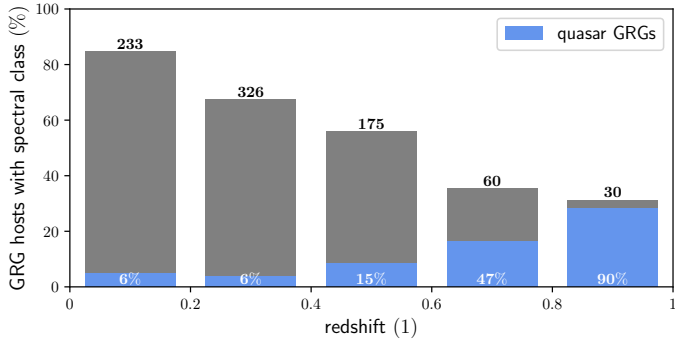
**Fig. 18.** PDF of the comoving GRG number density  $n_{\text{GRG}}$ . We mark the mean (vertical line) and the median-centred 80% credible interval (darker range). In the Local Universe, the average number of giants per comoving cube with 100 Mpc sides is  $4.6 \pm 2.4$ . We define giants through  $l_{p,\text{GRG}} = 0.7 \text{ Mpc}$ , and define the Local Universe to be up to cosmological redshift  $z_{\text{max}} = 0.5$ .

the GRG lobe VFF within clusters and filaments specifically is  $\text{VFF}_{\text{GRG}}(z = 0) = 5_{-2}^{+8} \times 10^{-6}$ . We conclude that, at each given moment, GRG lobes occupy just a small fraction of the WHIM and ICM. If the enrichment of the IGM by giants is to affect the WHIM and ICM on a large scale, mixing processes in the IGM are necessary and many galaxies must be able to form giants at some point in their evolution.

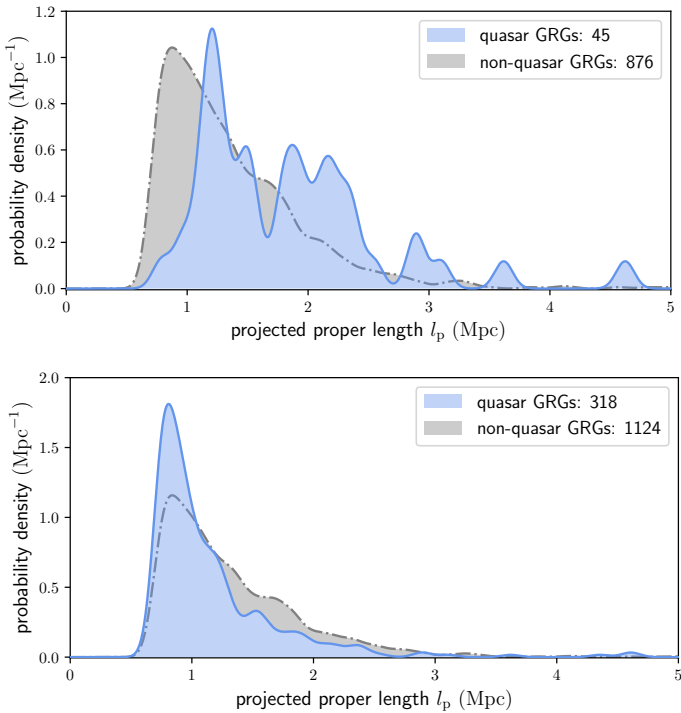
#### 4.4. Unification model constraints from quasar and non-quasar giants

In Sect. 2.11, we have predicted general ramifications of the basic unification model on a GRG sample. We constrained and tested this model with our LoTSS DR2 GRG sample.

First, for all 3198 giants with definitively identified hosts, we queried the SDSS DR12 spectral class spCl: a Boolean label indicating whether or not the host contains a quasar. As many hosts have no SDSS DR12 spectrum, or even fall outside of SDSS DR12 coverage, we retrieved host classifications for just 1442 of these giants (45%). Of these classified giants, 318 are quasar giants (22%) (of which 45 (14%) are discoveries presented in this work) and 1124 are non-quasar giants (78%) (of which 876 (78%) are discoveries presented in this work).



**Fig. 19.** Binary classification into quasar and non-quasar giants based on SDSS DR12 host spectra, for 5 redshift intervals. We selected all giants with definite hosts within the SDSS-covered sky patch bounded by right ascensions  $120^\circ$  and  $250^\circ$ , and declinations  $27^\circ$  and  $62^\circ$ . As redshift increases, the fraction of hosts with a known spectral class decreases (grey bars, with the absolute number of such hosts in black), whilst the fraction of quasar identifications increases (blue bars and white percentages).



**Fig. 20.** Observed projected proper length PDFs for SDSS-classified quasar and non-quasar giants, obtained through kernel density estimation. We used a Gaussian kernel with  $\sigma_{\text{KDE}} = 75$  kpc. For both panels, two-sample Kolmogorov–Smirnov tests yield  $p < 1\%$ . However, given the severe impact of selection effects, we could not reject the null hypothesis that quasar giants and non-quasar giants have the same underlying projected proper length distribution. *Top*: newly discovered (LoTSS DR2) giants with SDSS spectral class labels. *Bottom*: all known giants with SDSS spectral class labels.

Therefore, the apparent LoTSS DR2 quasar GRG fraction  $f_Q = \frac{45}{45+876} = 5\%$ . However, spectral class labels are preferentially available for GRG hosts with higher optical flux densities, such as those at low redshifts or those containing quasars, because they are more probable spectroscopic targets. Through Fig. 19, we demonstrate that the fraction of observed GRG hosts with spectral class labels indeed decreases with redshift, whilst the

fraction of quasar identifications increases. For each redshift interval, in white, we denote the fraction of quasar giants within the classified population. If spectral class labels would have been available for the non-classified populations as well, the quasar GRG fractions would probably have been lower. Assuming that all observed giants whose hosts have an unknown spectral class are non-quasar giants, we find quasar GRG fractions  $f_Q = 5\%$ ,  $4\%$ ,  $8\%$ ,  $16\%$ , and  $28\%$ , for redshift intervals 0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1, respectively. The true quasar GRG fraction for a given redshift interval might still differ from the aforementioned observed quasar GRG fraction: namely, if selection effects make a given quasar GRG easier (or harder) to detect than a given non-quasar GRG. At higher redshifts, quasar giants certainly appear easier to detect than non-quasar giants, as hosts without quasars often become too faint to optically identify. For the lowest redshift intervals, this problem does not exist, and we therefore consider the observed quasar GRG fraction  $f_Q = 5\%$  to be closest to the true one. A thorough analysis of the impact of selection effects on  $f_Q$  is a topic for future research.

Using Eq. (37), and assuming a quasar GRG probability  $p_Q = 5\%$ , we found a maximum inclination angle  $\theta_{\text{max}} = 39_{-3}^{+2}^\circ$ . For  $p_Q \sim \text{Uniform}(4\%, 6\%)$ , the result remained the same. In conclusion, if the basic unification model considered in Sect. 2.11 is correct, then observations of giants predict that quasars are AGN seen along lines-of-sight that make an angle of at most  $\theta_{\text{max}} = 39_{-3}^{+2}^\circ$  with the black hole rotation axis.

Finally, we tested whether the RV  $L_{\text{p,obs}} | L_{\text{p,obs}} \geq l_{\text{p,GRG}}$  has the same distribution for quasar and non-quasar giants, as predicted by the unification model. In the top panel of Fig. 20, we show PDFs approximated through kernel density estimation (KDE) for newly discovered (LoTSS DR2) SDSS-classified quasar giants and non-quasar giants up to  $z_{\text{max}} = \infty$ . Despite the small number of quasar giants,  $N_Q = 45$ , the two-sample Kolmogorov–Smirnov (KS) test yielded a low p-value  $p < 1\%$ : we rejected the null hypothesis that these distributions stem from a single underlying one. However, this does not mean that the unification model hypothesis should immediately be rejected as well, because the selective availability of SDSS DR12 spectral class labels induces a severe selection effect<sup>21</sup>. This effect can be tempered by choosing a lower  $z_{\text{max}}$ , but for choices such as  $z_{\text{max}} = 0.25$  and  $0.5$ , just  $N_Q = 10$  and  $16$  quasar giants remain – too few to extract meaningful information. The bottom panel of Fig. 20 again shows observed projected proper length KDE PDFs, but now for all SDSS-classified quasar giants and non-quasar giants up to  $z_{\text{max}} = \infty$ ; in this case,  $N_Q = 318$ . Again, the two-sample KS test yields  $p < 1\%$ , but this time quasar giants appear smaller than non-quasar giants. Because we aggregated samples here that have different selection effects imprinted, it becomes hard to draw clear conclusions. Limiting  $z_{\text{max}}$  reduces the severity of most selection effects, but of course comes at the cost of reducing the sample size. Interestingly, the corresponding quasar GRG and non-quasar GRG projected length distributions

<sup>21</sup> In fact, the top panel of Fig. 20 shows expected behaviour for a GRG search campaign with a (fuzzy) angular length threshold selection effect if the unification model is correct. At high redshifts, a GRG must have a larger projected length to pass the angular length threshold than at low redshifts. Thus, sampled high-redshift giants are physically larger than sampled low-redshift giants. Because the fraction of high-redshift quasar giants that is detectable and spectrally classifiable is higher than the fraction of high-redshift non-quasar giants that is detectable and spectrally classifiable, sampled quasar giants will be physically larger than sampled non-quasar giants. To draw this latter conclusion, we must also invoke the fact that, under the unification model, the projected length distributions of quasar and non-quasar giants are the same.

do become more alike; for instance for  $z_{\max} = 0.25$  and  $0.5$ , the two-sample KS test yields  $p = 4\%$  and  $1\%$ , respectively. In conclusion, because quasar giants are intrinsically rare and the availability of spectral class labels is biased towards low redshifts and hosts containing quasars, it is challenging to robustly test the unification model with current GRG observations. We refrain from drawing final conclusions, and recommend a careful future analysis.

## 5. Discussion

### 5.1. Radio galaxy length definitions

How large are radio galaxies? Despite the simplicity of this question and more than half a century of research on radio galaxies, their intrinsic length distribution has not yet been rigorously characterised. In this work, we have carried out the first precision analysis of the tail of the radio galaxy intrinsic length distribution. Precision analyses tend to raise questions; firstly, whether the studied observable is well defined, and secondly whether one could conceive of more informative observables – that is to say those that make it easier to reveal underlying physical mechanisms. This work’s main observable is the radio galaxy projected proper length; we argue that it is neither well defined nor maximally informative.

#### 5.1.1. The current length definition: survey-dependence

Contemporary research uses a survey-dependent definition for radio galaxy angular lengths, which then makes projected proper lengths survey-dependent too.

The angular length is canonically defined as the largest possible angular separation between two directions for which the RG’s specific intensity function  $I_{\nu, \text{RG}}$  exceeds  $b_{\nu, \text{th}}$ : some specified factor of order unity times the image noise  $\sigma_{I_\nu}$ . A complication is that not only  $I_{\nu, \text{RG}}$ , but also  $\sigma_{I_\nu}$  varies with observing frequency; the latter because of observational factors such as  $(u, v)$ -coverage, bandpass, radio-frequency interference (RFI), ionospheric weather, the sky density of bright calibrators and the performance of calibration algorithms.  $I_{\nu, \text{RG}}$  additionally depends on resolution, at least for point sources;  $\sigma_{I_\nu}$  additionally depends on resolution and integration time. As a result, both  $I_{\nu, \text{RG}}$  and the concrete value of  $b_{\nu, \text{th}}$  used in the angular length definition change from image to image. Each study thus far has therefore implicitly used a different definition for angular length, instead of a shared, absolute one. As the projected proper length follows from combining the angular length with the host redshift, it suffers from the same problem.

Whether the survey-dependence of the current length definition is problematic, depends on the radio galaxy. For archetypal FRII RGs, the angular length is roughly equal to the angular distance between the hotspots. These are an FRII RG’s brightest morphological components, and are thus the first to be picked up by a survey. In contrast, archetypal FRI RGs, which gradually fade with distance from the host, can have significantly larger angular lengths in surveys of higher sensitivity. The giants in the middle-right and bottom-left panel of Fig. 13 are good examples: these radio galaxies were known before the LoTSS DR2, but were not known to be giants; similarly, more sensitive surveys are poised to assign them even larger extents. If we are to move towards precision science, it therefore makes sense – at least for FRI RGs – to more explicitly recognise that the angular and projected proper lengths are functions of the observing frequency  $\nu_{\text{obs}}$  and a surface brightness threshold  $b_{\nu, \text{th}}$ . If cata-

logues would explicitly state for what combination  $(\nu_{\text{obs}}, b_{\nu, \text{th}})$  they provide  $\phi = \phi(\nu_{\text{obs}}, b_{\nu, \text{th}})$  and  $l_p = l_p(\nu_{\text{obs}}, b_{\nu, \text{th}})$ , it is possible to homogenise a collection of data sets by using universal angular and projected proper length definitions for all RGs<sup>22</sup>. If two angular lengths have been measured for the same RG, for instance  $\phi_1$  at  $(\nu_{\text{obs},1}, b_{\nu, \text{th},1})$  and  $\phi_2$  at  $(\nu_{\text{obs},2}, b_{\nu, \text{th},2})$ , then we can estimate  $\phi$  for any desired  $(\nu_{\text{obs}}, b_{\nu, \text{th}})$  through interpolation. For example, the interpolation formula for a symmetric radio galaxy with jets or lobes of constant spectral index  $\alpha$ , and with a specific intensity function contribution which fades to zero linearly with angular distance from the host, is

$$\phi = \max \left\{ \phi_1 + \frac{b_{\nu, \text{th}} - b_{\nu, \text{th},1} \left( \frac{\nu_{\text{obs}}}{\nu_{\text{obs},1}} \right)^\alpha}{b_{\nu, \text{th},2} \left( \frac{\nu_{\text{obs}}}{\nu_{\text{obs},2}} \right)^\alpha - b_{\nu, \text{th},1} \left( \frac{\nu_{\text{obs}}}{\nu_{\text{obs},1}} \right)^\alpha} (\phi_2 - \phi_1), 0 \right\}. \quad (41)$$

#### 5.1.2. The ideal length definition: physical relevance

Alcyoneus, shown in Fig. 12’s top-right panel, is a 5 Mpc giant whose ageing lobes are revealed for the first time by the LoTSS DR2 (Oei et al. 2022a). Future image sensitivity improvements shall reveal more hitherto unseen, fading lobes around known RGs that formed in the aftermath of earlier AGN activity episodes. Could a large fraction of sufficiently old RGs turn out to be giants, once such sensitivity improvements start providing evidence of earlier and earlier AGN activity episodes?

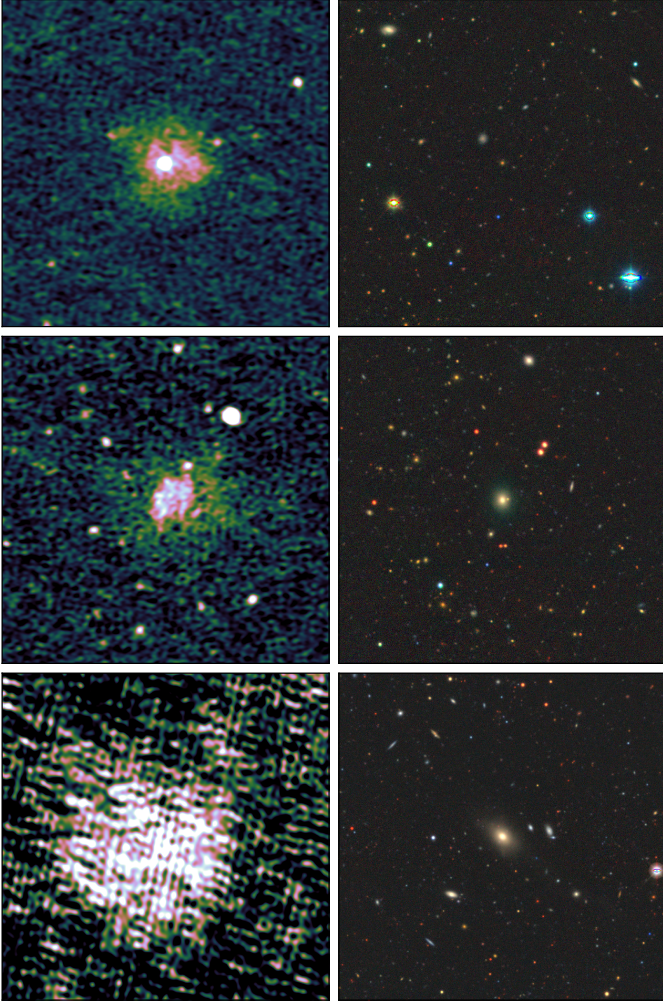
Up to now, it has been informative to include all visible plasma in the angular length measurement. In future images, some of the visible plasma might be of such low pressure that it has become physically insignificant, in the sense that it does not affect the thermodynamics of the surrounding IGM anymore; we propose to exclude such plasma from a radio galaxy length. Practically, one option is to introduce an absolute threshold: for example to include plasma of pressure  $P \geq 10^{-17}$  Pa only – this is the pressure of the warm-hot intergalactic medium (WHIM) in the  $\rho_{\text{BM}} \sim 10 \Omega_{\text{BM},0} \rho_{\text{c},0}$  and  $T \sim 5 \times 10^5$  K regime. Another option is to introduce an environment-dependent pressure threshold; this would mean that a cluster RG sees its length measured against a higher pressure threshold than a filament RG, because its plasma becomes thermodynamically irrelevant sooner. A problem with (equipartition or minimum energy) pressure-based length definitions is that pressure is a derived quantity: the specific intensity only determines the product of pressure and line-of-sight length through the lobe.

### 5.2. Moving beyond line segment projection

In this work, we have adopted a classical approach to treating projection, by modelling radio galaxies as line segments. When a radio galaxy’s inclination angle  $\theta$  is close to  $0^\circ$  or  $180^\circ$ , this approach predicts that the angular length  $\phi$  and thus the projected proper length  $l_p$  vanishes. In reality however, because radio galaxy lobes have non-zero volumes, the RV  $L_p$  will never tend to zero. Figure 21 illustrates this point through three sources interpreted to be radio galaxies aligned closely with the line of sight. In each case,  $\phi$  remains of arcminute scale.

A more realistic approach, outlined in Appendix A.3, moves beyond the simplistic line segment geometry by adding two lobes of radius  $R$  to the endpoints of the line segment, whose length is given by RV  $L$ . The ratio  $\frac{L}{l_p}$ , which was unbounded

<sup>22</sup> Our LoTSS DR2 GRG  $\phi$  and  $l_p$  correspond to  $b_{\nu, \text{th}} := 1 \times \sigma_{I_\nu}$ ; on average,  $\sigma_{I_\nu}(\nu_{\text{obs}} = 144 \text{ MHz}, \theta_{\text{FWHM}} = 6'') = 25 \text{ Jy deg}^{-2}$ .



**Fig. 21.** Suspected line-of-sight RGs in the radio and optical. We show  $5' \times 5'$  LoTSS DR2  $6''$  cutouts (*left column*) and corresponding DESI Legacy Imaging Surveys ( $g, r, z$ )-cutouts (*right column*). Due to its lobes, an RG whose axis aligns closely to the line of sight has a non-zero projected length. From top to bottom row, the SDSS host names are J125027.47+642034.3, J092220.72+560234.9, and J023100.61+032922.1.

under the classical approach, is now at most  $1 + \frac{1}{\eta}$ , where  $\eta := \frac{2R}{L}$ . (The classical approach simply is the limit  $\eta = 0$ .) We suggest a recalculation of this work's theoretical and applied results under this more realistic RG geometry as a direction for future research. To obtain the applied results, one must either fix  $\eta$  as a hyperparameter, or include it as an additional model parameter. We note that even using conservatively low values of  $\eta$ , such as  $\eta = \frac{1}{10}$ , will represent an improvement in realism over  $\eta = 0$ .

### 5.3. Unmodelled selection effects

In this work, we have modelled both an angular length and a surface brightness selection effect. Several other plausible selection effects have not been included in the forward model, as we have judged each to be of minor importance. However, in unison, they could have a non-negligible influence on the distribution of the observed projected proper GRG length RV  $L_{p,obs} | L_{p,obs} > l_{p,GRG}$ . Some of their influence might have been absorbed by the parameters of the included selection effects – that is by  $\frac{1}{2}(\phi_{max} - \phi_{min})$ ,  $b_{v,ref}$ , and  $\sigma_{ref}$  – or, worse still, by  $\xi$ .

One of these unmodelled selection effects is that RGs whose axes are oriented almost parallel to the line of sight are more likely to be rejected from a sample than RGs whose axes are closer to the plane of the sky, as the former do not always have a characteristic double-lobe appearance. For instance, perhaps not all readers would regard our identification of the sources in the left column of Fig. 21 as RGs convincing. Nevertheless, as remarked in Sect. 2.11, conditioning  $L_{p,obs} | L_{p,obs} > l_{p,GRG}$  on inclination angle does not affect its distribution. As a result, this selection effect does not necessitate forward model modifications.

Furthermore, as shown in Sect. 4.4, there is a selection effect at play that favours the selection of quasar giants over non-quasar giants at high redshift, as host galaxies with quasars are more luminous in the optical and therefore have a better chance to be picked up in optical imagery. This selection effect will get less severe once deeper photometric surveys become available.

We encountered three more selection effects during our LoTSS DR2 GRG search. The larger an RG – and especially an FR II RG – becomes, the harder it is for an observer to identify its host galaxy, as an increasing number of plausible host candidates can lie interspersed in the strip of sky between the lobes. The severity of this effect, which diminishes the prevalence of the largest giants in a sample, depends on the balance chosen between avoiding false discoveries and avoiding rejections of true discoveries. Another selection effect runs against RGs in galaxy clusters. Such environments can contain multiple adjacent RGs, making it at times unclear which lobe belongs to which RG. When no confident double-lobe associations can be made, the RGs involved fail to make it into the sample. If galaxy clusters contain primarily smaller giants, this selection effect induces a bias against smaller giants. A final unmodelled bias runs against RGs at the end of their life cycle. Once the AGN stops launching jets for a prolonged period, it becomes hard to identify the host galaxy, which will no longer present as a bright, compact radio source. This effect preferentially deselects larger giants, which are even more likely to approach the end of their life than smaller giants.

### 5.4. Does the choice of prior matter?

We have taken a conservative approach to constraining the posterior distribution through the prior: we have left tail index  $\xi$  and angular length selection half-width  $\frac{1}{2}(\phi_{max} - \phi_{min})$  fully unconstrained, and have adopted wide Gaussian priors for reference surface brightness parameters  $b_{v,ref}$  and  $\sigma_{ref}$ , despite measuring them explicitly under assumptions. We provide summary statistics of the posterior in Table 4 and visualise its one- and two-parameter marginals in Fig. 16. Does our choice of prior significantly affect the inferences? To explore this question, we chose a different reasonable prior and compared results. One such prior is the fully uniform prior, which equalises the posterior and the likelihood function. We provide analogous summary statistics of the likelihood function in Table E.1 and visualise analogous marginals in Fig. E.1. Reassuringly, no statistically significant parameter changes occur. In particular, for  $z_{max} = 0.5$ ,  $\xi = -3.5 \pm 0.5$  becomes  $\xi = -3.4 \pm 0.5$  upon changing to the uniform prior; for  $z_{max} = 0.25$ ,  $\xi = -3.5 \pm 0.4$  even remains the same. However, given the strong likelihood degeneracy between  $\xi$  and  $b_{v,ref}$  apparent in Fig. E.1, more stringent priors on  $b_{v,ref}$  are able to meaningfully shift  $\xi$ 's posterior mean. Such priors shall be appropriate only after studying the surface brightness properties of large radio galaxies – and the associated selection effect – in more detail.



### 5.5. Cosmological evolution of the GRG length distribution

In this work we have assumed that the parameter  $\xi$ , which fully characterises the intrinsic GRG length distribution under the ansatz of Paretianity, remains constant throughout the chosen redshift range  $[0, z_{\max}]$ . To test this assumption, we have analysed our LoTSS DR2 GRG sample in Sect. 4 up to both  $z_{\max} = 0.5$  and  $z_{\max} = 0.25$ . For  $z_{\max} = 0.5$ , we included giants that existed in the last 5 Gyr of the Universe's history, and found  $\xi = -3.5 \pm 0.5$ . Meanwhile, for  $z_{\max} = 0.25$ , we included giants that existed in the last 3 Gyr of the Universe's history only, and found the very similar  $\xi = -3.5 \pm 0.4$ . Thus, our analysis did not produce evidence that  $\xi$  evolves over cosmic time. However, given the large error bars, a modest time evolution cannot be excluded. Furthermore, the data sets that underlie these inferences are not disjoint: the 811 giants that inform the lower-maximum-redshift analysis make up 55% of the 1473 giants that inform the higher-maximum-redshift analysis.

Whether a time evolution of  $\xi$  is expected is presumably tied to whether giant radio galaxy growth varies with environmental density at a given epoch, because the combined effects of the Universe's expansion and ongoing large-scale structure formation can similarly change environmental density. Combining our LoTSS DR2 GRG sample with Cosmic Web reconstructions to explore giant growth as a function of environmental density is the topic of a forthcoming work (Oei et al., in prep.). Interestingly, the recent exploration by Lan & Prochaska (2021) that compared the environments of giants and non-giants did not find significant differences.

The most straightforward model extension is to again assume that RG lengths are Pareto distributed with tail index  $\xi$ , but now  $\xi = \xi(z)$ . This function's first-degree Maclaurin polynomial, which provides the linearisation at the present day, is

$$\xi(z) \approx \xi(z=0) + \frac{d\xi}{dz}(z=0) \cdot z. \quad (42)$$

One would adopt  $\xi(z=0)$  and  $\frac{d\xi}{dz}(z=0)$  as the parameters of interest, replacing what used to be a constant  $\xi$ ; the number of model parameters would thus increase by one. However, an attempt to infer the cosmic evolution of  $\xi$  appears promising only once the major selection effects are better constrained.

## 6. Conclusions

In this work, we have performed Bayesian inference on a LoTSS-derived sample of 2060 giant radio galaxy projected proper lengths, using a one-parameter model that assumes a spatially homogeneous, non-evolving population of radio galaxies with stick-like geometry, Pareto-distributed lengths, and isotropic inclination angles. Before fitting to data, we extended the forward model with two selection effects typical of contemporary manual GRG search campaigns. The best-fit survival function tightly reproduces the empirical one, leaving permille-scale absolute residuals. Having quantified the most important selection effects, we estimated the true comoving giant radio galaxy number density in the Local Universe.

1. We developed an analytical model through which statistical questions about radio galaxy (RG) lengths can be rigorously answered. In the current work, we applied this model to giant radio galaxies. We adopted the ansatz that the RG intrinsic proper length  $L$ , as measured in three spatial dimensions, is a random variable (RV) with a Pareto Type I distribution (i.e. a simple power-law distribution) characterised

by tail index  $\xi$ . Next, by assuming that RGs have no preferential orientation with respect to the observer, we derived the distribution of the RG projected proper length  $L_p$ . By conditioning, one obtains the version relevant for giants,  $L_p | L_p > l_{p,\text{GRG}}$  (where we chose  $l_{p,\text{GRG}} := 0.7 \text{ Mpc}$ ). This RV is again Paretian, with the same tail index  $\xi$ . In summary, for giant radio galaxies, projection retains Paretianity. Finally, observers face selection effects; we modelled the observed projected proper length  $L_{p,\text{obs}}$  by considering an angular length threshold selection effect and a surface brightness selection effect. The angular length threshold selection effect assumes a linearly increasing angular-length-dependent probability of sample inclusion around a particular pre-defined threshold, meant to emulate manual visual searches that only target RGs of some angular length and above. The surface brightness selection effect assumes that giants are self-similar, and that their lobes have lognormally distributed surface brightnesses which must be above-noise to secure sample inclusion. The GRG observed projected proper length  $L_{p,\text{obs}} | L_{p,\text{obs}} > l_{p,\text{GRG}}$  again follows through conditioning. We assumed our data to be realisations of this RV.

2. The model also yielded explicit expressions for the (posterior) distribution of  $L | L_p = l_p$ . This allows one to deproject RGs in a statistical sense, providing the intrinsic proper length given the projected proper length in the limit of negligible selection effects. We also present practical expressions for the mean and variance of  $L | L_p = l_p$ . To unravel the driving factors that allow some RGs to become giants, most authors search for correlations between host or environmental physical parameters and the GRG projected length. However, if there is a causal chain that connects host or environmental parameters to GRG length, the connection will be to the intrinsic length; the observer's vantage point does not play a role in the physics. Therefore, the projection effect merely serves as a multiplicative noise source. We suggest that future analyses should recognise the projection effect as such, and correlate host or environmental parameters with the intrinsic, rather than projected, proper length – using statistical deprojection.
3. Through a manual visual search of LoTSS DR2 pipeline products, which are part of the LOFAR's Northern Sky survey at 144 MHz, we discovered a population of 2060 previously unknown giants. This is the largest single contribution to the literature yet, and increases the community-wide census by a factor 2.6. We present 11 discoveries with  $l_p \geq 4 \text{ Mpc}$ , 53 with  $3 \leq l_p < 4 \text{ Mpc}$ , 291 with  $2 \leq l_p < 3 \text{ Mpc}$ , 1215 with  $1 \leq l_p < 2 \text{ Mpc}$ , and 490 with  $0.7 \leq l_p < 1 \text{ Mpc}$ . Our study extends the known breadth of the giant radio galaxy phenomenon. Among the findings are both the giant hosted by J081956.41+323537.6 and Alcyoneus (Oei et al. 2022a), at  $l_p = 5.1 \text{ Mpc}$  and  $l_p = 5.0 \text{ Mpc}$  the projectively largest giants ever found. We discover that multi-Mpc radio galaxies can be generated before redshift 1, despite the Universe's mean density being an order of magnitude higher, and by spiral galaxies, whose stellar masses are typically an order of magnitude lower than those of ellipticals. We discover giants whose hosts have a record-low stellar mass  $M_\star = 4.8 \times 10^{10} M_\odot$ . We also discover giants whose hosts have a record-high supermassive black hole mass  $M_\bullet \gtrsim 5 \times 10^{10} M_\odot$ ; interestingly, with  $l_p = 0.8 \text{ Mpc}$ , one of these giants is relatively small. We more than double the number of known giants with angular lengths exceeding that of the Moon; one discovery, at  $2.2^\circ$ , is the angularly largest

radio galaxy in the Northern Sky and the angularly largest giant overall. Excitingly, our LoTSS DR2 search has been far from exhaustive: many thousands of readily identifiable giants still await discovery in this public data set.

4. Using our LoTSS DR2 GRG sample up to  $z_{\max} = 0.5$ , we generated a posterior distribution over  $\xi$  and three selection effect parameters. Our model provides an excellent fit to the data, with absolute residuals being on average 2%. We inferred that the intrinsic proper length distribution of the largest radio galaxies resembles a Pareto distribution with tail index  $\xi = -3.5 \pm 0.5$ . Our analysis did not yield evidence for an evolving  $\xi$  in the last 5 Gyr of cosmic time.
5. The selection effect parameters estimated through the posterior are far from nuisance parameters, as they allowed us to statistically undo the selection effects imprinted on our LoTSS DR2 GRG data. As a result, we could for the first time estimate the true comoving giant radio galaxy number density  $n_{\text{GRG}}$  in the Local Universe up to  $z_{\max}$ . We relied on the crucial assumption that the surface brightness distribution of RGs with intrinsic length  $l_{\text{ref}} = 0.7 \text{ Mpc}$  at  $z = 0$  and frequency  $\nu_{\text{obs}} = 144 \text{ MHz}$  is unimodal – and lognormal in particular. We furthermore assumed a lobe spectral index  $\alpha = -1$  and self-similar growth. We found  $n_{\text{GRG}}(l_{\text{p,GRG}} = 0.7 \text{ Mpc}, z_{\max} = 0.5) = 5 \pm 2(100 \text{ Mpc})^{-3}$ . The implication is that giant radio galaxies are truly rare – not only from a current observational perspective, but also from a cosmological one. Current GRG lobes occupy just a few millionths of the IGM volume. At any given moment in time, most clusters and filaments – the building blocks of modern large-scale structure – do not harbour giants.

Giants embody the most extreme known mechanism by which galaxies can affect the Cosmic Web around them. Whereas this work has explored the geometric properties of giants, a thorough exploration of their Cosmic Web energisation and magnetisation potential is a future frontier. Excitingly, the interactions between giants and the ethereal intergalactic medium may also allow for new constraints on the thermodynamics in filaments.

*Acknowledgements.* M.S.S.L. Oei warmly thanks Frits Sweijen for coding the very useful <https://github.com/tikk3r/legacystamps> and for tireless ICT advice, and Aleksandar Shulevski for comments that have improved the manuscript. In dear memory of Nicoline. By staying strong despite life's challenges, you showed what it means to be a giant. M.S.S.L. Oei, R.J. van Weeren, and A. Botteon acknowledge support from the VIDI research programme with project number 639.042.729, which is financed by The Netherlands Organisation for Scientific Research (NWO). A. Drabent acknowledges support by the BMBF Verbundforschung under the grant 05A20STA. The LOFAR is the Low-Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and which are collectively operated by the 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; the Istituto Nazionale di Astrofisica (INAF), Italy. This research made use of the Dutch national e-infrastructure with support of the SURF Cooperative (e-infra 180169) and the LOFAR e-infra group. The Jülich LOFAR Long Term Archive and the German LOFAR network are both coordinated and operated by the Jülich Supercomputing Centre (JSC), and computing resources on the supercomputer JUWELS at JSC were provided by the Gauss Centre for Supercomputing e.V. (grant CHTB00) through the John von Neumann Institute for Computing (NIC). This research made use of the University of Hertfordshire high-performance computing facility and the LOFAR-UK computing facility located at the University of Hertfordshire and supported by STFC [ST/P000096/1], and of the Italian LOFAR IT computing infrastructure supported and operated by INAF, and by the Physics Department of the University of Turin (under an agreement with Consorzio Interuniversitario

per la Fisica Spaziale) at the C3S Supercomputing Centre, Italy. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. 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 Astrofísica 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 Pan-STARRS1 Surveys (PS1) and the PS1 public science archive 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, 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, Eötvös Loránd University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID #2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing–Arizona Sky Survey (BASS; NOAO Prop. ID #2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z-band Legacy Survey (MzLS; Prop. ID #2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministerio da Ciencia, Tecnologia e Inovacao, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. BASS is a key project of the Telescope Access Program (TAP), which has been funded by the National

Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program “The Emergence of Cosmological Structures” Grant # XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Program of Chinese Academy of Sciences (Grant # 114A11KY5B20160057), and Chinese National Natural Science Foundation (Grant # 11433005). The Legacy Survey team makes use of data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration. The Legacy Surveys imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; and by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to NOAO.

## References

- Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, *ApJS*, **203**, 21
- Alam, S., Albareti, F. D., Prieto, C. A., et al. 2015, *ApJS*, **219**, 12
- Amirkhanyan, V. R. 2016, *Astrophys. Bull.*, **71**, 384
- Andernach, H., Jiménez-Andrade, E. F., & Willis, A. G. 2021, *Galaxies*, **9**, 99
- Bagchi, J., Vivek, M., Vikram, V., et al. 2014, *ApJ*, **788**, 174
- Bassani, L., Ursini, F., Malizia, A., et al. 2021, *MNRAS*, **500**, 3111
- Beckmann, R. S., Devriendt, J., Slyz, A., et al. 2017, *MNRAS*, **472**, 949
- Blandford, R. D., & Rees, M. J. 1974, *MNRAS*, **169**, 395
- Bloom, J. S., Frail, D. A., & Sari, R. 2001, *AJ*, **121**, 2879
- Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, *A&AS*, **143**, 33
- Brüggen, M., Reiprich, T. H., Bulbul, E., et al. 2021, *A&A*, **647**, A3
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05560]
- Chang, Y.-Y., van der Wel, A., da Cunha, E., & Rix, H.-W. 2015, *ApJS*, **219**, 8
- Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, *AJ*, **115**, 1693
- Cooper, B. F. C., Price, R. M., & Cole, D. J. 1965, *Aust. J. Phys.*, **18**, 589
- Croton, D. J., Springel, V., White, S. D. M., et al. 2006, *MNRAS*, **365**, 11
- Dabhade, P., Gaikwad, M., Bagchi, J., et al. 2017, *MNRAS*, **469**, 2886
- Dabhade, P., Mahato, M., Bagchi, J., et al. 2020a, *A&A*, **642**, A153
- Dabhade, P., Röttgering, H. J. A., Bagchi, J., et al. 2020b, *A&A*, **635**, A5
- Dálya, G., Galgóczi, G., Dobos, L., et al. 2018, *MNRAS*, **479**, 2374
- Delhaize, J., Heywood, I., Prescott, M., et al. 2021, *MNRAS*, **501**, 3833
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, **157**, 168
- Di Matteo, T., Springel, V., & Hernquist, L. 2005, *Nature*, **433**, 604
- Fabian, A. C., Nulsen, P. E. J., & Canizares, C. R. 1984, *Nature*, **310**, 733
- Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., & Yepes, G. 2009, *MNRAS*, **396**, 1815
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Gaia Collaboration (Brown, A. G. A., et al.) 2021, *A&A*, **649**, A1
- Galvin, T. J., Huynh, M. T., Norris, R. P., et al. 2020, *MNRAS*, **497**, 2730
- Gürkan, G., Prandoni, I., O’Brien, A., et al. 2022, *MNRAS*, **512**, 6104
- Hardcastle, M. J. 2018, *MNRAS*, **475**, 2768
- Hardcastle, M. J., & Croston, J. H. 2020, *New A Rev.*, **88**, 101539
- Hart, R. E., Bamford, S. P., Willett, K. W., et al. 2016, *MNRAS*, **461**, 3663
- Ishwara-Chandra, C. H., Taylor, A. R., Green, D. A., et al. 2020, *MNRAS*, **497**, 5383
- Jansen, R. A., Franx, M., Fabricant, D., & Caldwell, N. 2000, *ApJS*, **126**, 271
- Jasche, J., & Lavaux, G. 2019, *A&A*, **625**, A64
- Jasche, J., & Wandelt, B. D. 2013, *MNRAS*, **432**, 894
- Jasche, J., Leclercq, F., & Wandelt, B. D. 2015, *JCAP*, **2015**, 036
- King, A. 2016, *MNRAS*, **456**, L109
- Kirk, J. G., & Schneider, P. 1987, *ApJ*, **315**, 425
- Kormendy, J., & Ho, L. C. 2013, *ARA&A*, **51**, 511
- Krause, M., Alexander, P., Riley, J., & Hopton, D. 2012, *MNRAS*, **427**, 3196
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- Kuźmicz, A., & Jamroz, M. 2021, *ApJS*, **253**, 25
- Lan, T.-W., & Prochaska, J. X. 2021, *MNRAS*, **502**, 5104
- Mahato, M., Dabhade, P., Saikia, D. J., et al. 2022, *A&A*, **660**, A59
- Malevergne, Y., Pisarenko, V., & Sornette, D. 2011, *Phys. Rev. E*, **83**, 036111
- Masini, A., Celotti, A., Grandi, P., Moravec, E., & Williams, W. L. 2021, *A&A*, **650**, A51
- McKinley, B., Tingay, S. J., Gaspari, M., et al. 2022, *Nat. Astron.*, **6**, 109
- McNamara, B. R., & Nulsen, P. E. 2012, *New J. Phys.*, **14**, 055023
- Mostert, R. I. J., Duncan, K. J., Röttgering, H. J. A., et al. 2021, *A&A*, **645**, A89
- Oei, M. S. S. L., van Weeren, R. J., Hardcastle, M. J., et al. 2022a, *A&A*, **660**, A2
- Oei, M. S. S. L., van Weeren, R. J., Vazza, F., et al. 2022b, *A&A*, **662**, A87
- Oei, M. S. S. L., van Weeren, R. J., Hardcastle, M. J., et al. 2023, *MNRAS*, **518**, 240
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Proctor, D. D. 2016, *ApJS*, **224**, 18
- Ringermacher, H. I., & Mead, L. R. 2009, *MNRAS*, **397**, 164
- Rybicki, G. B., & Lightman, A. P. 1986, *Radiative Processes in Astrophysics* (Wiley-VCH), 400
- Salim, S., Boquien, M., & Lee, J. C. 2018, *ApJ*, **859**, 11
- 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
- Shimwell, T. W., Hardcastle, M. J., Tasse, C., et al. 2022, *A&A*, **659**, A1
- Simonte, M., Andernach, H., Brüggen, M., et al. 2022, *MNRAS*, **515**, 2032
- Solovoyov, D. I., & Verkhodanov, O. V. 2011, *Astrophys. Bull.*, **66**, 416
- Solovoyov, D. I., & Verkhodanov, O. V. 2014, *Astrophys. Bull.*, **69**, 141
- Tang, H., Scaife, A. M. M., Wong, O. I., et al. 2020, *MNRAS*, **499**, 68
- Turner, R. J., & Shabala, S. S. 2015, *ApJ*, **806**, 59
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, **556**, A2
- Vazza, F., Brüggen, M., Gheller, C., et al. 2017, *CQG*, **34**, 234001
- Wen, Z. L., Han, J. L., & Liu, F. S. 2012, *ApJS*, **199**, 34
- Yang, H. Y. K., Gaspari, M., & Marlow, C. 2019, *ApJ*, **871**, 6
- Zheng, H., Tegmark, M., Dillon, J. S., et al. 2017, *MNRAS*, **464**, 3486
- Zou, H., Sui, J., Xue, S., et al. 2022, *Res. Astron. Astrophys.*, **22**, 065001

## Appendix A: Framework derivations and details

### A.1. Intrinsic proper length

Let  $f_L : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  be the PDF of the distribution of intrinsic (i.e. 3D) proper (i.e. not comoving) radio galaxy lengths  $L$ . We assume that  $f_L$  follows a power law between  $l_{\min}$  and  $l_{\max}$ :

$$f_L(l) = \begin{cases} f_L(l_{\text{ref}}) \left(\frac{l}{l_{\text{ref}}}\right)^\xi & l \in [l_{\min}, l_{\max}], \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.1})$$

Then, when  $\xi \neq -1$ ,

$$\begin{aligned} 1 &= \int_{\mathbb{R}} f_L(l) \, dl = \int_{l_{\min}}^{l_{\max}} f_L(l_{\text{ref}}) \left(\frac{l}{l_{\text{ref}}}\right)^\xi \, dl \\ &= \frac{f_L(l_{\text{ref}}) l_{\text{ref}}}{\xi + 1} \left( \left(\frac{l_{\max}}{l_{\text{ref}}}\right)^{\xi+1} - \left(\frac{l_{\min}}{l_{\text{ref}}}\right)^{\xi+1} \right). \end{aligned} \quad (\text{A.2})$$

This equation provides the normalisation factor  $f_L(l_{\text{ref}})$  given  $l_{\min}$ ,  $l_{\max}$ ,  $\xi$  and an arbitrary choice for  $l_{\text{ref}} \neq 0$ .

When  $l_{\max} = \infty$ ,  $L$  has a Pareto Type I distribution and we must have  $\xi < -1$  for the integrals of Eq. A.2 to converge. Throughout the remaining analysis, for the sake of simplicity, we assume  $l_{\max} = \infty$  and  $\xi < -1$  and choose  $l_{\text{ref}} = l_{\min}$ . Then  $f_L(l_{\text{ref}}) = f_L(l_{\min}) = -\frac{\xi+1}{l_{\min}}$ . The corresponding CDF is

$$F_L(l) = \begin{cases} 0 & \text{if } l \leq l_{\min} \\ 1 - \left(\frac{l}{l_{\min}}\right)^{\xi+1} & \text{if } l > l_{\min}. \end{cases} \quad (\text{A.3})$$

In reality, radio galaxies cannot become arbitrarily long (Hardcastle 2018): at some distance from the host, all energy carried by the jets will have been radiated away, used to perform work on the IGM, transferred to CMB photons through inverse Compton scattering, or converted into heat. However, at the moment of writing, the implied maximum length  $l_{\max}$  remains ill-constrained. As we see throughout Appendix A, a major advantage of simple model assumptions is that explicit and thus insightful analytic expressions can be derived. Such easy-to-evaluate expressions complement the results of expensive numerical simulations, which aim to maximise realism; our aim to maximise insight is best served by setting  $l_{\max} = \infty$ . From Eq. A.3, we see that for  $l_{\min} = 0.7$  Mpc and  $\xi = -3.5$ , the tail of the intrinsic proper length distribution ( $l > 5$  Mpc) contains less than 1% of probability. Thus, for most framework applications, our choice  $l_{\max} = \infty$  is unproblematic — except when higher powers of  $L$  are involved. For example, the volume-filling fraction calculations of Appendix A.10 necessitate considering  $L^3$ ; for realistic  $\xi$ , results exist only for finite  $l_{\max}$ .

Upon relabelling  $\xi \rightarrow -\alpha - 1$ ,  $l_{\min} \rightarrow k$ ,  $L \rightarrow X$  and  $l \rightarrow x$ , one obtains the literature's most common form of the Pareto Type I PDF:

$$f_X(x) = \begin{cases} \frac{\alpha k^\alpha}{x^{\alpha+1}} & x \geq k, \\ 0 & x < k. \end{cases} \quad (\text{A.4})$$

### A.2. Projected proper length

#### A.2.1. Distribution for RGs

Let  $f_{L_p} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  be the PDF of the distribution of projected proper radio galaxy lengths. The PDF  $f_{L_p}$  follows from the associated CDF  $F_{L_p} : \mathbb{R} \rightarrow [0, 1]$  through differentiation;  $F_{L_p}$  and  $f_L$

relate through

$$\begin{aligned} F_{L_p}(l_p) &:= \mathbb{P}(L_p \leq l_p) = \int_{\mathbb{R}} \mathbb{P}(L \sin \Theta \leq l_p \mid L = l) f_L(l) \, dl \\ &= \int_{\mathbb{R}} \mathbb{P}\left(\sin \Theta \leq \frac{l_p}{l}\right) f_L(l) \, dl. \end{aligned} \quad (\text{A.5})$$

We note that  $F_{L_p}(l_p)$  vanishes for  $l_p \leq 0$ :  $f_L(l)$  vanishes when  $l$  is negative, whilst  $\mathbb{P}(\sin \Theta \leq \frac{l_p}{l})$  vanishes when  $l$  is positive. Clearly, the interesting case is  $l_p > 0$ . We can differentiate between two cases:  $l_p \leq l_{\min}$ , and  $l_p > l_{\min}$ . In the first case, considering that  $f_L$  has support from  $l_{\min}$  onwards only, we have  $\frac{l_p}{l} \leq 1$ , and

$$\begin{aligned} F_{L_p}(l_p) &= \int_{l_{\min}}^{\infty} \left( \mathbb{P}\left(\Theta \leq \arcsin \frac{l_p}{l}\right) \right. \\ &\quad \left. + \mathbb{P}\left(\Theta \geq \pi - \arcsin \frac{l_p}{l}\right) \right) \frac{-(\xi+1)}{l_{\min}} \left(\frac{l}{l_{\min}}\right)^\xi \, dl \\ &= -\frac{\xi+1}{l_{\min}} \int_{l_{\min}}^{\infty} 2\mathbb{P}\left(\Theta \leq \arcsin \frac{l_p}{l}\right) \left(\frac{l}{l_{\min}}\right)^\xi \, dl \\ &= -\frac{\xi+1}{l_{\min}} \int_{l_{\min}}^{\infty} \left(1 - \sqrt{1 - \left(\frac{l_p}{l}\right)^2}\right) \left(\frac{l}{l_{\min}}\right)^\xi \, dl \\ &= 1 + (\xi+1) \int_{l_{\min}}^{\infty} \sqrt{\left(\frac{l}{l_{\min}}\right)^2 - \left(\frac{l_p}{l_{\min}}\right)^2} \left(\frac{l}{l_{\min}}\right)^{\xi-1} \frac{dl}{l_{\min}} \\ &= 1 + (\xi+1) \int_1^{\infty} \sqrt{\eta^2 - \left(\frac{l_p}{l_{\min}}\right)^2} \eta^{\xi-1} \, d\eta. \end{aligned} \quad (\text{A.6})$$

In the second case, we split up the integral in two:

$$\begin{aligned} F_{L_p}(l_p) &= \int_{l_{\min}}^{l_p} \mathbb{P}\left(\sin \Theta \leq \frac{l_p}{l}\right) f_L(l) \, dl \\ &\quad + \int_{l_p}^{\infty} \mathbb{P}\left(\sin \Theta \leq \frac{l_p}{l}\right) f_L(l) \, dl \\ &= \int_{l_{\min}}^{l_p} f_L(l) \, dl + \int_{l_p}^{\infty} \left(1 - \sqrt{1 - \left(\frac{l_p}{l}\right)^2}\right) f_L(l) \, dl \\ &= 1 - \left(\frac{l_p}{l_{\min}}\right)^{\xi+1} + \left(\frac{l_p}{l_{\min}}\right)^{\xi+1} \\ &\quad + (\xi+1) \int_{l_p}^{\infty} \sqrt{\left(\frac{l}{l_{\min}}\right)^2 - \left(\frac{l_p}{l_{\min}}\right)^2} \left(\frac{l}{l_{\min}}\right)^{\xi-1} \frac{dl}{l_{\min}} \\ &= 1 + (\xi+1) \int_{\frac{l_p}{l_{\min}}}^{\infty} \sqrt{\eta^2 - \left(\frac{l_p}{l_{\min}}\right)^2} \eta^{\xi-1} \, d\eta \\ &= 1 + (\xi+1) \frac{\sqrt{\pi}}{4} \left(\frac{l_p}{l_{\min}}\right)^{\xi+1} \frac{\Gamma\left(-\frac{\xi}{2} - \frac{1}{2}\right)}{\Gamma\left(-\frac{\xi}{2} + 1\right)}. \end{aligned} \quad (\text{A.7})$$

In summary,

$$F_{L_p}(l_p) = \begin{cases} 0 & \text{if } l_p \leq 0 \\ 1 + (\xi+1) \int_1^{\infty} \sqrt{\eta^2 - \left(\frac{l_p}{l_{\min}}\right)^2} \eta^{\xi-1} \, d\eta & \text{if } 0 < l_p < l_{\min} \\ 1 + (\xi+1) \frac{\sqrt{\pi}}{4} \left(\frac{l_p}{l_{\min}}\right)^{\xi+1} \frac{\Gamma\left(-\frac{\xi}{2} - \frac{1}{2}\right)}{\Gamma\left(-\frac{\xi}{2} + 1\right)} & \text{if } l_p \geq l_{\min}. \end{cases} \quad (\text{A.8})$$

Through differentiation,

$$f_{L_p}(l_p) = \begin{cases} 0 & \text{if } l_p \leq 0 \\ -\frac{\xi+1}{l_{\min}} \frac{l_p}{l_{\min}} I\left(\xi-1, \frac{l_p}{l_{\min}}\right) & \text{if } 0 < l_p < l_{\min} \\ \frac{(\xi+1)^2}{l_{\min}} \frac{\sqrt{\pi}}{4} \left(\frac{l_p}{l_{\min}}\right)^\xi \frac{\Gamma\left(-\frac{\xi}{2}-\frac{1}{2}\right)}{\Gamma\left(-\frac{\xi}{2}+1\right)} & \text{if } l_p \geq l_{\min}. \end{cases} \quad (\text{A.9})$$

where

$$I(a, b) := \int_1^\infty \frac{\eta^a d\eta}{\sqrt{\eta^2 - b^2}} \text{ for } a < 0, |b| < 1. \quad (\text{A.10})$$

Significantly, for  $l_p > l_{\min}$ , the projected length distribution follows a power law in  $l_p$  with the same exponent as the power law for the intrinsic length distribution:  $f_{L_p} \propto l_p^\xi$  (just as  $f_L \propto l^\xi$ ).<sup>23</sup> We compare  $f_L$  and  $f_{L_p}$  in Fig. 1.

### A.2.2. Distribution for giants

To derive the projected length distribution for giants, we consider the distribution of the conditioned RV  $L_p | L_p > l_{p,\text{GRG}}$ :

$$\begin{aligned} F_{L_p | L_p > l_{p,\text{GRG}}}(l_p) &:= \mathbb{P}(L_p \leq l_p | L_p > l_{p,\text{GRG}}) \\ &= 1 - \mathbb{P}(L_p > l_p | L_p > l_{p,\text{GRG}}) \\ &= 1 - \frac{\mathbb{P}(L_p > l_p, L_p > l_{p,\text{GRG}})}{\mathbb{P}(L_p > l_{p,\text{GRG}})}. \end{aligned} \quad (\text{A.11})$$

For  $l_p > l_{p,\text{GRG}}$ , this reduces to

$$F_{L_p | L_p > l_{p,\text{GRG}}}(l_p) = 1 - \frac{1 - F_{L_p}(l_p)}{1 - F_{L_p}(l_{p,\text{GRG}})}. \quad (\text{A.12})$$

Furthermore assuming  $l_{p,\text{GRG}} > l_{\min}$ , we twice use the bottom expression of Eq. A.8 to obtain the final CDF expression. As before, the corresponding PDF follows through differentiation. We find

$$F_{L_p | L_p > l_{p,\text{GRG}}}(l_p) = \begin{cases} 0 & \text{if } l_p \leq l_{p,\text{GRG}} \\ 1 - \left(\frac{l_p}{l_{p,\text{GRG}}}\right)^{\xi+1} & \text{if } l_p > l_{p,\text{GRG}}. \end{cases} \quad (\text{A.13})$$

$$f_{L_p | L_p > l_{p,\text{GRG}}}(l_p) = \begin{cases} 0 & \text{if } l_p \leq l_{p,\text{GRG}} \\ -\frac{\xi+1}{l_{p,\text{GRG}}} \left(\frac{l_p}{l_{p,\text{GRG}}}\right)^\xi & \text{if } l_p > l_{p,\text{GRG}}. \end{cases} \quad (\text{A.14})$$

Thus, the associated survival function is

$$\mathbb{P}(L_p > l_p | L_p > l_{p,\text{GRG}}) = \left(\frac{l_p}{l_{p,\text{GRG}}}\right)^{\xi+1}. \quad (\text{A.15})$$

The mean projected proper length of giants follows from the PDF by direct computation:

$$\begin{aligned} \mathbb{E}[L_p | L_p > l_{p,\text{GRG}}] &:= \int_{-\infty}^\infty f_{L_p | L_p > l_{p,\text{GRG}}}(l_p) \cdot l_p dl_p \\ &= l_{p,\text{GRG}} \frac{\xi+1}{\xi+2}. \end{aligned} \quad (\text{A.16})$$

<sup>23</sup> For  $l_{\max} < \infty$ , this statement does not hold exactly.

### A.3. Deprojection factor

Consider a radio galaxy (RG) with a projected proper length  $l_p$ . Let the inclination angle  $\theta$  denote the angle between the RG's central axis and the line of sight. The RG's inclination angle, projected proper length and intrinsic proper length  $l$  relate via  $l_p = l \sin \theta$ . Switching to random variable (RV) notation by using capital letters, the intrinsic proper length (which is the most physically relevant quantity) follows from the projected proper length (which can be measured) and the inclination angle (which is typically unknown), according to

$$L = \frac{1}{\sin \Theta} L_p. \quad (\text{A.17})$$

#### A.3.1. Without lobes

Calling the deprojection factor  $D := (\sin \Theta)^{-1}$ , we now calculate the distribution of  $D$  for  $f_\Theta(\theta) = \frac{1}{2} \sin \theta$ ;  $\theta \in [0, \pi]$ . The result is a continuous univariate distribution without parameters supported on the semi-infinite interval  $(1, \infty)$ . Let  $F_D : \mathbb{R} \rightarrow [0, 1]$  be the cumulative density function (CDF) of  $D$ . Then, for  $d > 1$ ,

$$\begin{aligned} F_D(d) &:= \mathbb{P}(D \leq d) = \mathbb{P}\left(\sin \Theta \geq \frac{1}{d}\right) \\ &= \mathbb{P}\left(\arcsin \frac{1}{d} \leq \Theta \leq \pi - \arcsin \frac{1}{d}\right) \\ &= F_\Theta\left(\pi - \arcsin \frac{1}{d}\right) - F_\Theta\left(\arcsin \frac{1}{d}\right) \\ &= \cos \arcsin \frac{1}{d} = \sqrt{1 - \frac{1}{d^2}}. \end{aligned} \quad (\text{A.18})$$

Meanwhile,  $F_D(d) = 0$  for  $d \leq 1$ . The quantile function  $F_D^{-1} : [0, 1] \rightarrow [1, \infty)$  follows from solving  $F_D(d) = p$  for  $d$ :

$$F_D^{-1}(p) = \frac{1}{\sqrt{1 - p^2}}. \quad (\text{A.19})$$

Thus, the minimum factor is  $F_D^{-1}(p=0) = 1$ , the median  $F_D^{-1}(p = \frac{1}{2}) = \frac{2}{\sqrt{3}}$ ,  $F_D^{-1}(p = \frac{1}{2}\sqrt{2}) = \sqrt{2}$ ,  $F_D^{-1}(p = \frac{1}{2}\sqrt{3}) = 2$ , and factors can grow arbitrarily large:  $F_D^{-1}(p) \rightarrow \infty$  as  $p \rightarrow 1$ . We conclude that half of all RGs have an intrinsic proper length more than  $\frac{2}{\sqrt{3}}$  their projected proper length.

By differentiating  $F_D$  to  $d$  we obtain  $f_D : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ , the probability density function (PDF) of  $D$ :

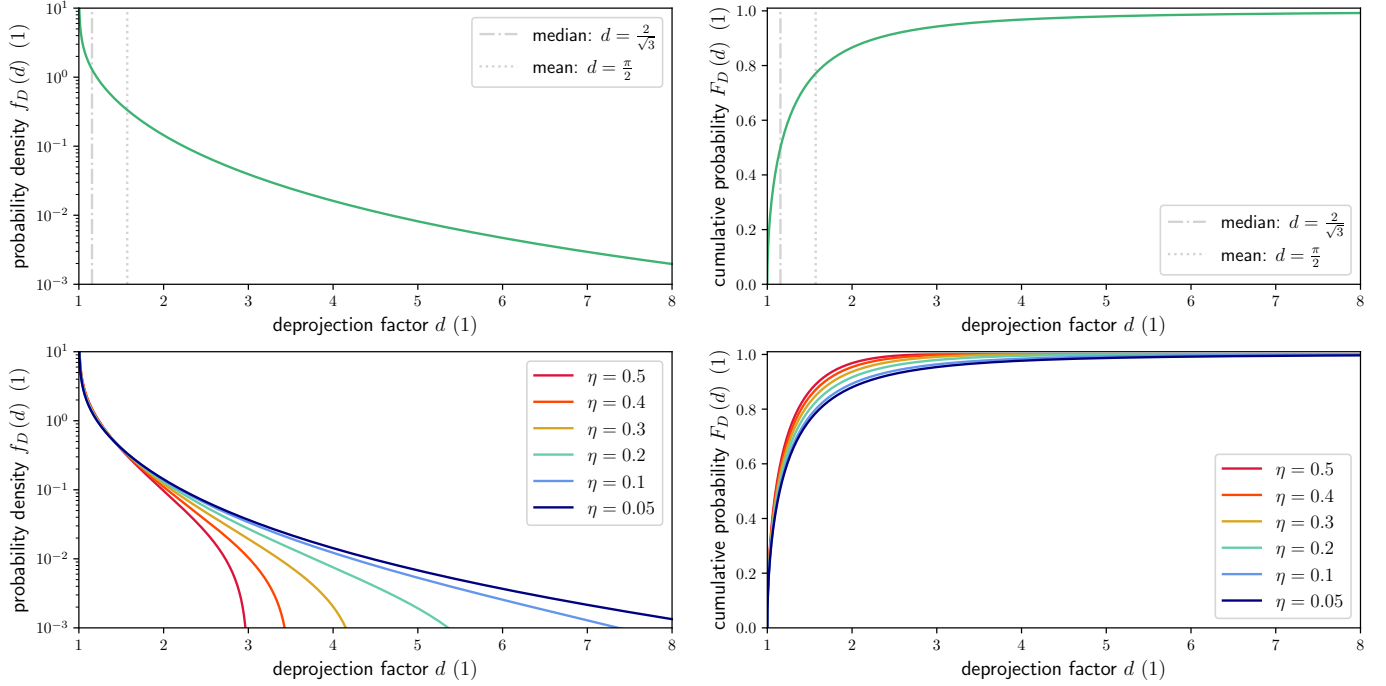
$$f_D(d) = \begin{cases} \frac{1}{d^2 \sqrt{d^2 - 1}} & \text{if } d > 1; \\ 0 & \text{if } d \leq 1. \end{cases} \quad (\text{A.20})$$

The mean of  $D$  is  $\mathbb{E}(D) = \frac{\pi}{2}$ ; the variance of  $D$  is undefined, as the corresponding integral diverges. Because  $f_D(d)$  has no maximum, the mode is undefined too. The PDF and CDF of  $D$  are shown in the upper left and right panels of Fig. A.1, respectively.

#### A.3.2. With lobes

The simplest model that includes a pair of lobes approximates them as spheres of radius  $R$ , whose centres are connected by a line segment of length  $L$ . Regardless of the viewing angle, the spheres retain their size, in contrast to the line segment connecting them. Calling  $\eta := \frac{2R}{L} \in \mathbb{R}_{\geq 0}$ , we have

$$D := \frac{L + 2R}{L \sin \Theta + 2R} = \frac{1 + \eta}{\sin \Theta + \eta}. \quad (\text{A.21})$$



**Fig. A.1.** PDFs (left column) and CDFs (right column) of the deprojection factor RV  $D$ . These functions quantify how much longer RGs are than their projected lengths suggest. *Top row:* model without lobes. Of the three canonical measures of central tendency, only the median and mean exist (and equal  $\frac{2}{\sqrt{3}}$  and  $\frac{\pi}{2}$ , respectively). *Bottom row:* model with spherical lobes. The distribution of  $D$  now depends on a parameter: the ratio  $\eta$  between the lobe diameter  $2R$  and the distance between the lobe centres  $L$ . The model without lobes is recovered in the limit  $\eta \rightarrow 0$ .

Again assuming  $f_{\Theta}(\theta) = \frac{1}{2} \sin \theta$ ;  $\theta \in [0, \pi]$ , we repeat the derivation and find

$$F_D(d) = \begin{cases} 0 & \text{if } d \leq 1; \\ \sqrt{1 - \left(\frac{1+\eta}{d} - \eta\right)^2} & \text{if } 1 < d < \frac{1}{\eta} + 1; \\ 1 & \text{if } d \geq \frac{1}{\eta} + 1. \end{cases} \quad (\text{A.22})$$

$$f_D(d) = \begin{cases} 0 & \text{if } d \leq 1; \\ \frac{(1+\eta)(1+\eta-\eta d)}{d^2 \sqrt{d^2 - (1+\eta-\eta d)^2}} & \text{if } 1 < d < \frac{1}{\eta} + 1; \\ 0 & \text{if } d \geq \frac{1}{\eta} + 1. \end{cases} \quad (\text{A.23})$$

The quantile function becomes

$$F_D^{-1}(p) = \frac{1 + \eta}{\sqrt{1 - p^2} + \eta}. \quad (\text{A.24})$$

The minimum factor remains  $F_D^{-1}(p = 0) = 1$ , but a maximum factor now exists:  $F_D^{-1}(p = 1) = \frac{1}{\eta} + 1$ . The median of  $D$  is

$$F_D^{-1}\left(p = \frac{1}{2}\right) = \frac{1 + \eta}{\frac{1}{2}\sqrt{3} + \eta}, \quad (\text{A.25})$$

which tends to  $\frac{2}{\sqrt{3}}$  for  $\eta \rightarrow 0$  (as before), and to 1 for  $\eta \rightarrow \infty$ : projection ceases to be an appreciable effect when the lobes are much larger than the line segment connecting their centres. If  $\eta \neq 0$ ,  $D$  has finite support, and thus the mean, variance and higher moments exist. The  $n$ -th non-central moment is

$$\mathbb{E}[D^n] = (1 + \eta)^2 I_{n-2}(\eta) - \eta(1 + \eta) I_{n-1}(\eta), \quad (\text{A.26})$$

where

$$I_n(\eta) := \int_1^{\frac{1}{\eta}+1} \frac{d^n dd}{\sqrt{d^2 - (1 + \eta - \eta d)^2}}. \quad (\text{A.27})$$

In particular,

$$I_{-1}(\eta) = \frac{\pi}{2(1 + \eta)}, \quad I_0(\eta) = \frac{\ln(\sqrt{1 - \eta^2} + 1) - \ln \eta}{\sqrt{1 - \eta^2}},$$

$$I_1(\eta) = \frac{1}{(1 - \eta)\eta} \left( 1 - \frac{2\eta^2 \operatorname{arcsinh} \sqrt{\frac{1-\eta}{2\eta}}}{\sqrt{1 - \eta^2}} \right). \quad (\text{A.28})$$

The expectation value of  $D$  is

$$\begin{aligned} \mathbb{E}[D] &= (1 + \eta)^2 I_{-1}(\eta) - \eta(1 + \eta) I_0(\eta) \\ &= (1 + \eta) \left( \frac{\pi}{2} - \frac{\eta}{\sqrt{1 - \eta^2}} \ln \left( \frac{\sqrt{1 - \eta^2}}{\eta} + \frac{1}{\eta} \right) \right). \end{aligned} \quad (\text{A.29})$$

The variance of  $D$  follows from combining

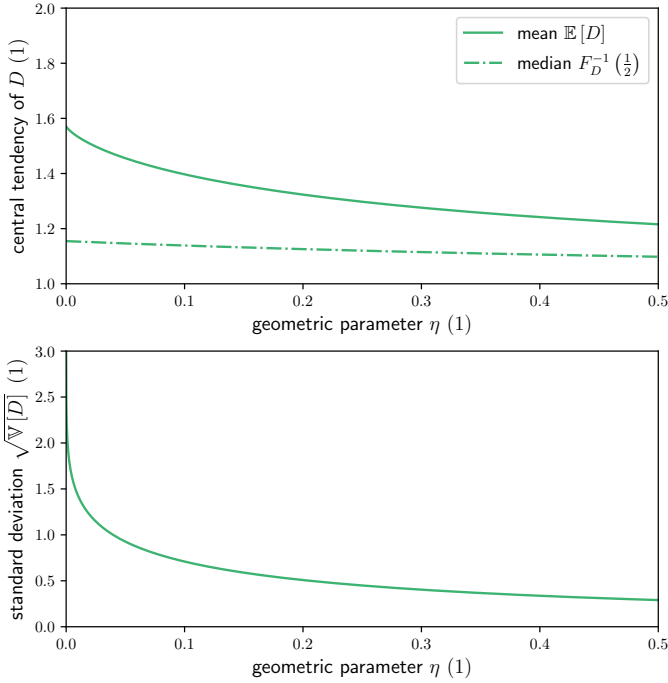
$$\mathbb{E}[D^2] = (1 + \eta)^2 I_0(\eta) - \eta(1 + \eta) I_1(\eta) \quad (\text{A.30})$$

and the identity  $\mathbb{V}[D] = \mathbb{E}[D^2] - \mathbb{E}^2[D]$ . The mode remains undefined. For typical values of  $\eta$ , we show the PDF and CDF of  $D$  in the bottom left and right panels of Fig. A.1, respectively. Figure A.2 shows the median, mean and standard deviation of  $D$  as a function of  $\eta$ .

#### A.4. Intrinsic proper length posterior and its moments

Our next objective is to find the posterior PDF  $f_{L|L_p=l_p}(l)$  through Bayes' theorem:

$$f_{L|L_p=l_p}(l) = \frac{f_{L_p|L=l}(l_p) f_L(l)}{f_{L_p}(l_p)}. \quad (\text{A.31})$$



**Fig. A.2.** Summary statistics for the deprojection factor RV  $D$  under the spherical lobe model. *Top:* the  $\eta$ -dependency of two measures of central tendency of  $D$ . For  $\eta = 0$ , the mean  $\mathbb{E}[D] = \frac{\pi}{2}$  and the median  $F_D^{-1}(\frac{1}{2}) = \frac{2}{\sqrt{3}}$ . *Bottom:* the  $\eta$ -dependency of the standard deviation of  $D$ . As  $\eta \rightarrow 0+$ ,  $\sqrt{\mathbb{V}[D]} \rightarrow \infty$ .

Our line of attack will be to first calculate the likelihood CDF  $F_{L_p|L=l}(l_p)$ , and then through differentiation the likelihood PDF  $f_{L_p|L=l}(l_p)$ . Of course, we always consider  $l > 0$ , and

$$F_{L_p|L=l}(l_p) := \mathbb{P}(L_p \leq l_p | L = l) = \mathbb{P}(l \sin \Theta \leq l_p) = \begin{cases} 0 & \text{if } l_p \leq 0 \\ 2 \mathbb{P}(\Theta \leq \arcsin \frac{l_p}{l}) & \text{if } 0 < l_p \leq l \\ 1 & \text{if } l_p > l, \end{cases} \quad (\text{A.32})$$

where we have invoked Eq. A.17 and the fact that  $f_\Theta(\theta)$  is symmetric in  $\theta = \frac{\pi}{2}$  for the case  $0 < l_p < l$ . Concretising this case further yields

$$F_{L_p|L=l}(l_p) = 2 F_\Theta\left(\arcsin \frac{l_p}{l}\right) = 1 - \cos \arcsin \frac{l_p}{l} = 1 - \sqrt{1 - \left(\frac{l_p}{l}\right)^2}. \quad (\text{A.33})$$

Through differentiation,

$$f_{L_p|L=l}(l_p) = \begin{cases} 0 & \text{if } l_p \leq 0 \\ \frac{l_p}{l^2} \frac{1}{\sqrt{1 - \left(\frac{l_p}{l}\right)^2}} & \text{if } 0 < l_p < l \\ 0 & \text{if } l_p \geq l. \end{cases} \quad (\text{A.34})$$

Having found the likelihood, the posterior PDF follows directly through Bayes' theorem and the PDFs computed hitherto.

In concreto, when  $0 < l_p \leq l_{\min}$ ,

$$f_{L|L_p=l_p}(l) = \begin{cases} 0 & \text{if } l \leq l_{\min} \\ \frac{1}{I(\xi-1, \frac{l_p}{l_{\min}})} \frac{1}{l_p} \frac{1}{\sqrt{\left(\frac{l}{l_p}\right)^2 - 1}} \left(\frac{l}{l_{\min}}\right)^{\xi-1} & \text{if } l > l_{\min}. \end{cases} \quad (\text{A.35})$$

whereas for  $l_p > l_{\min}$ ,

$$f_{L|L_p=l_p}(l) = \begin{cases} 0 & \text{if } l \leq l_p \\ -\frac{\xi}{2^{1+\xi}\pi} \frac{\Gamma^2(-\frac{\xi}{2})}{\Gamma(-\xi)} \frac{1}{l_p} \frac{1}{\sqrt{\left(\frac{l}{l_p}\right)^2 - 1}} \left(\frac{l}{l_p}\right)^{\xi-1} & \text{if } l > l_p. \end{cases} \quad (\text{A.36})$$

We note that in both cases, for  $l \gg l_p$ ,  $f_{L|L_p=l_p}(l) \propto \left(\frac{l}{l_p}\right)^{\xi-2}$ : the posterior probability density follows a power law in  $l$  with exponent  $\xi - 2$ .

For  $l_p \leq l_{\min}$ , the posterior mean is

$$\mathbb{E}[L | L_p = l_p] = l_{\min} \frac{I\left(\xi, \frac{l_p}{l_{\min}}\right)}{I\left(\xi - 1, \frac{l_p}{l_{\min}}\right)}. \quad (\text{A.37})$$

The posterior variance  $\mathbb{V}[L | L_p = l_p]$  follows from considering the second non-central moment:

$$\mathbb{E}[L^2 | L_p = l_p] = l_{\min}^2 \frac{I\left(\xi + 1, \frac{l_p}{l_{\min}}\right)}{I\left(\xi - 1, \frac{l_p}{l_{\min}}\right)}, \quad (\text{A.38})$$

so that

$$\mathbb{V}[L | L_p = l_p] = \mathbb{E}[L^2 | L_p = l_p] - \mathbb{E}^2[L | L_p = l_p] = l_{\min}^2 \left( \frac{I\left(\xi + 1, \frac{l_p}{l_{\min}}\right)}{I\left(\xi - 1, \frac{l_p}{l_{\min}}\right)} - \frac{I^2\left(\xi, \frac{l_p}{l_{\min}}\right)}{I^2\left(\xi - 1, \frac{l_p}{l_{\min}}\right)} \right). \quad (\text{A.39})$$

For  $l_p > l_{\min}$ ,

$$\mathbb{E}[L | L_p = l_p] = l_p \cdot \frac{-\xi}{2^{\xi+1}\pi} \frac{\Gamma^2(-\frac{\xi}{2})}{\Gamma(-\xi)} \int_1^\infty \frac{\eta^\xi d\eta}{\sqrt{\eta^2 - 1}} = l_p \cdot \frac{-\xi}{2^{2\xi+3}\pi} \frac{\Gamma^4(-\frac{\xi}{2})}{\Gamma^2(-\xi)}. \quad (\text{A.40})$$

Proceeding analogously,

$$\mathbb{E}[L^2 | L_p = l_p] = l_p^2 \frac{\xi}{\xi + 1}, \quad (\text{A.41})$$

so that

$$\mathbb{V}[L | L_p = l_p] = l_p^2 \left( \frac{\xi}{\xi + 1} - \frac{\xi^2}{2^{4\xi+6}\pi^2} \frac{\Gamma^8(-\frac{\xi}{2})}{\Gamma^4(-\xi)} \right). \quad (\text{A.42})$$

In this case, both the mean and standard deviation of  $L | L_p = l_p$  are proportional to  $l_p$ . In the table below, we list the mean and standard deviation in multiples of  $l_p$  for several values of  $\xi$ . Since we assume  $\xi < -1$ , the mean and variance are guaranteed to exist. The existence of higher-order moments is  $\xi$ -dependent; in concreto, the highest defined order is  $\lceil -\xi \rceil$ .

We prove that  $L_p$  and  $D$  are not independent by contradiction. If we assume that  $L_p$  and  $D$  are independent, then  $\mathbb{E}[L_p D] = \mathbb{E}[L_p] \mathbb{E}[D]$ , or  $\mathbb{E}[L] = \mathbb{E}[L] \mathbb{E}[\sin \Theta] \mathbb{E}[D]$  by the independence of  $L$  and  $\sin \Theta$ . In other words, if  $L_p$  and  $D$  are independent, then  $1 = \mathbb{E}[\sin \Theta] \mathbb{E}[D]$ . However,  $\mathbb{E}[\sin \Theta] = \frac{\pi}{4}$  and  $\mathbb{E}[D] = \frac{\pi}{2}$ ; because  $1 \neq \frac{\pi^2}{8}$ , the initial assertion must be wrong.

### A.5. GRG inclination angle

We derive the inclination angle distribution for giants. The probability that a GRG has inclination angle  $\theta$  is

$$\begin{aligned} f_{\Theta | L_p > l_{p,\text{GRG}}}(\theta) d\theta &= \frac{\mathbb{P}(\Theta = \theta, L \sin \Theta > l_{p,\text{GRG}})}{\mathbb{P}(L_p > l_{p,\text{GRG}})} \\ &= \frac{f_{\Theta}(\theta) \cdot \mathbb{P}(L > \frac{l_{p,\text{GRG}}}{\sin \theta})}{\mathbb{P}(L_p > l_{p,\text{GRG}})}, \end{aligned} \quad (\text{A.43})$$

where we make use of the fact that the numerator's joint probability factorises because  $L$  and  $\Theta$  are independent RVs. We thus find

$$f_{\Theta | L_p > l_{p,\text{GRG}}}(\theta) = \frac{(1 - F_L(\frac{l_{p,\text{GRG}}}{\sin \theta})) f_{\Theta}(\theta)}{1 - F_{L_p}(l_{p,\text{GRG}})}. \quad (\text{A.44})$$

Under the Paretian assumption for  $L$ , we have

$$f_{\Theta | L_p > l_{p,\text{GRG}}}(\theta) = \frac{2}{-(\xi + 1) \sqrt{\pi}} \frac{\Gamma(-\frac{\xi}{2} + 1)}{\Gamma(-\frac{\xi}{2} - \frac{1}{2})} \sin^{-\xi} \theta. \quad (\text{A.45})$$

### A.6. GRG angular length

The GRG angular length RV  $\Phi | L_p > l_{p,\text{GRG}}$  relates to the GRG projected proper length RV  $L_p | L_p > l_{p,\text{GRG}}$  and the comoving distance RV  $R$  as

$$\Phi | L_p > l_{p,\text{GRG}} = L_p | L_p > l_{p,\text{GRG}} \cdot \frac{1 + z(R)}{R}. \quad (\text{A.46})$$

The model predicts the distribution of GRG angular lengths in the Local Universe up to comoving distance  $r_{\text{max}}$ . If the GRG number density is constant in the Local Universe,

$$F_R(r) = \begin{cases} 0 & \text{if } r \leq 0 \\ \left(\frac{r}{r_{\text{max}}}\right)^3 & \text{if } 0 < r < r_{\text{max}} \\ 1 & \text{if } r \geq r_{\text{max}} \end{cases} \quad (\text{A.47})$$

$$f_R(r) = \begin{cases} 0 & \text{if } r \leq 0 \\ \frac{3}{r_{\text{max}}} \left(\frac{r}{r_{\text{max}}}\right)^2 & \text{if } 0 < r < r_{\text{max}} \\ 0 & \text{if } r \geq r_{\text{max}}. \end{cases} \quad (\text{A.48})$$

Because GRG life cycles are shorter than the age of the Universe,  $L_p | L_p > l_{p,\text{GRG}}$  and  $R$  are independent RVs. In a non-expanding universe,  $z(R) = 0$ , and the distribution of  $\Phi | L_p > l_{p,\text{GRG}}$  can be calculated analytically. From a well-known ratio distribution identity,

$$\begin{aligned} f_{\Phi | L_p > l_{p,\text{GRG}}}(\phi) &= \int_{-\infty}^{\infty} |r| \cdot f_{L_p | L_p > l_{p,\text{GRG}}}(\phi r) \cdot f_R(r) dr \\ &= \int_0^{r_{\text{max}}} f_{L_p | L_p > l_{p,\text{GRG}}}(\phi r) \cdot 3 \left(\frac{r}{r_{\text{max}}}\right)^3 dr. \end{aligned} \quad (\text{A.49})$$

The integrand is non-zero only when  $\phi r \geq l_{p,\text{GRG}}$ , suggesting a lower integration limit of  $\frac{l_{p,\text{GRG}}}{\phi}$ . The integral vanishes altogether when  $\frac{l_{p,\text{GRG}}}{\phi} \geq r_{\text{max}}$ . Calling  $\phi_{\text{GRG}} := \frac{l_{p,\text{GRG}}}{r_{\text{max}}}$ , we have  $f_{\Phi | L_p > l_{p,\text{GRG}}}(\phi) = 0$  for  $\phi \leq \phi_{\text{GRG}}$ . For  $\phi > \phi_{\text{GRG}}$ ,

$$\begin{aligned} f_{\Phi | L_p > l_{p,\text{GRG}}}(\phi) &= \int_{\frac{l_{p,\text{GRG}}}{\phi}}^{r_{\text{max}}} \frac{\xi + 1}{l_{p,\text{GRG}}} \left(\frac{\phi r}{l_{p,\text{GRG}}}\right)^{\xi} \cdot 3 \left(\frac{r}{r_{\text{max}}}\right)^3 dr \\ &= -3 \frac{\xi + 1}{\xi + 4} \cdot \frac{1}{\phi_{\text{GRG}}} \left( \left(\frac{\phi}{\phi_{\text{GRG}}}\right)^{\xi} - \left(\frac{\phi}{\phi_{\text{GRG}}}\right)^{-4} \right), \end{aligned} \quad (\text{A.50})$$

where we assume  $\xi \neq -4$ . This PDF depends on  $\xi$ ,  $l_{p,\text{GRG}}$ , and  $r_{\text{max}}$  only. The CDF follows from direct integration:

$$F_{\Phi | L_p > l_{p,\text{GRG}}}(\phi) = \begin{cases} 0 & \text{if } \phi \leq \phi_{\text{GRG}} \\ 1 + \frac{1}{\xi + 4} \left( -3 \left(\frac{\phi}{\phi_{\text{GRG}}}\right)^{\xi + 1} - (\xi + 1) \left(\frac{\phi}{\phi_{\text{GRG}}}\right)^{-3} \right) & \text{if } \phi > \phi_{\text{GRG}}, \end{cases} \quad (\text{A.51})$$

just as the mean:

$$\mathbb{E}[\Phi | L_p > l_{p,\text{GRG}}] = 3 \frac{\xi + 1}{\xi + 4} \left( \frac{1}{\xi + 2} + \frac{1}{2} \right) \cdot \phi_{\text{GRG}}, \quad (\text{A.52})$$

which exists only for  $\xi < -2$ . By solving

$$\frac{df_{\Phi} | L_p > l_{p,\text{GRG}}}{d\phi}(\phi_{\text{mode}}) = 0 \quad (\text{A.53})$$

for the GRG angular length mode  $\phi_{\text{mode}}$ , we find

$$\phi_{\text{mode}} = \left( \frac{-4}{\xi} \right)^{\frac{1}{\xi + 4}} \cdot \phi_{\text{GRG}}. \quad (\text{A.54})$$

There is no explicit expression for the associated median.

For a general non-Euclidean universe, no simple analytic form for the GRG angular length PDF appears to exist. We find an approximation valid at low redshifts by considering the Maclaurin polynomial of degree 1 for  $z(r)$ , the relation between comoving distance and cosmological redshift. One finds  $z(r) \approx \frac{H_0}{c} r$ . As a result, for the Local Universe, Eq. A.46 becomes

$$\begin{aligned} \Phi | L_p > l_{p,\text{GRG}} &\approx L_p | L_p > l_{p,\text{GRG}} \cdot \frac{1 + \frac{H_0}{c} R}{R} \\ &= L_p | L_p > l_{p,\text{GRG}} \cdot \left( \frac{1}{R} + \frac{1}{d_H} \right), \end{aligned} \quad (\text{A.55})$$

where we use that the Hubble distance  $d_H := \frac{c}{H_0}$ . From a well-known inverse distribution identity,

$$f_{\frac{1}{k}}(k) = \begin{cases} 0 & \text{if } k \leq \frac{1}{r_{\text{max}}} \\ \frac{3}{r_{\text{max}}^3 k^4} & \text{if } k > \frac{1}{r_{\text{max}}}, \end{cases} \quad (\text{A.56})$$

so that

$$f_{\frac{1}{k} + \frac{1}{d_H}}(k) = \begin{cases} 0 & \text{if } k \leq \frac{1}{r_{\text{max}}} + \frac{1}{d_H} \\ \frac{3}{r_{\text{max}}^3 \left(k - \frac{1}{d_H}\right)^4} & \text{if } k > \frac{1}{r_{\text{max}}} + \frac{1}{d_H}. \end{cases} \quad (\text{A.57})$$

Combining a well-known product distribution identity with the fact that  $L_p | L_p > l_{p,\text{GRG}}$  and  $\frac{1}{R} + \frac{1}{d_H}$  are independent RVs, we find

$$f_{\Phi | L_p > l_{p,\text{GRG}}}(\phi) = \frac{-3(\xi + 1) \phi^{\xi}}{r_{\text{max}}^3 l_{p,\text{GRG}}^{\xi + 1}} \int_{\frac{1}{r_{\text{max}}} + \frac{1}{d_H}}^{\frac{\phi}{l_{p,\text{GRG}}}} \frac{dk}{k^{\xi + 1} \left(k - \frac{1}{d_H}\right)^4} \quad (\text{A.58})$$

when  $\phi > \phi_{\text{GRG}} + \frac{l_{p,\text{GRG}}}{d_H}$ . When  $\phi \leq \phi_{\text{GRG}} + \frac{l_{p,\text{GRG}}}{d_H}$ ,  $f_{\Phi | L_p > l_{p,\text{GRG}}}(\phi) = 0$ . It is easy to verify that in the Euclidean limit,  $\frac{1}{d_H} \rightarrow 0$ , this expression reduces to that of Eq. A.50.



### A.7. Maximum likelihood estimation of the tail index

How can we estimate  $\xi$  from a sample of  $N$  giants? Let  $\xi_{\text{MLE}}$  be the maximum likelihood estimate (MLE) of  $\xi$ . This RV is a function of  $N$  IID RVs  $\{L_{p,1}, \dots, L_{p,N}\} \sim L_p \mid L_p > l_{p,\text{GRG}}$ . Define the following likelihood and log-likelihood functions:

$$\mathcal{L}(\xi) := \prod_{i=1}^N f_{L_p \mid L_p > l_{p,\text{GRG}}}(L_{p,i}) \quad (\text{A.59})$$

$$\begin{aligned} \tilde{\mathcal{L}}(\xi) &:= \ln(\mathcal{L}(\xi)) l_{p,\text{GRG}}^N \\ &= N \ln(-(\xi + 1)) + \xi \sum_{i=1}^N \ln \frac{L_{p,i}}{l_{p,\text{GRG}}}. \end{aligned} \quad (\text{A.60})$$

We note the necessity to include a factor  $l_{p,\text{GRG}}^N$  in the definition of the log-likelihood to avoid a dimensionality error. The second derivative of  $\tilde{\mathcal{L}}$  to  $\xi$  is

$$\frac{d^2 \tilde{\mathcal{L}}}{d\xi^2} = -\frac{N}{(\xi + 1)^2} < 0. \quad (\text{A.61})$$

Thus, if there exists a solution to the equation  $\frac{d\tilde{\mathcal{L}}}{d\xi} = 0$ , it must correspond to a global *maximum* of the likelihood and log-likelihood functions:

$$\frac{d\tilde{\mathcal{L}}}{d\xi}(\xi = \xi_{\text{MLE}}) = 0, \text{ or } \xi_{\text{MLE}} = -\frac{N}{\sum_{i=1}^N \ln \frac{L_{p,i}}{l_{p,\text{GRG}}}} - 1. \quad (\text{A.62})$$

### A.8. Observed projected proper length

#### A.8.1. General considerations

The model can be extended to incorporate observational selection effects. The relevant effects to consider vary from RG search campaign to RG search campaign, although some formulae apply in all cases. We derive these here.

To keep our extensions simple, we assume that the survey sensitivity is sufficient to detect RGs up to some redshift  $z_{\text{max}}$  only. In addition we assume that the projected proper length distribution does not evolve between  $z = z_{\text{max}}$  and  $z = 0$ . We let  $p_{\text{obs}}(l_p, z)$  denote the probability that an RG of projected proper length  $l_p$  at cosmological redshift  $z$  is detected during the campaign. Also,  $r$  is the radial comoving distance and  $n$  is the total RG number density, counting the intrinsic number of RGs (irrespective of length) per unit of comoving volume. The observed number of RGs with projected proper length between  $l_p$  and  $l_p + dl_p$  throughout a survey covering a solid angle  $\Omega$  is  $dN_{L_p, \text{obs}}(l_p, \Omega)$ , where

$$\begin{aligned} dN_{L_p, \text{obs}}(l_p, \Omega) \\ = \frac{\Omega}{4\pi} \int_0^{z_{\text{max}}} n \cdot f_{L_p}(l_p) dl_p \cdot p_{\text{obs}}(l_p, z) 4\pi r^2(z) \frac{dr}{dz} dz. \end{aligned} \quad (\text{A.63})$$

The total number of RGs with projected proper length between  $l_p$  and  $l_p + dl_p$  throughout a survey with solid angle  $\Omega$  is  $dN_{L_p}(l_p, \Omega)$ , where

$$dN_{L_p}(l_p, \Omega) = \frac{\Omega}{4\pi} \int_0^{z_{\text{max}}} n \cdot f_{L_p}(l_p) dl_p \cdot 4\pi r^2(z) \frac{dr}{dz} dz. \quad (\text{A.64})$$

We define the completeness  $C(l_p, z_{\text{max}})$  to be

$$\begin{aligned} C(l_p, z_{\text{max}}) &:= \frac{dN_{L_p, \text{obs}}(l_p, \Omega)}{dN_{L_p}(l_p, \Omega)} \\ &= \frac{\int_0^{z_{\text{max}}} p_{\text{obs}}(l_p, z) r^2(z) E^{-1}(z) dz}{\int_0^{z_{\text{max}}} r^2(z) E^{-1}(z) dz}. \end{aligned} \quad (\text{A.65})$$

The completeness only depends on the function  $p_{\text{obs}}(l_p, z)$  and our choice of  $z_{\text{max}}$ .

Let the RV  $O$  denote whether an RG picked at random within  $z < z_{\text{max}}$  is detected during the search campaign. We have  $O \sim \text{Bernoulli}(C(l_p, z_{\text{max}}))$ ; the parameter that determines the distribution of  $O$  is itself an RV. It immediately follows that  $\mathbb{P}(O = 1 \mid L_p = l_p) = C(l_p, z_{\text{max}})$ .

If  $L_p$  would be discrete,

$$\begin{aligned} \mathbb{P}(L_p = l_p \mid O = 1) &= \frac{\mathbb{P}(O = 1, L_p = l_p)}{\mathbb{P}(O = 1)} \\ &= \frac{\mathbb{P}(O = 1 \mid L_p = l_p) \mathbb{P}(L_p = l_p)}{\sum_{l_p} \mathbb{P}(O = 1 \mid L_p = l_p) \mathbb{P}(L_p = l_p)}. \end{aligned} \quad (\text{A.66})$$

Let  $L_{p, \text{obs}}$  be the observed projected proper length RV; that is  $L_{p, \text{obs}} := L_p \mid O = 1$ . Its PDF is given by the continuous analogon of the preceding equation:

$$f_{L_{p, \text{obs}}}(l_p) = \frac{C(l_p, z_{\text{max}}) f_{L_p}(l_p)}{\int_0^{\infty} C(l_p, z_{\text{max}}) f_{L_p}(l_p) dl_p}. \quad (\text{A.67})$$

Obviously,  $F_{L_{p, \text{obs}}}(l_p) = \int_0^{l_p} f_{L_{p, \text{obs}}}(l'_p) dl'_p$ . The CDF  $F_{L_{p, \text{obs}} \mid L_{p, \text{obs}} > l_{p, \text{GRG}}}(l_p)$  follows from an analogon of Eq. A.12.

We note that multiplying  $p_{\text{obs}}(l_p, z)$  with an  $l_p$ - and  $z$ -independent factor changes  $C(l_p, z_{\text{max}})$  by the same factor, but leaves  $f_{L_{p, \text{obs}}}$  unaltered.

#### A.8.2. Fuzzy angular length threshold

Here we illustrate a simple extension. When performing visual searches for GRG candidates, a natural criterion is to only inspect sources with an angular length that is larger than some threshold. Researchers determine the threshold based on the amount of time available to them: lower thresholds will lead to more complete samples, but will take more time to collect. For humans, it is hard to estimate a source's angular length precisely by eye; as a result, some of the GRG candidates included in the project's GRG candidate catalogue will be sources with an angular length below the threshold, whilst others will be sources with an angular length above the threshold. We idealise this situation by asserting that sources with angular length  $\phi_{\text{min}}$  or below are included in the catalogue with probability 0 (i.e. never), and that sources with an angular length  $\phi_{\text{max}}$  or above are included in the catalogue with probability 1 (i.e. always). We assume a linear increase in probability as a function of  $\phi$  for intermediate angular lengths: a source with angular length  $\phi_{\text{min}} < \phi < \phi_{\text{max}}$  is included in the catalogue with probability

$$p_{\text{obs}}(l_p, z) = \min \left\{ \max \left\{ \frac{\phi(l_p, z) - \phi_{\text{min}}}{\phi_{\text{max}} - \phi_{\text{min}}}, 0 \right\}, 1 \right\}. \quad (\text{A.68})$$

In a flat Friedmann–Lemaître–Robertson–Walker (FLRW) universe, the angular length of an RG with projected proper length  $l_p$  at cosmological redshift  $z$  is

$$\phi(l_p, z) = \frac{l_p(1+z)}{r(z)}. \quad (\text{A.69})$$

### A.8.3. Surface brightness limitations

**Fanaroff–Riley class II** If the surface brightness  $B_v$  of the lobes is proportional to some negative power  $\zeta$  of the GRG's proper length  $L$ , that is

$$B_v = b_{v,\text{ref}} \left( \frac{L}{l_{\text{ref}}} \right)^\zeta, \quad (\text{A.70})$$

then  $B_v$  is Pareto distributed — just like  $L$ :

$$F_{B_v}(b) = \begin{cases} 0 & \text{if } b \leq 0 \\ \left( \frac{b}{b_{v,\text{ref}}} \right)^{\frac{\zeta+1}{\zeta}} & \text{if } 0 < b < b_{v,\text{ref}} \\ 1 & \text{if } b \geq b_{v,\text{ref}}, \end{cases} \quad (\text{A.71})$$

$$f_{B_v}(b) = \begin{cases} 0 & \text{if } b \leq 0 \\ \frac{\zeta+1}{\zeta} \frac{1}{b_{v,\text{ref}}} \left( \frac{b}{b_{v,\text{ref}}} \right)^{\frac{\zeta+1}{\zeta}-1} & \text{if } 0 < b < b_{v,\text{ref}} \\ 0 & \text{if } b \geq b_{v,\text{ref}}. \end{cases} \quad (\text{A.72})$$

If RGs are self-similar in shape, then their lobe volumes are proportional to  $L^3$ . RGs appear to retain constant lobe luminosity density over most of their lifetime, so that their lobe monochromatic emission coefficients (Rybicki & Lightman 1986) are proportional to  $L^{-3}$ . As line-of-sight lengths through the lobes are proportional to  $L$ , surface brightness is proportional to  $L^{-2}$ . These arguments thus suggest  $\zeta = -2$ .

It is a poor approximation to assume that all giants of proper length  $l_{\text{ref}}$  have the same surface brightness  $b_{v,\text{ref}}$ : observations suggest a variability of several orders of magnitude. A better description is

$$B_v = b_{v,\text{ref}} \left( \frac{L}{l_{\text{ref}}} \right)^\zeta S, \quad (\text{A.73})$$

where  $S$  is a lognormal RV whose median is 1. The PDF of  $S$  is then determined by parameter  $\sigma_{\text{ref}}$ :

$$f_S(s) = \frac{1}{\sqrt{2\pi}\sigma_{\text{ref}}s} \exp\left(-\frac{\ln^2 s}{2\sigma_{\text{ref}}^2}\right). \quad (\text{A.74})$$

The surface brightness CDF and PDF now become

$$F_{B_v}(b) = \begin{cases} 0 & \text{if } b \leq 0 \\ F_S\left(\frac{b}{b_{v,\text{ref}}}\right) + \left(\frac{b}{b_{v,\text{ref}}}\right)^{\frac{\zeta+1}{\zeta}} \int_{\frac{b}{b_{v,\text{ref}}}}^{\infty} s^{-\frac{\zeta+1}{\zeta}} f_S(s) ds & \text{if } b > 0, \end{cases} \quad (\text{A.75})$$

$$f_{B_v}(b) = \begin{cases} 0 & \text{if } b \leq 0 \\ \frac{\zeta+1}{\zeta} \frac{1}{b_{v,\text{ref}}} \left( \frac{b}{b_{v,\text{ref}}} \right)^{\frac{\zeta+1}{\zeta}-1} \int_{\frac{b}{b_{v,\text{ref}}}}^{\infty} s^{-\frac{\zeta+1}{\zeta}} f_S(s) ds & \text{if } b > 0. \end{cases}$$

We note that  $B_v$  is not exactly Pareto distributed anymore.

In a relativistic rather than Euclidean universe, surface brightness is not constant with distance. To describe RGs beyond  $z = 0$ , we introduce a final model refinement:

$$B_v = \frac{b_{v,\text{ref}} \cdot S}{(1+Z)^{3-\alpha}} \left( \frac{L}{l_{\text{ref}}} \right)^\zeta, \quad (\text{A.76})$$

where the RV  $Z$  denotes cosmological redshift and  $\alpha$  is the spectral index of the lobes. We interpret  $b_{v,\text{ref}} \cdot S$  as the (lognormally distributed) lobe surface brightness for RGs of intrinsic proper length  $l_{\text{ref}}$  at  $z = 0$ . In this case, the CDF and PDF of  $B_v$  are most easily determined through sampling. To sample  $Z$ , we can first sample the comoving distance RV  $R$  instead, and subsequently use  $Z = z_{\text{M}}(R)$ . We stress that this conversion depends on cosmological parameters  $\mathfrak{M}$ .

To forward model a survey's surface brightness selection effect, we must compute

$$\begin{aligned} p_{\text{obs}}(l_p, z) &= \mathbb{P}(B_v > b_{v,\text{th}} \mid L_p = l_p, Z = z) \\ &= 1 - F_{B_v \mid L_p=l_p, Z=z}(b_{v,\text{th}}), \end{aligned} \quad (\text{A.77})$$

where  $b_{v,\text{th}} > 0$  is the surface brightness threshold. Typically,  $b_{v,\text{th}}$  is comparable to the survey's RMS noise. What is  $F_{B_v \mid L_p=l_p, Z=z}(b)$ ? In the simplest case, devoid of  $S$ - and  $Z$ -dependence,

$$\begin{aligned} F_{B_v \mid L_p=l_p, Z=z}(b) &= \mathbb{P}\left(b_{v,\text{ref}} \left( \frac{L_p}{\sin \Theta \cdot l_{\text{ref}}} \right)^\zeta \leq b \mid L_p = l_p\right) \\ &= \mathbb{P}\left(\sin^{-\zeta} \Theta \leq \frac{b}{b_{v,\text{ref}}} \left( \frac{l_p}{l_{\text{ref}}} \right)^{-\zeta}\right) \\ &= \mathbb{P}\left(\sin \Theta \leq \tilde{s}^{-\frac{1}{\zeta}}\right) \\ &= \begin{cases} 0 & \text{if } b \leq 0 \\ 2F_\Theta\left(\arcsin\left(\tilde{s}^{-\frac{1}{\zeta}}\right)\right) & \text{if } 0 < b < b_{v,\text{ref}} \left( \frac{l_p}{l_{\text{ref}}} \right)^\zeta \\ 1 & \text{if } b \geq b_{v,\text{ref}} \left( \frac{l_p}{l_{\text{ref}}} \right)^\zeta \end{cases} \\ &= 1 - \sqrt{1 - \min\left\{\tilde{s}^{-\frac{2}{\zeta}}(\max\{b, 0\}), 1\right\}}, \end{aligned} \quad (\text{A.78})$$

where

$$\tilde{s} = \tilde{s}(b) := \frac{b}{b_{v,\text{ref}}} \left( \frac{l_p}{l_{\text{ref}}} \right)^{-\zeta}. \quad (\text{A.79})$$

Therefore,

$$p_{\text{obs}}(l_p, z) = \sqrt{1 - \min\left\{\tilde{s}^{-\frac{2}{\zeta}}(b_{v,\text{th}}), 1\right\}}. \quad (\text{A.80})$$

In the most refined case, for  $b > 0$ ,

$$\begin{aligned} F_{B_v \mid L_p=l_p, Z=z}(b) &= \mathbb{P}\left(\frac{b_{v,\text{ref}} \cdot S}{(1+Z)^{3-\alpha}} \left( \frac{L_p}{\sin \Theta \cdot l_{\text{ref}}} \right)^\zeta \leq b \mid L_p = l_p, Z = z\right) \\ &= \mathbb{P}\left(\sin^{-\zeta} \Theta \cdot S \leq \frac{b}{b_{v,\text{ref}}} \left( \frac{l_p}{l_{\text{ref}}} \right)^{-\zeta} (1+z)^{3-\alpha}\right) \\ &= \int_0^\infty \mathbb{P}\left(\sin^{-\zeta} \Theta \leq \frac{b}{b_{v,\text{ref}}} \frac{1}{s} \left( \frac{l_p}{l_{\text{ref}}} \right)^{-\zeta} (1+z)^{3-\alpha}\right) f_S(s) ds \\ &= \int_0^\infty \mathbb{P}\left(\sin \Theta \leq \left(\frac{\tilde{s}}{s}\right)^{-\frac{1}{\zeta}}\right) f_S(s) ds \\ &= \int_0^{\tilde{s}} f_S(s) ds + \int_{\tilde{s}}^\infty 2F_\Theta\left(\arcsin\left(\left(\frac{\tilde{s}}{s}\right)^{-\frac{1}{\zeta}}\right)\right) f_S(s) ds \\ &= \int_0^{\tilde{s}} f_S(s) ds + \int_{\tilde{s}}^\infty f_S(s) ds \end{aligned}$$

$$\begin{aligned}
 & - \int_{\tilde{s}}^{\infty} \cos \arcsin \left( \left( \frac{\tilde{s}}{s} \right)^{-\frac{1}{\zeta}} \right) f_S(s) ds \\
 & = 1 - \int_{\tilde{s}}^{\infty} \sqrt{1 - \left( \frac{\tilde{s}}{s} \right)^{-\frac{2}{\zeta}}} f_S(s) ds,
 \end{aligned} \tag{A.81}$$

where

$$\tilde{s} = \tilde{s}(b) := \frac{b}{b_{v,\text{ref}}} \left( \frac{l_p}{l_{\text{ref}}} \right)^{-\zeta} (1+z)^{3-\alpha}. \tag{A.82}$$

The minimum value of  $S$  for which an RG of projected length  $l_p$  at redshift  $z$  is detectable, is  $s_{\text{min}}(l_p, z) := \tilde{s}(b_{v,\text{th}})$ . For brevity, we simply write  $s_{\text{min}}$  instead. We find

$$p_{\text{obs}}(l_p, z) = \int_{s_{\text{min}}}^{\infty} \sqrt{1 - \left( \frac{s_{\text{min}}}{s} \right)^{-\frac{2}{\zeta}}} f_S(s) ds. \tag{A.83}$$

The following approximation might facilitate the numerical evaluation of this integral. We note that for  $s \gg s_{\text{min}}$ , the square root factor in the integral approaches 1. Now split up the original integral in two parts, where  $\eta$  governs the approximation's accuracy:

$$\begin{aligned}
 p_{\text{obs}} & \approx \hat{p}_{\text{obs}} \\
 & = \int_{s_{\text{min}}}^{\eta s_{\text{min}}} \sqrt{1 - \left( \frac{s_{\text{min}}}{s} \right)^{-\frac{2}{\zeta}}} f_S(s) ds + \int_{\eta s_{\text{min}}}^{\infty} f_S(s) ds \\
 & = \int_{s_{\text{min}}}^{\eta s_{\text{min}}} \sqrt{1 - \left( \frac{s_{\text{min}}}{s} \right)^{-\frac{2}{\zeta}}} f_S(s) ds + 1 - F_S(\eta s_{\text{min}}),
 \end{aligned} \tag{A.84}$$

where we substitute numerically integrating to infinity for an evaluation of the CDF of the lognormally distributed RV  $S$ . The approximation error is bounded from above:

$$\begin{aligned}
 \hat{p}_{\text{obs}} - p_{\text{obs}} & = \int_{\eta s_{\text{min}}}^{\infty} \left( 1 - \sqrt{1 - \left( \frac{s_{\text{min}}}{s} \right)^{-\frac{2}{\zeta}}} \right) f_S(s) ds \\
 & < \left( 1 - \sqrt{1 - \left( \frac{s_{\text{min}}}{\eta s_{\text{min}}} \right)^{-\frac{2}{\zeta}}} \right) \int_{\eta s_{\text{min}}}^{\infty} f_S(s) ds \\
 & = \left( 1 - \sqrt{1 - \eta^{\frac{2}{\zeta}}} \right) (1 - F_S(\eta s_{\text{min}})) \\
 & < 1 - \sqrt{1 - \eta^{\frac{2}{\zeta}}}.
 \end{aligned} \tag{A.85}$$

Let us assume  $\zeta = -2$ . For  $\eta = 100$ , the approximation error is at most 0.005, and for  $\eta = 1000$ , the approximation error is at most 0.0005.

**Fanaroff–Riley class I** The simplest correction in which FRI RGs retain a well-defined notion of length assumes a linearly decreasing surface brightness, from some value  $b_v(0)$  at the core to zero at the RG's two endpoints. For a symmetric FRI RG, the surface brightness profile along one of the jets is  $b_v: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ , and depends on the projected proper distance from the core  $r_p$  as

$$b_v(r_p) = \begin{cases} b_v(0) \left( 1 - \frac{2r_p}{l_p} \right) & \text{if } 0 \leq r_p < \frac{l_p}{2} \\ 0 & \text{if } r_p \geq \frac{l_p}{2}. \end{cases} \tag{A.86}$$

Now we consider an FRI GRG, whose projected proper length  $l_p > l_{p,\text{GRG}}$  would only be observed in full in the absence of

noise. In actual observations, this GRG is detected as a GRG if and only if

$$b_v \left( \frac{l_{p,\text{GRG}}}{2} \right) > b_{v,\text{th}}, \quad \text{or } b_v(0) \left( 1 - \frac{l_{p,\text{GRG}}}{l_p} \right) > b_{v,\text{th}}. \tag{A.87}$$

Under our assumption of a linear surface brightness profile, the mean surface brightness along the jet axis  $\langle b_v \rangle = b_v \left( \frac{1}{2} \cdot \frac{l_p}{2} \right)$ . As this must be half of the surface brightness at the core,

$$\langle b_v \rangle = b_v \left( \frac{l_p}{4} \right) = \frac{b_v(0)}{2}. \tag{A.88}$$

Combining Eqs. A.87 and A.88, we find that the GRG is detected as such if

$$\langle b_v \rangle > \frac{b_{v,\text{th}}}{2 \left( 1 - \frac{l_{p,\text{GRG}}}{l_p} \right)}. \tag{A.89}$$

Now regarding  $\langle b_v \rangle$  as an RV and recognising that it might behave exactly as in Eq. A.76, we find that the surface brightness selection effect for FRI giants may be modelled as

$$p_{\text{obs}}(l_p, z) = \mathbb{P} \left( B_v > \frac{b_{v,\text{th}}}{2 \left( 1 - \frac{l_{p,\text{GRG}}}{l_p} \right)} \mid L_p = l_p, Z = z \right). \tag{A.90}$$

We see that the full formulaic structure of  $p_{\text{obs}}(l_p, z)$  is the same for FRI and FRII giants, except that for FRI giants a change

$$b_{v,\text{th}} \rightarrow \frac{b_{v,\text{th}}}{2 \left( 1 - \frac{l_{p,\text{GRG}}}{l_p} \right)} \tag{A.91}$$

is necessary. There is no change for  $l_p = 2l_{p,\text{GRG}}$ .

### A.9. GRG number density

A statistic of major interest is the number density of giants in the contemporary Universe. Let  $n_{\text{GRG}}$  be the comoving GRG number density, so that

$$n_{\text{GRG}} = n \mathbb{P} \left( L_p > l_{p,\text{GRG}} \right) = n \left( 1 - F_{L_p} \left( l_{p,\text{GRG}} \right) \right). \tag{A.92}$$

$$N_{\text{GRG,obs}}(\Omega, z_{\text{max}}) := \int_{l_{p,\text{GRG}}}^{\infty} dN_{L_p, \text{obs}}(l_p, \Omega). \tag{A.93}$$

After invoking Eq. A.63 and isolating  $n$ , we obtain

$$n = \frac{\frac{4\pi}{\Omega} N_{\text{GRG,obs}}(\Omega, z_{\text{max}})}{\int_{l_{p,\text{GRG}}}^{\infty} \int_0^{z_{\text{max}}} f_{L_p}(l_p) p_{\text{obs}}(l_p, z) 4\pi r^2(z) \frac{dr}{dz} dz dl_p}. \tag{A.94}$$

Combining Eqs. A.8, A.9, A.92 and A.94 for  $l_{p,\text{GRG}} > l_{\text{min}}$ , we arrive at

$$\begin{aligned}
 n_{\text{GRG}} & = - \frac{l_{p,\text{GRG}}^{\xi+1}}{\xi+1} \cdot \frac{\frac{4\pi}{\Omega} N_{\text{GRG,obs}}(\Omega, z_{\text{max}})}{\int_{l_{p,\text{GRG}}}^{\infty} l_p^{\xi} \int_0^{z_{\text{max}}} p_{\text{obs}}(l_p, z) 4\pi r^2(z) \frac{dr}{dz} dz dl_p}.
 \end{aligned} \tag{A.95}$$

From observations, we know  $\Omega$  and can — for a given  $z_{\text{max}}$  — simply count  $N_{\text{GRG,obs}}(\Omega, z_{\text{max}})$ . Moreover, we can fit  $\xi$  and the parameters that occur in  $p_{\text{obs}}(l_p, z)$  (e.g.  $\frac{1}{2}(\phi_{\text{max}} - \phi_{\text{min}})$ ,  $b_{v,\text{ref}}$ , and  $\sigma_{\text{ref}}$ ) to the ECDF of  $L_{p,\text{obs}} \mid L_{p,\text{obs}} > l_{p,\text{GRG}}$ . We note that  $n_{\text{GRG}}$  does not depend on  $l_{\text{min}}$ , which drops out through the division. However,  $n_{\text{GRG}}$  does depend on cosmological parameters through the relation between cosmological redshift  $z$  and radial comoving distance  $r$ .

### A.10. GRG lobe volume-filling fraction

Assuming self-similar growth, the combined proper volume  $V$  of an RG's lobes and its intrinsic proper length  $l$  obey  $V \propto l^3$ . The constant of proportionality varies per RG and depends on the shape of the lobes; we treat it as an RV  $\Upsilon := \frac{V}{L^3}$ . Then the proper VFF of GRG lobes  $\text{VFF}_{\text{GRG}}(z) = \text{VFF}_{\text{GRG}}(z=0) \cdot (1+z)^3$ , where

$$\begin{aligned} \text{VFF}_{\text{GRG}}(z=0) &:= n_{\text{GRG}} \cdot \mathbb{E}[V | L_p > l_{p,\text{GRG}}] \\ &= n_{\text{GRG}} \cdot \mathbb{E}\left[\frac{V}{L^3} \cdot L^3 | L_p > l_{p,\text{GRG}}\right] \\ &= n_{\text{GRG}} \cdot \mathbb{E}[\Upsilon \cdot L^3 | L_p > l_{p,\text{GRG}}]. \end{aligned} \quad (\text{A.96})$$

Assuming that RGs grow self-similarly, so that shape does not reveal length,  $\Upsilon$  and  $L^3$  are conditionally independent given  $L_p > l_{p,\text{GRG}}$ :  $\Upsilon \perp L^3 | L_p > l_{p,\text{GRG}}$ . As a result,

$$\begin{aligned} \text{VFF}_{\text{GRG}}(z=0) &= n_{\text{GRG}} \cdot \mathbb{E}[\Upsilon | L_p > l_{p,\text{GRG}}] \cdot \mathbb{E}[L^3 | L_p > l_{p,\text{GRG}}] \\ &= n_{\text{GRG}} \cdot \mathbb{E}[\Upsilon] \cdot \mathbb{E}[L^3 | L_p > l_{p,\text{GRG}}]. \end{aligned} \quad (\text{A.97})$$

To obtain this last line, we once more exploit self-similarity:  $\Upsilon | (L_p > l_{p,\text{GRG}}) = \Upsilon$ . We can approximate  $\mathbb{E}[\Upsilon]$  by taking the mean of some  $\frac{V}{l^3}$  deduced from observations. A technical complication arises from the fact that  $\mathbb{E}[L^3 | L_p > l_{p,\text{GRG}}]$  does not exist for  $\xi \geq -4$  under our model. This is an artefact of the Pareto distribution assumption for  $L$ , which unrealistically features support over an infinitely long part of the real line:  $\{l \in \mathbb{R} | l > l_{\min}\}$ . This causes the expectation value integral to diverge for  $\xi \geq -4$ . An approximation to  $\text{VFF}_{\text{GRG}}(z=0)$  that works for all  $\xi$  is the lower bound

$$\text{VFF}_{\text{GRG}}(z=0) > n_{\text{GRG}} \cdot \mathbb{E}[\Upsilon] \cdot \mathbb{E}^3[L | L_p > l_{p,\text{GRG}}], \quad (\text{A.98})$$

which follows from Jensen's inequality. Here

$$\mathbb{E}[L | L_p > l_{p,\text{GRG}}] = l_{p,\text{GRG}} \frac{\Gamma(-\frac{\xi}{2} - 1) \Gamma(-\frac{\xi}{2} + 1)}{\Gamma(-\frac{\xi}{2} - \frac{1}{2}) \Gamma(-\frac{\xi}{2} + \frac{1}{2})}. \quad (\text{A.99})$$

Alternative approximation formulae, which use  $\Upsilon_p := \frac{V}{L_p^3}$ , are

$$\text{VFF}_{\text{GRG}}(z=0) > n_{\text{GRG}} \cdot \mathbb{E}[\Upsilon_p] \cdot \mathbb{E}^3[L_p | L_p > l_{p,\text{GRG}}] \quad (\text{A.100})$$

and

$$\text{VFF}_{\text{GRG}}(z=0) > n_{\text{GRG}} \cdot \mathbb{E}[\Upsilon_p] \cdot m_{L_p^3 | L_p > l_{p,\text{GRG}}}, \quad (\text{A.101})$$

where

$$m_{L_p^3 | L_p > l_{p,\text{GRG}}} = l_{p,\text{GRG}}^3 \cdot 2^{\frac{3}{-(\xi+1)}} \quad (\text{A.102})$$

is the median of the cubed projected proper length for giants. An advantage of these latter expressions is that there are more data available to estimate  $\mathbb{E}[\Upsilon_p]$  than there are to estimate  $\mathbb{E}[\Upsilon]$ .

### A.11. Unification model constraints from quasar and non-quasar giants

The quasar GRG probability  $p_Q$  is

$$\begin{aligned} p_Q &:= \frac{\mathbb{P}(L_{p,\text{obs}} \geq l_{p,\text{GRG}}, \sin \Theta \leq \sin \theta_{\max})}{\mathbb{P}(L_{p,\text{obs}} \geq l_{p,\text{GRG}})} \\ &= \mathbb{P}(\sin \Theta \leq \sin \theta_{\max} | L_{p,\text{obs}} \geq l_{p,\text{GRG}}) \\ &= \int_0^{\sin \theta_{\max}} \frac{\mathbb{P}(L_{p,\text{obs}} \geq l_{p,\text{GRG}} | \sin \Theta = x) f_{\sin \Theta}(x) dx}{\mathbb{P}(L_{p,\text{obs}} \geq l_{p,\text{GRG}})} \\ &= \int_0^{\sin \theta_{\max}} \frac{\mathbb{P}(L_p \geq l_{p,\text{GRG}} | \sin \Theta = x) f_{\sin \Theta}(x) dx}{\mathbb{P}(L_p \geq l_{p,\text{GRG}})} \\ &= \int_0^{\sin \theta_{\max}} \frac{(1 - F_L(\frac{l_{p,\text{GRG}}}{x})) f_{\sin \Theta}(x) dx}{1 - F_{L_p}(l_{p,\text{GRG}})}. \end{aligned} \quad (\text{A.103})$$

Now using Eqs. A.3 and A.8 and  $f_{\sin \Theta}(x) = \frac{x}{\sqrt{1-x^2}}$  over the domain of integration,

$$p_Q = \frac{4\Gamma(-\frac{\xi}{2} + 1)}{-(\xi + 1) \sqrt{\pi} \Gamma(-\frac{\xi}{2} - \frac{1}{2})} \int_0^{\sin \theta_{\max}} \frac{x^{-\xi} dx}{\sqrt{1-x^2}}. \quad (\text{A.104})$$

### A.12. Extreme giants in a sample

An interesting feature of the model is its ability to predict the occurrence of giants with extreme projected proper lengths in a GRG sample of, say, size  $N$ . Now consider some  $l_p > l_{p,\text{GRG}}$  — what is the probability  $p_{>l_p}$  that an observed GRG will have a projected proper length exceeding  $l_p$ ? Proceeding as in the derivation of Eq. A.12, we find

$$\begin{aligned} p_{>l_p}(l_p) &:= \mathbb{P}(L_{p,\text{obs}} > l_p | L_{p,\text{obs}} > l_{p,\text{GRG}}) \\ &= \frac{1 - F_{L_{p,\text{obs}}}(l_p)}{1 - F_{L_{p,\text{obs}}}(l_{p,\text{GRG}})}. \end{aligned} \quad (\text{A.105})$$

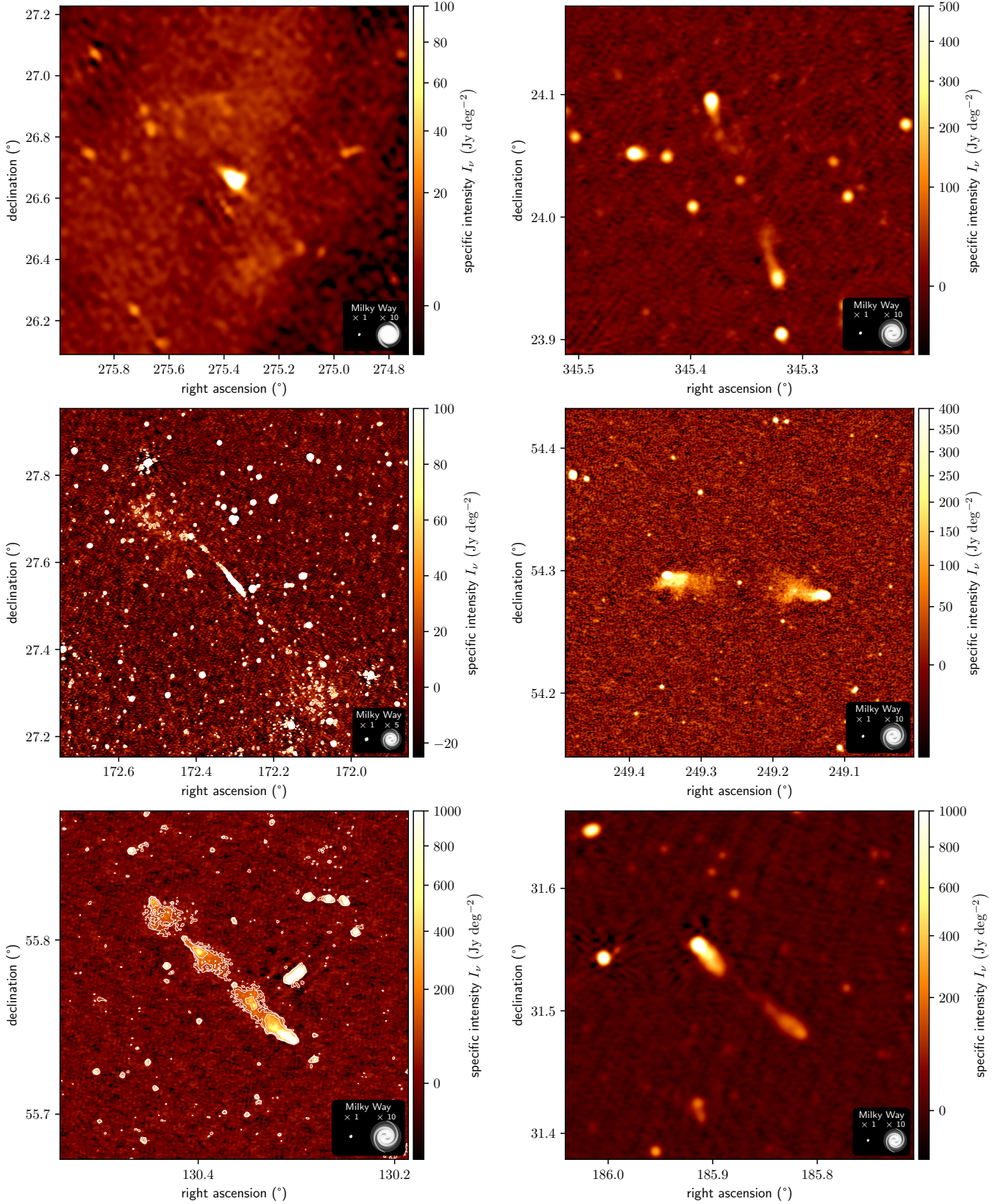
In the absence of selection effects,  $p_{>l_p}(l_p)$  is given by Eq. A.15. The number of giants with extreme projected proper lengths  $N_{>l_p} \sim \text{Binom}(N, p_{>l_p}(l_p))$ .

Interesting questions can be answered readily. For example, the probability that the sample contains at least one GRG with projected proper length  $l_p$  or larger, is

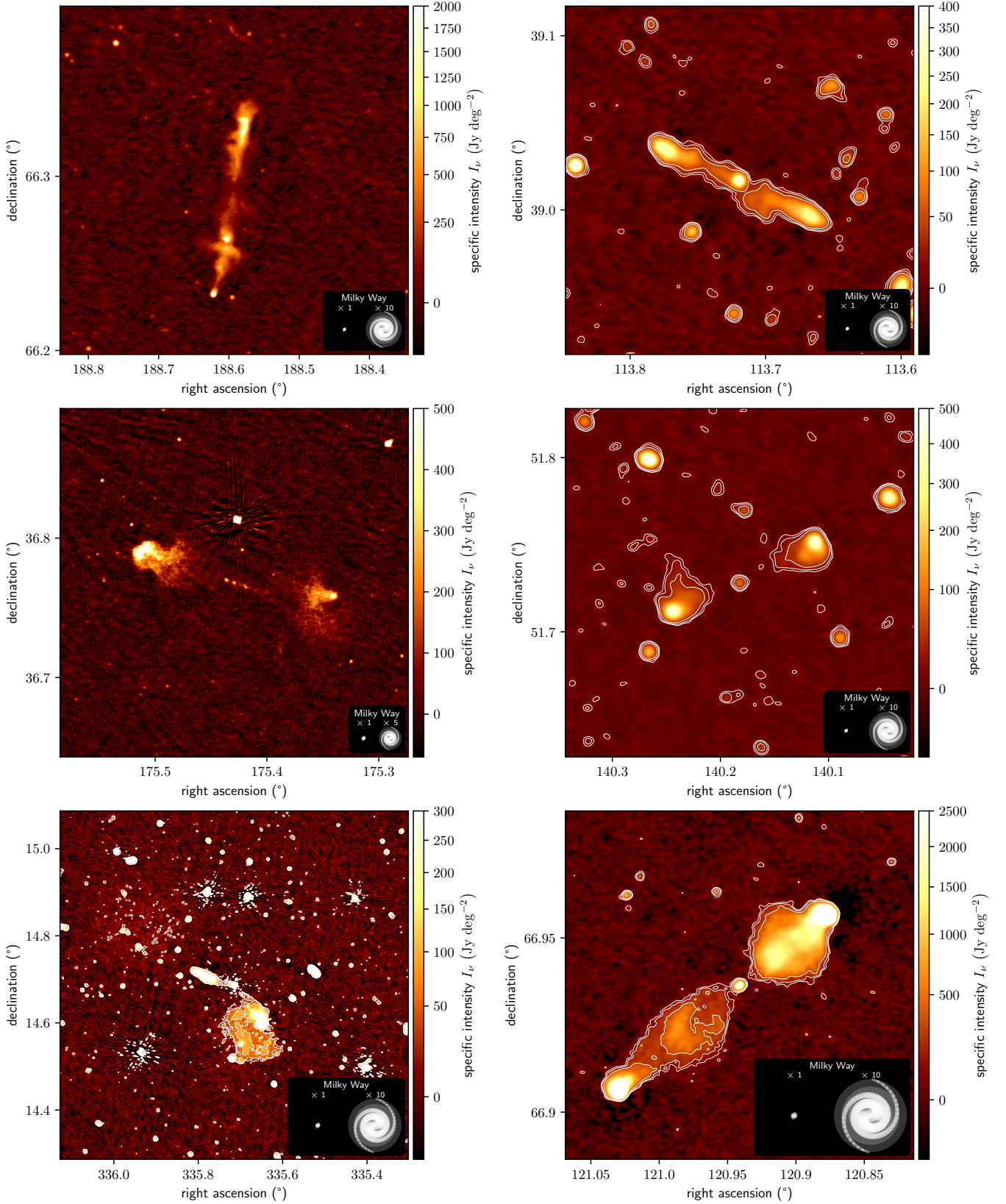
$$\mathbb{P}(N_{>l_p} \geq 1) = 1 - \mathbb{P}(N_{>l_p} = 0) = 1 - (1 - p_{>l_p}(l_p))^N. \quad (\text{A.106})$$

## Appendix B: Additional images

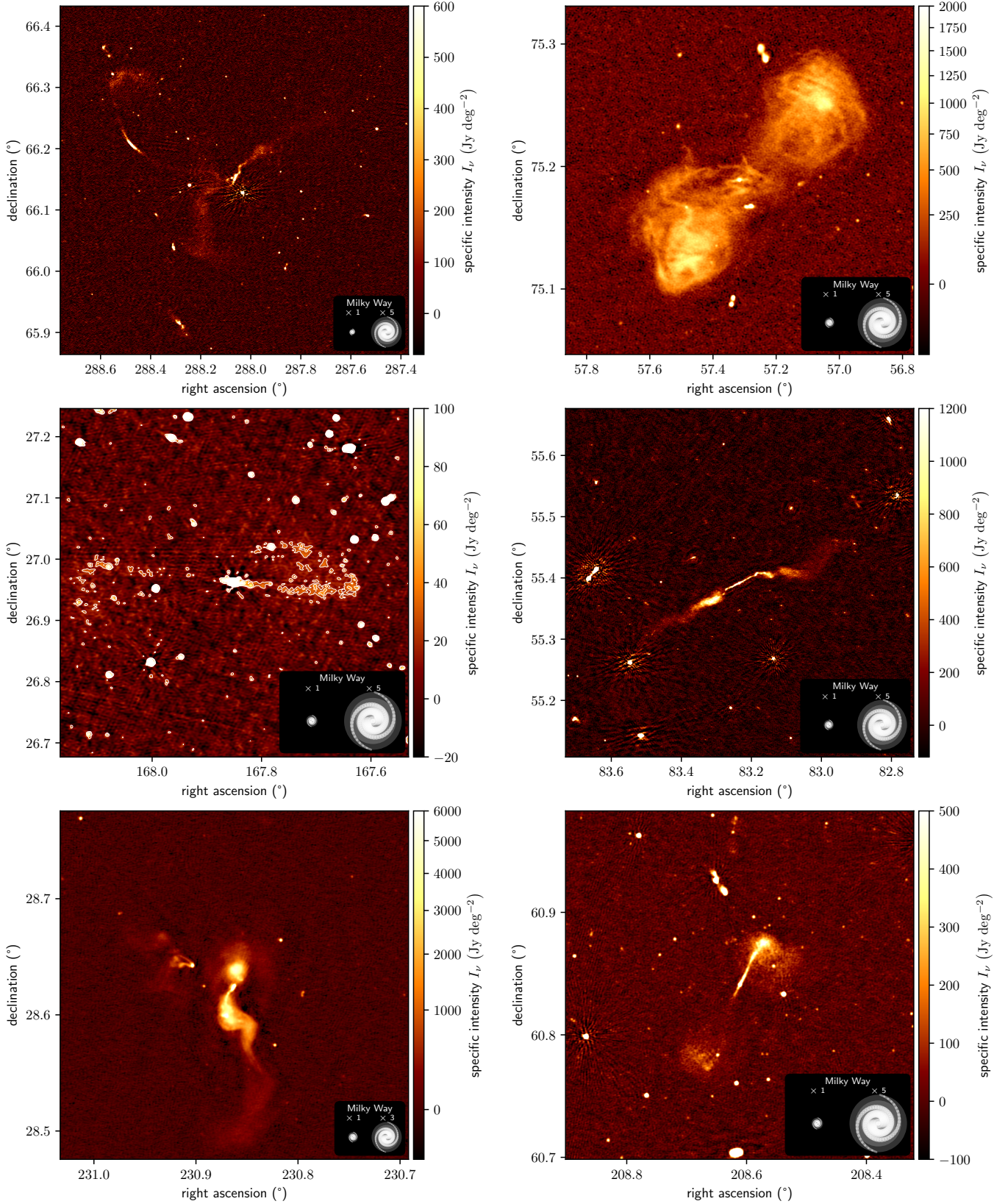
In this appendix, as in Fig. 12, we show newly discovered giants. These cover the projected length range  $l_p \in [0.7 \text{ Mpc}, 4 \text{ Mpc}]$ .



**Fig. B.1.** Details of the LoTSS DR2-estimated specific intensity function  $I_\nu(\hat{r})$  at central observing frequency  $\nu_{\text{obs}} = 144$  MHz and resolutions  $\theta_{\text{FWHM}} \in \{6'', 20'' 90''\}$ , centred around the hosts of newly discovered giants. Row-wise from left to right, from top to bottom, the projected proper length  $l_p$  is 3.9 Mpc, 3.5 Mpc, 3.3 Mpc, 3.3 Mpc, 3.3 Mpc, and 3.2 Mpc; in the same order,  $\theta_{\text{FWHM}}$  is  $90''$ ,  $20''$ ,  $20''$ ,  $6''$ ,  $6''$ , and  $20''$ . The giants in the top-left and middle-left panels appear larger in the sky than the Moon. Contours signify 3, 5 and 10 sigma-clipped standard deviations above the sigma-clipped median. For scale, we show the stellar Milky Way disk (with a diameter of 50 kpc) generated using the [Ringermacher & Mead \(2009\)](#) formula, alongside a 5 or 10 times inflated version.



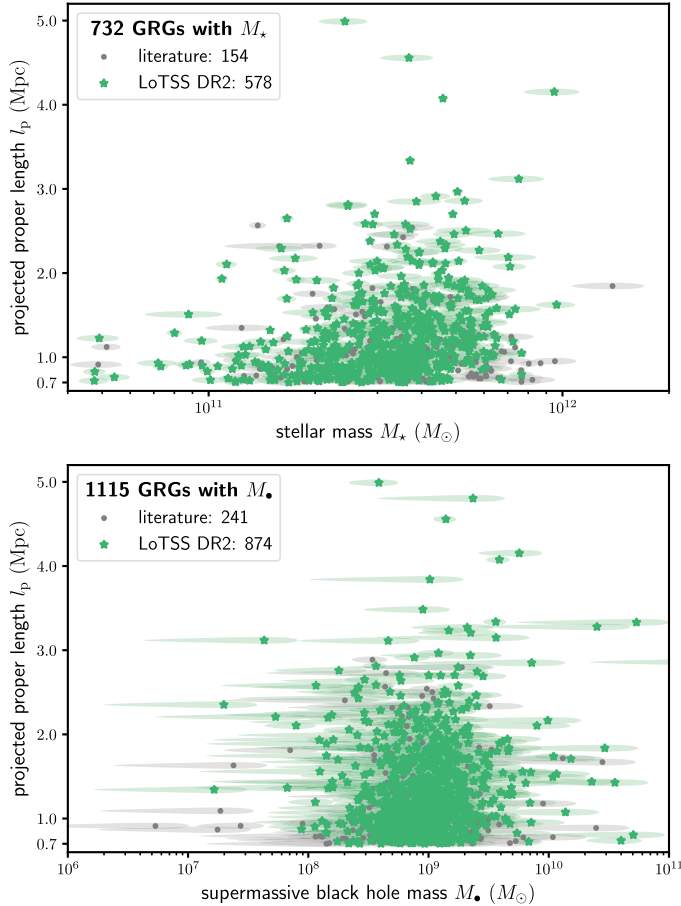
**Fig. B.2.** Details of the LoTSS DR2-estimated specific intensity function  $I_\nu(\hat{r})$  at central observing frequency  $\nu_{\text{obs}} = 144$  MHz and resolutions  $\theta_{\text{FWHM}} \in \{6'', 20''\}$ , centred around the hosts of newly discovered giants. Row-wise from left to right, from top to bottom, the projected proper length  $l_p$  is 2.8 Mpc, 2.6 Mpc, 2.6 Mpc, 2.2 Mpc, 2.1 Mpc, and 2.1 Mpc; in the same order,  $\theta_{\text{FWHM}}$  is  $6''$ ,  $20''$ ,  $6''$ ,  $20''$ ,  $20''$ , and  $6''$ . The GRG in the bottom-left panel appears larger in the sky than the Moon. Contours signify 3, 5, and 10 sigma-clipped standard deviations above the sigma-clipped median. For scale, we show the stellar Milky Way disk (with a diameter of 50 kpc) generated using the [Ringermacher & Mead \(2009\)](#) formula, alongside a 5 or 10 times inflated version.



**Fig. B.3.** Details of the LoTSS DR2-estimated specific intensity function  $I_\nu(\hat{r})$  at central observing frequency  $\nu_{\text{obs}} = 144$  MHz and resolutions  $\theta_{\text{FWHM}} \in \{6'', 20''\}$ , centred around the hosts of newly discovered giants. Row-wise from left to right, from top to bottom, the projected proper length  $l_p$  is 1.6 Mpc, 1.5 Mpc, 1.3 Mpc, 1.2 Mpc, 1.1 Mpc, and 0.7 Mpc; in the same order,  $\theta_{\text{FWHM}}$  is  $6''$ ,  $6''$ ,  $20''$ ,  $6''$ ,  $6''$  and  $6''$ . Contours signify 3, 5, and 10 sigma-clipped standard deviations above the sigma-clipped median. For scale, we show the stellar Milky Way disk (with a diameter of 50 kpc) generated using the [Ringermacher & Mead \(2009\)](#) formula, alongside a 3 or 5 times inflated version.

## Appendix C: Stellar and supermassive black hole masses

In Fig. C.1, we present the SDSS-derived relations between host stellar mass and projected proper length, and between host supermassive black hole mass and projected proper length, for all giants in our final catalogue. We obtained the data as in Section 3.7 of Oei et al. (2022a).



**Fig. C.1.** Observed relations between host stellar mass  $M_*$  and GRG projected length  $l_p$  (top) and between host SMBH mass  $M_*$  and GRG projected length  $l_p$  (bottom). Our LoTSS DR2 sample confirms that luminous giants typically have  $M_* \in 10^{11}$ – $10^{12} M_\odot$  and  $M_* \in 10^8$ – $10^{10} M_\odot$ . The sample effects an almost fivefold increase in the number of giants with SDSS-derived host stellar masses and SMBH masses. We do not show or count giants for which only a nearest host candidate could be determined.

## Appendix D: Surface brightness prior

Consider a radio galaxy at cosmological redshift  $z$  of intrinsic proper length  $l$  bounded by spherical lobes of radius  $R$  and spectral index  $\alpha$ . If a fraction  $f_l$  of the radio galaxy's central axis lies within the lobes, then

$$R = \frac{1}{4} f_l \cdot l. \quad (\text{D.1})$$

If a fraction  $f_{L_\nu}$  of the radio galaxy's total luminosity density  $L_\nu$  comes from the lobes, then the monochromatic emission coefficient

$$j_\nu = \frac{L_\nu f_{L_\nu}}{4\pi} \frac{1}{2} \frac{1}{V}, \quad (\text{D.2})$$

where the lobe volume  $V = \frac{4}{3}\pi R^3$ . A formula of practical value features projected proper length  $l_p$  instead of  $l$ . Given the approximate nature of our approach, we therefore simply assume  $l \approx \mathbb{E}[D](\eta(f_l)) \cdot l_p$ , with the deprojection factor expectation  $\mathbb{E}[D]$  given in Eq. A.29;  $\eta(f_l) = \frac{f_l}{2-f_l}$ . The maximum surface brightness of the lobes as seen by an observer is

$$b_{\nu, \max} = \frac{3f_{L_\nu} \cdot L_\nu}{\pi^2 \cdot \mathbb{E}[D](\eta(f_l)) \cdot f_l^2 \cdot l_p^2 \cdot (1+z)^{3-\alpha}}, \quad (\text{D.3})$$

valid for the line of sight that pierces through a lobe along a diameter. The average line of sight length  $d$  within a lobe is smaller than  $d_{\max} = 2R$ , though, and given by

$$\langle d \rangle := \frac{\int_0^R 2\sqrt{R^2 - x^2} 2\pi x dx}{\pi R^2} = \frac{4}{3}R. \quad (\text{D.4})$$

Therefore, the mean surface brightness of the lobes as seen by an observer is

$$\langle b_\nu \rangle = \frac{\langle d \rangle}{d_{\max}} b_{\nu, \max} = \frac{2}{3} b_{\nu, \max}. \quad (\text{D.5})$$

## Appendix E: Likelihood function

In Table E.1, we present maximum likelihood and likelihood mean and standard deviation estimates for the inference described in Section 4.1.4; one may compare the results to those in Table 4. In Fig. E.1, we visually summarise the likelihood function; one may compare to Fig. 16.

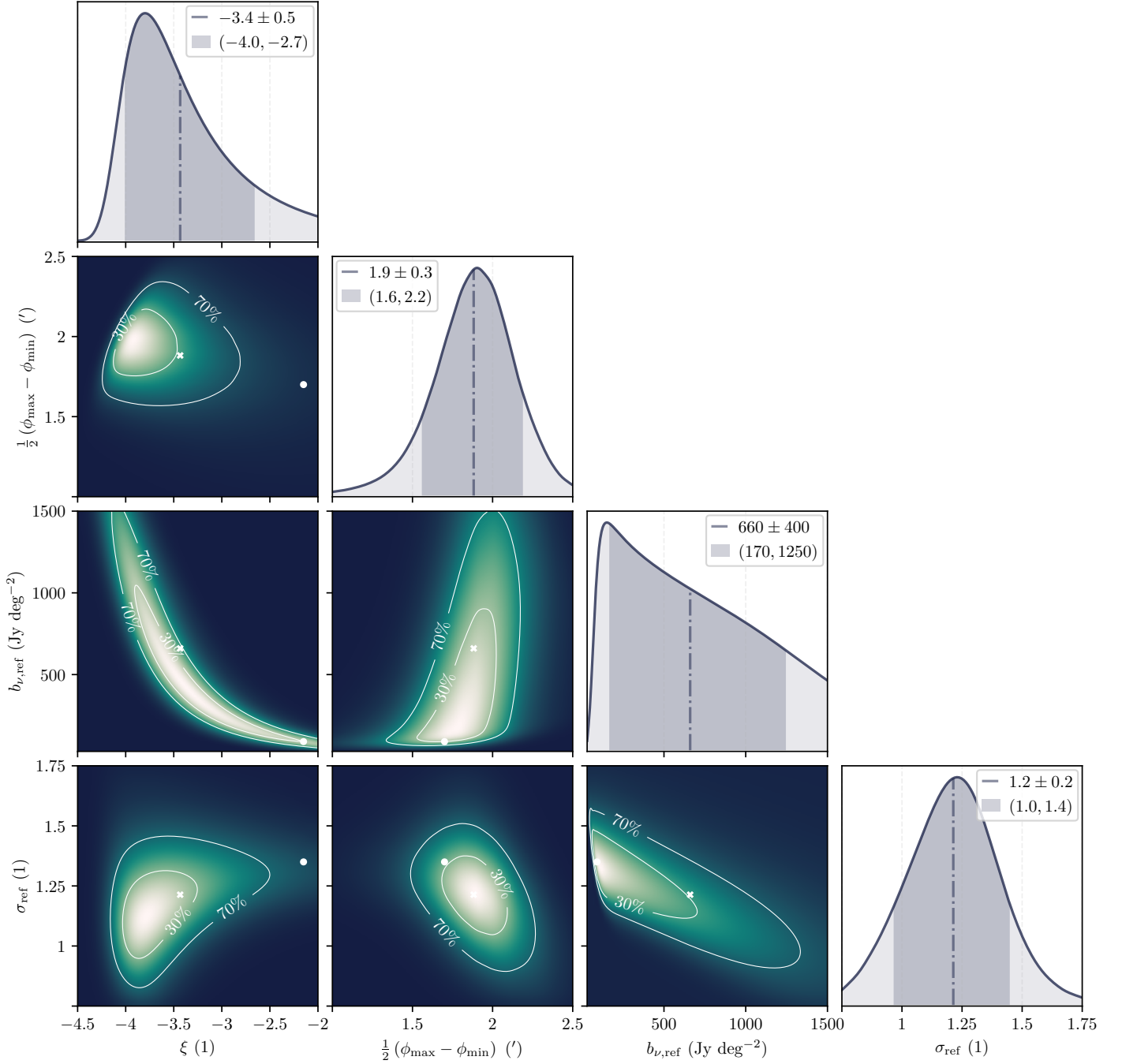
The strong degeneracy between  $\xi$  and  $b_{\nu, \text{ref}}$ , directly apparent from the central two-parameter marginal in the leftmost column of Fig. E.1, translates to a ridge of essentially constant likelihood that extends from  $\xi \approx -4$  to  $\xi \approx -2$ . Compared to a non-degenerate case, this makes the maximum likelihood parameters both intrinsically less meaningful and more prone to numerical approximation errors.

**Table E.1.** Maximum likelihood estimate (MLE) and likelihood mean and standard deviation (SD) estimates of the free parameters in intrinsic GRG length distribution inference.<sup>1</sup>

$z_{\max} = 0.5$		
parameter	MLE	likelihood mean and SD
$\xi$	-2.15	$-3.4 \pm 0.5$
$\frac{1}{2}(\phi_{\max} - \phi_{\min})$	1.7'	$1.9 \pm 0.3'$
$b_{\nu, \text{ref}}$	90 Jy deg <sup>-2</sup>	$660 \pm 400$ Jy deg <sup>-2</sup>
$\sigma_{\text{ref}}$	1.35	$1.2 \pm 0.2$
$z_{\max} = 0.25$		
parameter	MLE	likelihood mean and SD
$\xi$	-2.2	$-3.5 \pm 0.4$
$\frac{1}{2}(\phi_{\max} - \phi_{\min})$	2.25'	$1.7 \pm 0.3'$
$b_{\nu, \text{ref}}$	150 Jy deg <sup>-2</sup>	$1050 \pm 560$ Jy deg <sup>-2</sup>
$\sigma_{\text{ref}}$	1.25	$1.4 \pm 0.4$

<sup>1</sup> The model assumes  $\xi$  is constant for  $z \in [0, z_{\max}]$ . We determined the likelihood function twice: for  $z_{\max} = 0.5$ , using 1473 giants, and for  $z_{\max} = 0.25$ , using 811 giants. We caution that, in this case, the MLE parameters are a poor measure of central tendency.





**Fig. E.1.** Joint likelihood function over  $\xi$  — the parameter of interest — and  $\frac{1}{2}(\phi_{\max} - \phi_{\min})$ ,  $b_{\nu,\text{ref}}$  and  $\sigma_{\text{ref}}$  — the selection effect parameters, based on 1473 projected lengths of LoTSS DR2 giants up to  $z_{\max} = 0.5$ . We show all two-parameter marginals of the likelihood function, with contours enclosing 30% and 70% of total probability. We mark the maximum likelihood parameters (white circle) and the likelihood mean parameters (white cross). The single-parameter marginals again show the estimated mean, now marked by a vertical line, alongside shaded median-centred 80% credible intervals. The likelihood function is the posterior for a uniform prior. To compare the likelihood function to the posterior actually chosen, see Fig. 16.

## Appendix F: Properties of newly discovered giants

Table 2 provides properties of the 50 projectively longest giants discovered during this work's LoTSS DR2 search campaign. We share these data, alongside those for the other 2010 (98%)

giants in our sample, in Flexible Image Transport System (FITS) format through the Centre de Données astronomiques de Strasbourg (CDS).

For our final catalogue, which also includes literature giants, please contact the authors.