Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **553,** 1–14 (2026)
Advance Access publication 2026 October 19

https://doi.org/10.1093/mnras/stag1947

A MIGHTEE robust measurement of the star formation rate–radio correlation

Charlotte L. Jackson [1]★ Matt J. Jarvis [1,2] Imogen H. Whittam [1,2] James H. Matthews [1] Johannes Buchner,[3,4] Catherine L. Hale [1,5] Joel Hamlett [1] S. Lyla Jung [1] Mara Salvato [3,4] Daniel J. B. Smith [6] Natalia Stylianou [1] and Rohan G. Varadaraj [1]

[1]*Department of Physics, Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[2]*Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, 7535 Bellville, Cape Town, South Africa*
[3]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85741 Garching, Germany*
[4]*Excellence Cluster Universe, Boltzmannstr 2, D-85748 Garching, Germany*
[5]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[6]*Centre for Astrophysics Research, University of Hertfordshire, Hatfield, Herts AL10 9AB, UK*

Accepted 2026 October 8. Received 2026 September 25; in original form 2026 July 1

ABSTRACT

Determining the relationship between star formation rate (SFR) and the radio luminosity ($L_{1.4}$) is critical if we are to trace the star formation history of the Universe dust-agnostically using current and future radio facilities. However, until now, such work has relied on potentially biased binary classifications of sources to remove contaminating active galactic nuclei (AGNs). We present a new, statistically driven methodology for deriving the SFR–$L_{1.4}$ relation, removing the need for problematic cuts. We use a Bayesian hierarchical mixture model fit to the radio-detected sources in the deep MIGHTEE COSMOS DR1 catalogue, incorporating the full SFR posterior probability distributions generated by state-of-the-art spectral energy distribution fitting code GRAHSP. This allows us to probabilistically determine a mean SFR–$L_{1.4}$ relation for the SF-dominated galaxies, whilst accounting for changing fractions of SF dominated sources across redshift, radio luminosity, and stellar mass ranges. We find that the SFR–radio luminosity correlation exhibits a significant dependence on redshift, but a stellar mass dependence that is weaker than previous studies. Our resultant SFR–radio correlation is $\log_{10}(\mathrm{SFR}/\mathrm{M}_\odot\,\mathrm{yr}^{-1}) = 0.790 \times (\log_{10}(L_{1.4}/\mathrm{W\,Hz}^{-1}) - 23) + 1.244 \times (1+z)^{0.122} - 0.033 \times (\log_{10}(M_*/\mathrm{M}_\odot) - 10)$, with an intrinsic scatter of 0.178 dex. We show that this redshift evolution could be explained by a moderate evolution in the radio spectral index of SF galaxies. We attribute the lack of observed strong dependence on stellar mass, compared to recent studies, to the novel statistical approach that does not rely on cuts to remove AGNs.

Key words: galaxies: evolution – galaxies: star formation – radio continuum: galaxies.

1 INTRODUCTION

Observations of the radio continuum emission from galaxies offer powerful insights into important physical processes, unaffected by intervening dust. In (radio loud) active galactic nuclei (AGNs), emission from jets and outflows generally dominate these wavelengths (A. Merloni & S. Heinz 2007; M. J. Hardcastle & L. W. Looney 2008; N. L. Zakamska & J. E. Greene 2014; S. V. White et al. 2017; F. Panessa et al. 2019; V. A. Fawcett et al. 2020; A. L. Rankine et al. 2021; C. L. Jackson et al. 2026), whereas in star-forming galaxies (SFG), it is primarily star formation (SF) activity that gives rise to the radio emission (see J. J. Condon 1992, for a review). The connection between SF activity and radio emission arises due to the fact that the most luminous, short-lived stars end in supernovae, and these hugely energetic events accelerate cosmic rays (CRs) to relativistic speeds. The CR electrons and positrons then interact with diffuse galactic magnetic fields, producing synchrotron radiation (B. T. Draine 2011; R. S. Klessen & S. C. O. Glover 2016). Therefore, 1.4 GHz radio observations can be used as a delayed tracer of SF activity and have been shown on many occasions to strongly correlate with the total star formation rate (SFR) in a galaxy.

There has been considerable work over many decades to calibrate the empirical relationship between SFR and 1.4 GHz radio emission, first often relying on the far-infrared (FIR) emission as a tracer of star formation through the FIR–radio correlation (e.g. G. Helou, B. T. Soifer & M. Rowan-Robinson 1985; T. Jong et al. 1985; M. S. Yun, N. A. Reddy & J. J. Condon 2001; P. N. Appleton et al. 2004; E. Ibar et al. 2008; R. J. Ivison et al. 2010; M. J. Jarvis et al. 2010; J. Delhaize et al. 2017; H. S. B. Algera et al. 2020; I. Delvecchio et al. 2021), or using dust-corrected emission lines (e.g. K. J. Duncan et al. 2020; N. Seymour et al. 2026), and also multiwavelength estimates of the SFR (e.g. L. Cram et al. 1998; E. F. Bell et al. 2003; J. A. Hodge et al. 2008; T. Garn et al. 2009; E. J. Murphy et al. 2011; F. S. Tabatabaei et al. 2017; G. Gürkan et al.

★ E-mail: charlotte.jackson@physics.ox.ac.uk

© The Author(s) 2026.
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

2018; D. J. B. Smith et al. 2021; V. Heesen et al. 2022; R. H. W. Cook et al. 2024). It is generally agreed that the SFR–radio relationship is approximately linear in log–log space, however there is a lack of consensus regarding the universality of any such relation across redshifts, stellar masses and morphology (e.g. J. Delhaize et al. 2017; D. C. Molnár et al. 2018; I. Delvecchio et al. 2021; D. C. Molnár et al. 2021; D. J. B. Smith et al. 2021; R. H. W. Cook et al. 2024).

There are multiple physical reasons to expect some kind of evolution of the infrared–radio correlation (IRRC) and/or the SFR–radio correlation with redshift and stellar mass (M_*). The connection between the infrared (IR) and radio emission in a star forming galaxy hinges on the ability of CR electrons to propagate through the galaxy, and it would be reasonable to assume that this propagation depends on, for example, the mass of that galaxy (see e.g. D. J. B. Smith et al. 2021). Furthermore, the calculation of radio luminosity, L_ν, from flux, S_ν, depends on a spectral index parameter such that $S_\nu \propto \nu^\alpha$, where α is often taken to be -0.7 for non-thermal synchrotron emission (M. Novak et al. 2017; V. Smolčić et al. 2017; R. Kondapally et al. 2022), which is dominant at 1.4 GHz. Any evolution of average radio spectral shape with redshift, either from a steepening from synchrotron cooling (e.g. F. An et al. 2021, 2024; O. Bait et al. 2026), the increasing contribution from thermal free–free emission (e.g. G. Calistro Rivera et al. 2017) or inverse Compton scattering off of cosmic microwave background photons at higher redshifts (G. De Zotti et al. 2024; I. H. Whittam et al. 2025), would therefore induce a redshift dependence in the IRRC and the SFR–radio correlations. There are many such potential explanations for an observed redshift evolution in either the IRRC and the SFR–radio correlation, however as the majority of the recent work on this topic has focussed on the IRRC rather than a direct SFR–radio correlation, it is unclear whether any observed evolution is driven by a decoupling of the radio or the IR as a reliable SFR indicator.

Furthermore, as work in this area has begun to benefit from deep, wide-field radio surveys carried out on instruments such as MeerKAT (J. L. Jonas 2009; J. Jonas 2018), LOFAR (M. P. van Haarlem et al. 2013), and GMRT (G. Swarup 1991; Y. Gupta et al. 2017), the issue of cleanly identifying truly SF-dominated sources within catalogues comprised of both faint AGNs and SFGs has become increasingly pertinent. Any SFR–radio correlation (or associated IRRC) must of course be calibrated only on sources where it can be safely assumed that the radio emission is a direct result of SF activity. In order to select appropriate samples free from contamination from AGNs, where the radio emission is likely dominated by jets and other outflows, selection criteria based on IR colours, optical emission line diagnostics, very-long baseline interferometry, or X-ray emission have all been employed in the past to eliminate AGNs from radio samples (e.g. G. P. Szokoly et al. 2004; J. L. Donley et al. 2012; N. Herrera Ruiz et al. 2017; I. H. Whittam et al. 2022; A. B. Drake et al. 2024; M. I. Arnaudova et al. 2025; G. Mountrichas et al. 2025; G. Peluso et al. 2026).

The use of the IRRC to calibrate an SFR–radio correlation is intrinsically highly assumption-driven. In particular, such an approach assumes that the IR is a perfect tracer of SFR for all galaxies – which is very unlikely to be true. For example, in low metallicity and high redshift galaxies, the UV emission of a galaxy can be as important to constraining SFR as IR (K. E. Whitaker et al. 2017), which is unaccounted for in the use of an IRRC. Mass and redshift dependent deviations between the IR and radio emission of a galaxy can also induce evolution in the IRRC that is unrelated to SFR. Furthermore, it is likely that highly obscured AGNs, which show no visible-wavelength signs of an AGN component, can heat the dust within both a putative torus and the wider interstellar medium, thus elevating the FIR emission above the level one would expect from SF alone (e.g. D. G. Bonfield et al. 2011; M. Symeonidis et al. 2016; N. Maddox et al. 2017; M. Symeonidis et al. 2022; A. D. Sokol et al. 2023). However, whilst multiwavelength spectral energy distribution (SED) fitting (e.g. G. Calistro Rivera et al. 2016; G. Yang et al. 2022) can be used to help determine this level of emission for bright, disc-dominated AGNs, there remains the tricky subject of identifying those AGNs that do not exhibit typical AGN characteristics (e.g. accretion discs, broad-line regions) other than a powerful radio jet, often referred to as low-excitation radio galaxies, or LERGs. Past studies have resorted to removing sources that show an excess of radio emission relative to their IR (i.e. have a particularly low $q_{\rm IR}$) *before then fitting for the IRRC*. This methodology runs the risk of being circular, i.e. a given SFR–radio correlation is assumed and sources with 'excess radio emission' are removed, and then the correlation is measured, and is also dependant on the (subjective) choice of cut location.

In this paper, we aim to address some of these critical issues surrounding the calibration of the correlation, and present an improved measurement of the SFR–radio relationship using a rich dataset combining deep radio imaging, exceptional multiwavelength ancillary data, and state-of-the-art SED fitting. We choose not to employ any form of binary SFG–AGN pre-classification, and instead use a hierarchical Bayesian mixture model to assign sources a likelihood of belonging to a 'SF-dominated' or 'not SF-dominated' cluster. We also use an SFR derived from detailed SED modelling, using multiwavelength data from the FUV-FIR. We marginalize over the posterior density distributions of the derived SFRs, include considerations of observational and modelling uncertainties, and are able to identify degeneracies between any mass and redshift evolution of the derived SFR–radio correlation.

In Section 2, we describe the observational data used in our analysis and outline the SED fitting procedure. In Section 3, we explain the details of the model used, and in Section 4 present our improved form of the SFR–radio correlation, and discuss any derived evolution with galaxy properties. We assume a Planck Collaboration VI (2020) flat Λ cold dark matter cosmology with $\Omega_{\rm m} = 0.31$, $\Omega_\Lambda = 0.69$, $H_0 = 67.66\,{\rm km\,s^{-1}\,Mpc^{-1}}$. Energies, frequencies, and wavelengths are given in the rest frame.

2 OBSERVATIONAL DATA

2.1 Multiwavelength measurements

The MeerKAT International Tiered GHz Extragalactic Exploration (MIGHTEE; M. J. Jarvis et al. 2016) is a deep radio survey underway using the MeerKAT radio telescope. The survey is being conducted over 20 deg^2 of sky across four different well-studied fields, namely the Cosmic Evolution Survey (COSMOS), XMM Large-Scale Structure (XMM-LSS), European Large-Area Infrared Space Observatory S1 (ELAIS-S1), and Extended Chandra Deep Field South (CDFS) fields. The resulting images have central rms sensitivities of $\sim$1.3–2.7 μJy beam^{-1} at $\sim$1.2–1.3 GHz – see C. L. Hale et al. (2025) for more details on the Data Release 1 (DR1). We use the host-galaxy associated catalogue (C. L. Hale et al. 2026) of the COSMOS field continuum DR1, which contains 20 859 sources in total. Of these, 13 782 lie within the multiwavelength cross-matching region described in C. L. Hale et al. (2026), and subsequently 13 155 have a successful host-galaxy association

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

(95.45 per cent) and 13 117 sources have a redshift assigned. Approximately 55 per cent of sources have spectroscopic redshifts from the catalogues of M. Vaccari (2015, 2026),[1] which contain a compilation of measurements from a variety of spectroscopic catalogues, e.g. L. J. M. Davies et al. (2025) and A. A. Khostovan et al. (2026). 637 sources have their redshifts taken from the Physics of the Accelerating Universe (PAU) survey (A. Alarcon et al. 2021). Where neither spectroscopic nor PAU redshifts are available, then photometric redshifts from N. Stylianou et al. (2026) are used, many of which have their origins in the catalogues of P. W. Hatfield et al. (2022). We do not sample the posteriors of the photometric redshift measurements in our analysis, and refer the interested reader to C. L. Hale et al. (2026) for a full description of the redshift assignment process of the sample, and P. W. Hatfield et al. (2022) and N. Stylianou et al. (2026) for more detail about the photometric redshift methodology.

The Stylianou et al. catalogue contains Visible Infrared Survey Telescope for Astronomy (VISTA) IR imaging from UltraVISTA Data Release 6 (H. J. McCracken et al. 2013), and optical *grizy* band photometry from DR3 of the HSC-SSP survey (H. Aihara et al. 2022), as well as *ugriz* data from the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS: J.-C. J. Cuillandre et al. 2012), and is magnitude limited to $K_s = 24.8$. We then supplement the catalogue further with mid- and far-IR data from the *Herschel* Extragalactic Legacy Project (HELP; R. Shirley et al. 2021). Specifically, we use the HELP XID+ photometry data, which have been probabilistically deblended using XID+, as described in Section 4.3 of R. Shirley et al. (2021). The advantage of this is that it mitigates issues regarding the combination of low resolution IR data with much higher resolution UV/optical data. We use the *Spitzer* Multiband Image Photometer (MIPS) 24 μm band, the *Herschel* Photoconductor Array Camera and Spectrometer (PACS) 100 and 160 μm bands, and the *Herschel* Spectral and Photometric Imaging Receiver (SPIRE) 250, 350, and 500 μm bands. 83 per cent of our sample are detected have MIPS and PACS data, and 87 per cent have detections in all three SPIRE bands. Finally, we access *Spitzer* Infrared Array Camera (IRAC) bands at 3.6, 4.5, 5.8, and 8.0 μm (available for 70 per cent, 71 per cent, 72 per cent, and 72 per cent of sources, respectively), and near and far ultraviolet (NUV/FUV) Galaxy Evolution Explorer (GALEX) bands at 0.23 and 0.15 μm from the Cosmic Evolution Survey (COSMOS) 2020 catalogue `Farmer` catalogue (J. R. Weaver et al. 2022). We use the COSMOS2020 `Farmer` catalogue, which uses photometry measured by fitting parametric galaxy profiles, taking into account variations in the point-spread function, depth, and source crowding. NUV (FUV) detections are available for 21 (38) per cent of sources due to limitations in GALEX sensitivity, particularly at higher redshifts where the Lyman-break is redshifted through the GALEX filters. To check the robustness of our work against the effect of non-detections, particularly at the far-IR wavelengths, we also repeated the analysis using only sources with full IR coverage from 1.02 to 500 μm (7482 sources, 57 per cent of our full sample) and our conclusions do not change. Indeed, the values parametrizing our final SFR–radio relationship are consistent within uncertainties.

[1]Using the 2025 March 20 version.

2.2 SED fitting

We conduct multi-band, FUV-FIR SED fitting with the inclusion of a flexible AGN component, using the new SED fitting code `GRAHSP` (J. Buchner et al. 2024) to derive galaxy properties such as M_* and SFR. `GRAHSP` builds upon the foundations of the well-established SED-fitting code `CIGALE` (M. Boquien et al. 2019; G. Yang et al. 2022), introducing a flexible, empirically validated AGN model component that is comprised of emission from an accretion disc in the form of a power-law continuum, along with broad and narrow lines and an Fe II forest, an IR torus component and appropriate galactic and AGN attenuation. This is combined with models of nebular emission and galactic stellar populations, the latter of which makes use of modules developed for `CIGALE`.

`GRAHSP` employs a Bayesian approach to constrain model parameters using a nested sampling inference procedure. For this study, we adopt a delayed-τ star formation history (the `sfhdelayed` module), a G. Chabrier (2003) initial mass function with a solar metallicity ($Z = 0.02$), and G. Bruzual & S. Charlot (2003) for the simple stellar population (SSP) module. We add nebular emission using the `nebular` module (J. A. Villa-Vélez et al. 2021). Dust attenuation is handled by the `biattenuation` module, which treats the AGN and galaxy components separately. Energy balance is enforced in that absorbed galactic luminosity is remitted in the infrared using the `galdale2014` module, following D. A. Dale et al. (2014). For the AGN components, we use the `activatepl` module to model the power-law continuum emission from the accretion disc, the `activatelines` module to account for both narrow and broad AGN emission lines, and the `activategtorus` module to add infrared emission associated with reprocessing of UV light by a dust torus. A complete description of the SED fitting carried out on the full MIGHTEE DR1 sample will be given in a following paper, Jackson et al. (in preparation).

3 METHOD

The aim of this work is to agnostically and self-consistently define the SFR–radio relationship using the radio-detected MIGHTEE DR1 sample. To do this, we use star formation rates derived from `GRAHSP` SED fitting, marginalizing over their full posterior probability distributions, and the 1.4 GHz monochromatic radio luminosities, $L_{1.4}$, calculated from the observed MIGHTEE flux densities as

$$L_{1.4} = 4\pi d_L^2 S_{\rm obs}(1+z)^{-\alpha-1}\left(\frac{1.4\,\text{GHz}}{\nu_{\rm obs}}\right)^{\alpha}\ \text{W Hz}^{-1}, \quad (1)$$

where d_L is the luminosity distance, $S_{\rm obs}$ is the observed flux density, z is the redshift, $\nu_{\rm obs}$ is the observed effective frequency quoted in the MIGHTEE DR1 catalogue, and α is the spectral index. Throughout, we assume $S_\nu \propto \nu^\alpha$ with $\alpha = -0.7$, unless stated otherwise. The analysis in this paper was repeated for $\alpha = -0.75$ and $\alpha = -0.8$, and it was verified that the conclusions were unchanged. For clarity, throughout the paper, $L_{1.4}$ is given in units of W Hz^{-1}, SFR is given in units of M$_\odot$ yr^{-1}, and M_* is given in units of M$_\odot$, unless stated otherwise. Fig. 1 shows the redshift versus radio luminosity for the sample used in this paper.

We show in Fig. 2 that the distribution of sources in $\log_{10}(L_{1.4})$–$\log_{10}$(SFR) space produces a clear sequence that corresponds to the expected SFR–radio correlation in SF-dominated sources, alongside a background of radio-excess sources whose radio emission is likely dominated by AGN activity. To preserve our agnosticism and allow our derived relationship to be entirely

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

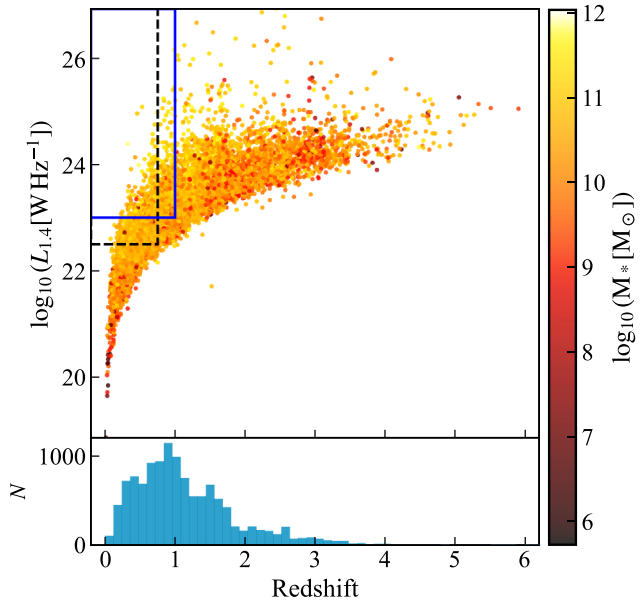


Figure 1. Top: 1.4 GHz radio luminosity as a function of redshift for the MIGHTEE DR1 (COSMOS field) sample used in this work, coloured by stellar mass. The solid and dashed lines indicate the regions used in the volume-limited test performed in Section 4.5. Bottom: Distribution of source redshifts. The median (mean) redshift of the sample is $z = 0.99$ ($z = 1.16$), and the 16th (84th) percentile is $z = 0.46$ ($z = 1.76$). 73 sources lie below $z < 0.1$, and 72 beyond $z > 4$.

data-driven, we choose to model both components simultaneously, and infer an SFG–AGN separation probabilistically, rather than removing sources that are likely to be AGN 'by hand'. The following sections detail how our model and associated SFR–$L_{1.4}$ relationship is computed.

3.1 Dataset preparation

Whilst we wish to remain as agnostic to our sample as possible, it is however necessary to remove sources with unphysical or poorly constrained physical properties. Therefore, we restrict our sample to the redshift range $0.01 < z < 6$, and remove any sources with an extremely low median SFR, $<10^{-5}$ $\mathrm{M_\odot\,yr^{-1}}$. The exact location of the lower SFR limit does not impact our conclusions. We also quantify the relative size of the SFR posteriors, $\frac{\Delta\mathrm{SFR}}{\mathrm{SFR}}$, as the difference between the 84th and 16th percentile values, scaled by the median, and remove any source with $\log_{10}\left(\frac{\Delta\mathrm{SFR}}{\mathrm{SFR}}\right) > 1$. The vast majority of sources are well-constrained, with 87.6 per cent having $\log_{10}\left(\frac{\Delta\mathrm{SFR}}{\mathrm{SFR}}\right) < 0.1$, and 95.7 per cent of sources have $\log_{10}\left(\frac{\Delta\mathrm{SFR}}{\mathrm{SFR}}\right) < 0.5$, and the $\log_{10}\left(\frac{\Delta\mathrm{SFR}}{\mathrm{SFR}}\right)$ cut removes 239 sources (1.8 per cent of the sample). The upper and lower redshift bounds remove 9 and 11 sources (0.07/0.08 per cent of the total sample), respectively, and 306 sources do not meet SFR median criteria (2.3 per cent). The final sample meeting all quality criteria contains 12 083 sources. We discuss the effects of completeness in our sample in Section 4.5.

Then, for numerical stability and to reduce fitted parameter degeneracies, observed quantities are centred around typical values for the population:

$$L' = \log_{10}(L_{1.4}) - 23, \quad M' = \log_{10}(M_*) - 10. \tag{2}$$

Redshift is not centred, as it will ultimately be given a non-linear parametrization. We want to preserve non-Gaussianity in the SFR posterior distribution whilst improving computational efficiency, so compress each posterior by drawing $R = 40$ representative samples (though the overall results are insensitive to this choice), and subsequently assigning each sample a relative weight of $w_{i,r} = 1/R$, for the ith source and rth sample.

3.2 Defining the Bayesian hierarchical mixture model

We model the distribution as a two-component hierarchical Bayesian mixture: a primary SF-dominated relation and a background (non-SF dominated) component. The SF-dominated component is described by one-dimensional Gaussian scatter in SFR around a mean relation, and the background as a one-dimensional Gaussian in SFR space. We verified that the results do not change, but equally the overall fit is not improved when the background Gaussian component is allowed to be skewed (i.e. with the skew as an additional free parameter), so we elect to use a standard Gaussian for simplicity.

A crucial element of the model design is how to define the parametric mean of the SF-dominated component, μ_{SF}, as this ultimately will become our formalized SFR–radio relationship. In its most basic form, this depends only on $\log_{10}(L_{1.4})$ simply given by

$$\mu_{\mathrm{SF}} = \log_{10}(\mathrm{SFR}) = \eta_L L' + \zeta, \tag{3}$$

where sampling will optimize for the slope η_L and intercept ζ. For clarity, all freely varying coefficients are labelled $\eta_{\mathrm{parameter}}$ and exponents $\gamma_{\mathrm{parameter}}$. There is a range of literature that suggest that the SFR–radio relationship should include a stellar mass or redshift dependence, or both (e.g. J. Delhaize et al. 2017; G. Gürkan et al. 2018; I. Delvecchio et al. 2021; D. C. Molnár et al. 2021; D. J. B. Smith et al. 2021). These two properties are inherently observationally linked for flux limited samples, but may also display their own bona fide evolution. Therefore, to test if either of these properties are required to evolve to explain the connection between SFR and radio emission in our sample, we optionally introduce a stellar mass evolution term, $\eta_m M'$, and/or a redshift evolution term, $\eta_z(1+z)^{\gamma_z}$, into the SF-dominated component's mean relation. A constant intercept, ζ, is not necessary when both stellar mass and redshift terms are used additively, as this introduces unreconcilable degeneracies in the fitting. A model with both a stellar mass and redshift dependence would therefore be given by

$$\log_{10}(\mathrm{SFR}) = \eta_L L' + \eta_z(1+z)^{\gamma_z} + \eta_m M'. \tag{4}$$

The intrinsic scatter of the SF-dominated component is treated as a free parameter, σ_{int}, and is folded into the calculation of the scatter of the SF-dominated Gaussian component,

$$\sigma^2_{\mathrm{SF}} = \sigma^2_{\mathrm{int}} + \eta^2_L \sigma^2_L, \tag{5}$$

where σ_L is the measurement uncertainty in $\log_{10}(L_{1.4})$, taken directly from the MIGHTEE DR1 catalogue flux errors.

The background component is modelled as a one-dimensional Gaussian in $\log_{10}(L_{1.4})$ with mean μ_{bkg} and scatter σ_{bkg}, which are both free parameters. The final element of model design is then the prescription of the fraction of sources in the SF-dominated component compared to the background, f. The sample used in this work spans wide redshift, stellar mass, and radio luminosity ranges, and therefore it is expected that the population of star forming and active galaxies will not be in constant proportion across the entire parameter space. Failing to account for this

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

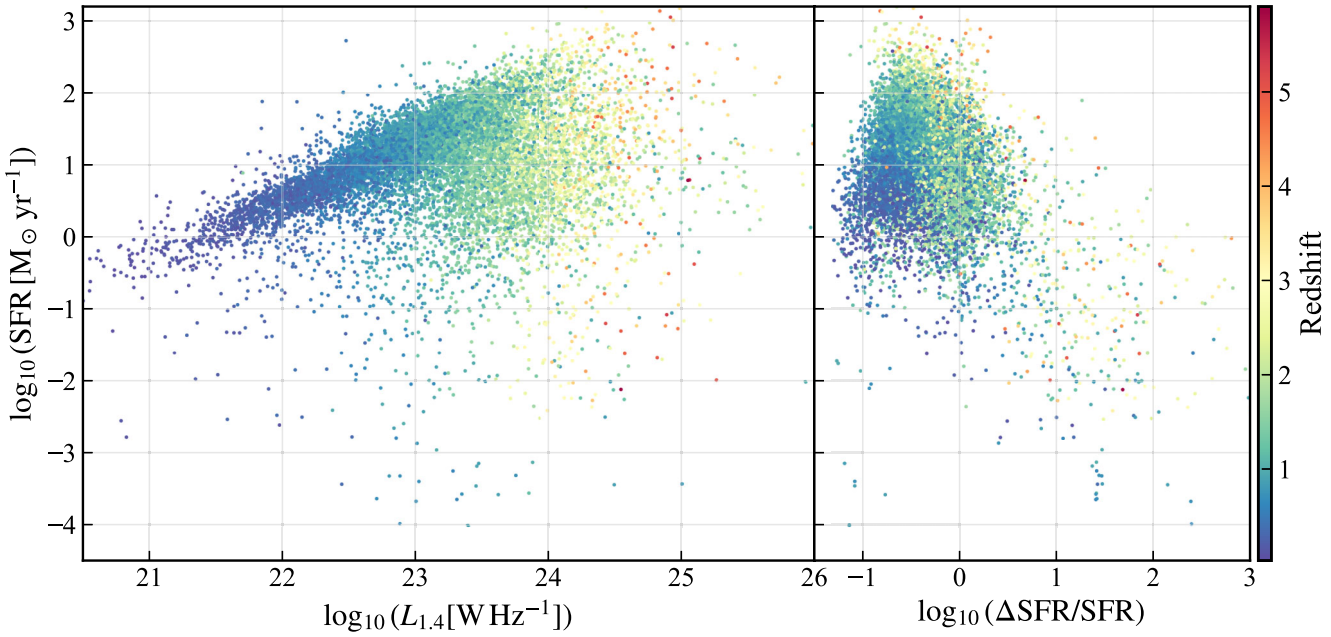


Figure 2. (Left panel) Distribution of intrinsic 1.4 GHz radio luminosities from MIGHTEE DR1 and SFR from GRAHSP SED fitting for all sources meeting basic quality cuts described in Section 3.1, coloured by redshift. (Right panel) The SFR shown as a function of the log of the size of the SFR posterior, as defined by the difference between the 84th and 16th percentile values, scaled by the median. Most sources are well-constrained, with 87.6 per cent having $\log_{10}\left(\frac{\Delta\mathrm{SFR}}{\mathrm{SFR}}\right) < 0.1$, and 95.7 per cent of sources have $\log_{10}\left(\frac{\Delta\mathrm{SFR}}{\mathrm{SFR}}\right) < 0.5$.

would have serious consequences for the robustness of any inferred trends in these properties, as a result of untreated selection effects impacting both populations differently. Therefore, we choose to implement a novel evolving background fraction, and test whether the evolution is best captured by a radio luminosity, redshift, or stellar mass dependence (or some combination of all three). The fraction f is thus parametrized as a logistic function with the form

$$f = \frac{1}{1 + \exp\left[-(\phi + \sum_x \psi_X X)\right]}, \tag{6}$$

where ϕ is a constant, X is some set of radio luminosity, redshift, and/or stellar mass, and ψ_X are the associated coefficients of those variables.

We test an exhaustive suite of models, comprised of all possible combinations of mean SF-dominated and background fraction parametrizations, summarized in Table 1. We adopt a 'flat-top Gaussian' prior distribution for all free parameters, which is defined as a uniform region centred on 0 with a half-width of 2, outside of which there are Gaussian tails of width 0.1, in normalized units. We choose these priors to combine the non-informativeness of standard uniform priors with improved numerical stability when it comes to sampling, due to the removal of discontinuities in probability space.

Table 1. Summary of the models considered in this work. For brevity, the functional form of each model is presented with its dependences in square brackets, i.e. an additive μ_{SF} model that depends on centred stellar mass and redshift is indicated as $\mu_{\mathrm{SF}}[M' + z]$ rather than the full written form, $\log_{10}(\mathrm{SFR}) = \eta_L L' + \eta_z(1+z)^{\gamma_z} + \eta_m M'$. Multiplicative models refer to the form e.g. $\eta \times L' \times M'^{,\gamma_m} \times z^{\gamma_z}$. All μ_{SF} models must depend on $L_{1.4}$, so this is omitted for brevity.

SF-dominated mean description, μ_{SF}	
Functional form	Description
μ_{SF}[base]	$L_{1.4}$-only baseline
$\mu_{\mathrm{SF}}[M']$	Includes M_* term
$\mu_{\mathrm{SF}}[z]$	Includes z term
$\mu_{\mathrm{SF}}[M' + z]$	Full additive model
$\mu_{\mathrm{SF}}[\mathrm{mult} : z]$	Multiplicative model including z
$\mu_{\mathrm{SF}}[\mathrm{mult} : M']$	Multiplicative model including M_*
$\mu_{\mathrm{SF}}[\mathrm{mult} : M' \times z]$	Full multiplicative model
Population fraction description denominator, $1/f$	
Functional form	Description
f[base]	Constant fraction
$f[L']$	$L_{1.4}$ dependence
$f[z]$	z dependence
$f[M']$	M_* dependence
$f[L' + z]$	$L_{1.4}$ and z dependence
$f[M' + z]$	M_* and z dependence
$f[L' + M']$	$L_{1.4}$ and M_* dependence
$f[L' + z + M']$	$L_{1.4}$, M_*, and z dependence

3.3 Likelihood calculations

Each source i has an associated set of SFR values, $\mathrm{SFR}_{i,r}$, sampled from its SED-derived posterior distribution, where r runs from 0 to $R = 40$. The probability of $\mathrm{SFR}_{i,r}$ being drawn from each model component (i.e. the SF-dominated or background) is given by a Gaussian likelihood,

$$\log p_{\mathrm{comp}}(\mathrm{SFR}_{i,r}|\theta) = -\log\sigma_{\mathrm{comp},i} - \frac{\log(2\pi)}{2} - \frac{(\mathrm{SFR}_{i,r} - \mu_{\mathrm{comp},i})^2}{2\sigma^2_{\mathrm{comp},i}}, \tag{7}$$

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

where θ denotes the full free parameter set, and $\mu_{\mathrm{comp},i}$ and $\sigma_{\mathrm{comp},i}$ are the corresponding mean and standard deviation of each component (e.g. SF-dominated or background), as given in Section 3.2.

The per-source-per-sample probability is then marginalized over $\mathrm{SFR}_{i,r}$ to get a per-source probability of being associated with each component,

$$\log p_{\mathrm{comp}}(\mathrm{SFR}_i|\theta) = \log\left[\sum_{r=1}^{R} w_{i,r} \times p_{\mathrm{comp}}(\mathrm{SFR}_{i,r}|\theta)\right]. \quad (8)$$

The total per-source likelihood is thus computed as a sum of both components, each weighted by the fractional occupation of each component as described by the fraction f:

$$p_{\mathrm{mixture}}(\mathrm{SFR}_i|\theta) = f_i p_{\mathrm{SF}}(\mathrm{SFR}_i|\theta) + (1-f_i)p_{\mathrm{bkg}}(\mathrm{SFR}_i|\theta). \quad (9)$$

The overall log likelihood of a model is then given by the sum over all sources:

$$\log \mathcal{L}(\theta) = \sum_{i=1}^{N} \log p_{\mathrm{mixture}}(\mathrm{SFR}_i|\theta), \quad (10)$$

where N is the number of sources.

Depending on the exact specification of μ_{SF} and f used in a given model, θ can contain up to 13 free parameters. To efficiently explore our large suite of possible models, we first perform an initial model screening using a maximum a posteriori (MAP) estimation from the `PyMC find_MAP` function, which acts to maximize the value of $\log \mathcal{L}(\theta) + \log p(\theta)$, where $p(\theta)$ represents the prior. The function uses the L-BFGS-B algorithm (C. Zhu et al. 1997), and multiple random initializations per model in order to mitigate convergence to local maxima. The resulting maximum log-likelihood, $\log \hat{\mathcal{L}}(\theta)$, is used to estimate the Akaike Information Criterion (AIC; H. Akaike 1987), using the point estimates. All possible model configurations are ranked by their AIC score, and the five best (i.e. lowest) scoring model are shortlisted for full posterior sampling.

For those shortlisted models, we perform Bayesian inference using a version of a Hamiltonian Monte Carlo algorithm called the 'No U-Turns Sampler' (NUTS; M. D. Hoffman et al. 2014) from the PYTHON package PYMC (O. Abril-Pla et al. 2023). Each run draws 2000 samples, uses 2000 tuning iterations, samples with 4 chains and has a target acceptance of 0.975. It also allows the posterior probability of a source belonging to each component, $p(\mathrm{comp}|\mathrm{SFR}_i)$, to be calculated. For the primary SF-dominated component, $p(\mathrm{SF\text{-}dominated}|\mathrm{SFR}_i)$ [referred to herein as $p(\mathrm{SF})$ for clarity] corresponds to the probability of a source's radio luminosity being entirely attributable to star formation given its SFR as informed by our mean relation, and is defined as

$$p(\mathrm{SF}) = \frac{f_i p_{\mathrm{SF}}(\mathrm{SFR}_i)}{f_i p_{\mathrm{SF}}(\mathrm{SFR}_i) + (1-f_i)p_{\mathrm{bkg}}(\mathrm{SFR}_i)}. \quad (11)$$

4 RESULTS AND DISCUSSION

In this section, we present the results of both the model selection procedure, and the sampling performed with the optimal model configuration. We further discuss population evolution as described by our evolving background model, investigate the potential impact of an evolution of spectral index with redshift, and test for possible selection effects using a volume-limited subsample of sources.

4.1 Model selection

The models that produce the five best AIC scores from the initial MAP screening procedure and thus undergo full posterior sampling are detailed in Table 2. These highest ranking models have a great deal in common: four out of the top five contain a redshift evolution term in the μ_{SF} equation, and all five have both a radio luminosity and stellar mass dependence in the fraction of SFG–AGN. Differences in the top scoring models arise from whether a stellar mass term is needed for the SFR-radio relation, and whether a redshift term is preferred in the population fraction, but it is decisive that the data are best described by models with a redshift-dependent SFR–radio relationship and an SFG–AGN fraction that depends on stellar mass and radio luminosity. In order to robustly compare the possible model configurations, we report the expected log-predictive density (ELPD) calculated from the widely applicable information criterion (WAIC; S. Watanabe & M. Opper 2010) and the Pareto-smoothed importance sampling leave-one-out cross validation (PSIS-LOO-CV; A. Vehtari, A. Gelman & J. Gabry 2017; A. Vehtari et al. 2024). The WAIC is similar to the AIC, but generalized to include the full posterior distribution from a Bayesian model, penalizing models with greater complexity. PSIS-LOO-CV is an efficient implementation of the more general LOO-CV, which quantifies how well a model performs on unseen data. We implement both of these with `ArviZ` (R. Kumar et al. 2019).

Both the WAIC and PSIS-LOO-CV produce the same rankings, given in Table 2, preferring overall the model with both a redshift and stellar mass dependence, with the functional form

$$\log_{10}(\mathrm{SFR}) = \eta_L L' + \eta_z (1+z)^{\gamma_z} + \eta_m M', \quad (12)$$

and an evolving population fraction that goes as

$$f = \frac{1}{1+\exp\left[-(\phi + \psi_L L' + \psi_m M')\right]}. \quad (13)$$

However, there is only an $\sim 1\sigma$ preference between the highest and second highest scoring models, meaning they cannot be strongly statistically distinguished. These two models have identical μ_{SF} forms given by equation (14), with the only difference between them being simply the addition of an explicit redshift term to the population fraction, f. Given that the addition of an independent redshift evolution in the background sources does not significantly improve predictive performance, it suggests that any apparent redshift evolution in the fraction of sources that have their radio emission dominated by SF-related processes is instead induced by the evolution of $L_{1.4}$ and M_*, insofar as can be constrained by the data. Beyond these two best scoring models, the WAIC or PSIS-LOO-CV ELPD becomes dramatically worse. Both models without an M_* term in μ_{SF} also produce highly degenerate, bimodal posterior distributions. Therefore, due to the preference of the WAIC and PSIS-LOO-CV for the $\mu_{\mathrm{SF}}[M'+z]$, $f[L'+M']$ formulation, we choose to retain this as our primary model.

Explicitly, the data are best described by an SFR–radio relation that evolves as a function of both stellar mass and redshift, and the fraction of SF-dominated sources is dictated by a combination of stellar mass and radio luminosity.

4.2 The SFR–radio relation

Our optimal model configuration is then sampled using `PyMC.sample`, yielding the identification of two components

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

Table 2. Comparison of candidate models using the ELPD as calculated by the WAIC and PSIS-LOO-CV. Differences are reported relative to the highest-ranked model. Uncertainties on the ELPDs correspond to the estimated standard errors, whilst uncertainties on the differences correspond to the standard error of the pairwise difference. Model notation is as defined in Table 1. The best performing model has a redshift evolution and weak mass evolution, with a mass and luminosity dependent background fraction.

Rank	Model	k	$\mathrm{ELPD_{WAIC}}$	$\Delta\mathrm{ELPD_{WAIC}}$	$\mathrm{ELPD_{LOO}}$	$\Delta\mathrm{ELPD_{LOO}}$
1	$\mu_{\rm SF}[M'+z];\, f[L'+M']$	10	-9399.5 ± 141.9	0.0	-9399.5 ± 141.9	0.0
2	$\mu_{\rm SF}[M'+z];\, f[L'+z+M']$	11	-9400.4 ± 141.9	0.90 ± 0.77	-9400.4 ± 141.9	0.89 ± 0.77
3	$\mu_{\rm SF}[z];\, f[L'+M']$	10	-9404.5 ± 141.8	5.02 ± 4.83	-9404.5 ± 141.8	11.78 ± 4.83
4	$\mu_{\rm SF}[M'];\, f[L'+M']$	11	-9417.9 ± 141.1	18.39 ± 7.62	-9417.86 ± 141.1	18.38 ± 7.63
5	$\mu_{\rm SF}[z];\, f[L'+z+M']$	11	-11404.6 ± 142.0	2005.10 ± 39.64	-11254.2 ± 134.1	1854.70 ± 35.16

Table 3. Posterior constraints on the best-fitting model, $\mu_{\rm SF}[M'+z];\, f\ [L'+M']$. Values are medians, and the uncertainties stated are 68 per cent credible intervals.

$\eta_{\rm l}$	$\eta_{\rm m}$	η_z	γ_z	ϕ	$\psi_{\rm l}$	$\psi_{\rm m}$	$\sigma_{\rm int}$	$\mu_{\rm bkg}$	$\sigma_{\rm bkg}$
$0.790^{+0.010}_{-0.010}$	$-0.033^{+0.009}_{-0.010}$	$1.244^{+0.016}_{-0.016}$	$0.122^{+0.019}_{-0.019}$	$-0.006^{+0.046}_{-0.046}$	$-1.971^{+0.055}_{-0.053}$	$1.288^{+0.054}_{-0.052}$	$0.178^{+0.004}_{-0.004}$	$0.769^{+0.013}_{-0.013}$	$0.762^{+0.009}_{-0.008}$

Table 4. Posterior constraints for the second and third best performing model configurations. Values are medians with 68 per cent credible intervals. Some variables are not present in both models, so are only quoted when relevant. It is important to note that model $\mu_{\rm SF}[z];\, f[L'+M']$ shows bimodal, highly degenerate posterior distributions, so the quoted medians and uncertainties should be treated with caution.

Parameter	$\mu_{\rm SF}[M'+z];\, f[L'+z+M']$	$\mu_{\rm SF}[z];\, f[L'+M']$
$\eta_{\rm l}$	$0.789^{+0.010}_{-0.010}$	$0.776^{+0.011}_{-0.012}$
$\eta_{\rm m}$	$-0.032^{+0.009}_{-0.009}$	–
η_z	$1.241^{+0.017}_{-0.017}$	$0.974^{+0.765}_{-1.311}$
γ_z	$0.126^{+0.020}_{-0.020}$	$0.113^{+0.100}_{-1.092}$
ζ	–	$-0.094^{+1.574}_{-0.627}$
ϕ	$0.045^{+0.096}_{-0.096}$	$0.320^{+0.052}_{-0.052}$
$\psi_{\rm l}$	$-2.022^{+0.082}_{-0.074}$	$-2.077^{+0.055}_{-0.055}$
$\psi_{\rm m}$	$1.277^{+0.057}_{-0.054}$	$1.307^{+0.054}_{-0.053}$
ψ_z	$-0.051^{+0.080}_{-0.083}$	–
$\sigma_{\rm int}$	$0.178^{+0.004}_{-0.003}$	$0.180^{+0.003}_{-0.004}$
$\mu_{\rm bkg}$	$0.770^{+0.013}_{-0.014}$	$0.837^{+0.014}_{-0.013}$
$\sigma_{\rm bkg}$	$0.762^{+0.011}_{-0.008}$	$0.716^{+0.009}_{-0.008}$

as shown in Fig. 3. As expected, the model recovers a tight SF-dominated component, alongside a scattered background. The sampling returns posteriors for the parameters that can be used to derive a mean relation for the SFR–radio correlation, producing the following SFR–radio correlation:

$$\mu_{\rm SF} = \log_{10}(\mathrm{SFR}) = 0.790L' + 1.244(1+z)^{0.122} - 0.033M' \quad (14)$$

and shown in Fig. 3 as the solid black line. The optimized parameters for equation (14) are presented in Table 3 , and the posterior distributions for those parameters are shown in Fig. 4. Most parameters are fit with little degeneracy, except between η_L, η_z, and γ_z. Notably, the sampling does not find any significant degeneracy between the parameters describing stellar mass and redshift dependence.

Parameter uncertainties are given in Table 3 . A key conclusion from our result is that a redshift evolution in the SFR–radio relationship is strongly preferred by the data. All of the highest scoring model configurations require a redshift term in the $\mu_{\rm SF}$ relation, and in our best fitting solution the exponent on the $(1+z)$ term is 6.4σ above zero. In the left panel of Fig. 3, we demonstrate the effect of varying redshift, showing that at higher redshifts, the same SFR is associated with lower radio emission. In our suite of tested models, we included a multiplicative form that would have allowed redshift to vary the slope of the relation rather than the intercept, but this was strongly disfavoured. On the other hand, the mass dependence found in Table 3 is only 3.7σ above zero and as can be seen in Fig. 3, even an increase of multiple orders of magnitude in stellar mass yields virtually indistinguishable results. Indeed, a change in M_* of four orders of magnitude produces only a 0.1 dex difference in $\log_{10}$(SFR) in our best-fitting model, much less than even the intrinsic scatter in the relationship. Therefore, whilst the formal version of our derived SFR–radio includes a mass evolution term, in practise it could be safely omitted from calculation at little cost. For completeness, we also give the optimal parameter values for our second and third best performing models in Table 4. We are confident that the observed evolutionary trend with redshift is not induced through an increasing mass incompleteness at higher redshifts, as our evolving background fraction acts to absorb any changing completeness in radio luminosity or stellar mass, as discussed in Section 4.3. Furthermore, even when the fitting is performed without a redshift term (i.e. as per $\mu_{\rm SF}[M'];\, f[L'+M']$, our fourth best ranking model in Table 2), a very mild mass dependence of $\eta_{\rm m} = -0.038^{+0.009}_{-0.010}$ is found, and instead $\eta_{\rm l}$ steepens to a value of $0.843^{+0.006}_{-0.007}$. Therefore, using our entirely data-driven methodology, we can conclude that the SFR–radio correlation shows very limited evolution with stellar mass. Interestingly, if we apply a strict χ^2 cut to our SED fitting results, the mass evolution term becomes weaker still, perhaps suggesting that the derived mass dependence of previous studies may be influenced by poorly constrained physical properties of sources, particularly if AGNs have not been properly accounted for.

4.3 Population fraction evolution

Our NUTS sampling also reveals an optimal population fraction evolution described by

$$f = \frac{1}{1+\exp\left[-(0.006 - 1.971 \times L' + 1.288 \times M')\right]}, \quad (15)$$

where all parameters and their uncertainties are given in Table 5. The physical implication of this is that the fraction of SF-

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

Table 5. Posterior constraints on the model $\mu_{\rm SF}[M' + z]$; f $[L' + M']$ with an evolving spectral index, $\alpha = \alpha_0 - \eta_\alpha z$. Values are medians, and the uncertainties stated are 68 per cent credible intervals.

η_α	$\eta_{\rm l}$	$\eta_{\rm m}$	ζ	ϕ	$\psi_{\rm l}$	$\psi_{\rm m}$	$\sigma_{\rm int}$	$\mu_{\rm bkg}$	$\sigma_{\rm bkg}$
$-0.068^{+0.025}_{-0.024}$	$0.830^{+0.008}_{-0.008}$	$-0.037^{+0.009}_{-0.009}$	$1.329^{+0.009}_{-0.009}$	$0.059^{+0.047}_{-0.047}$	$-1.929^{+0.058}_{-0.057}$	$1.274^{+0.054}_{-0.051}$	$0.180^{+0.004}_{-0.004}$	$0.764^{+0.009}_{-0.009}$	$0.768^{+0.013}_{-0.013}$

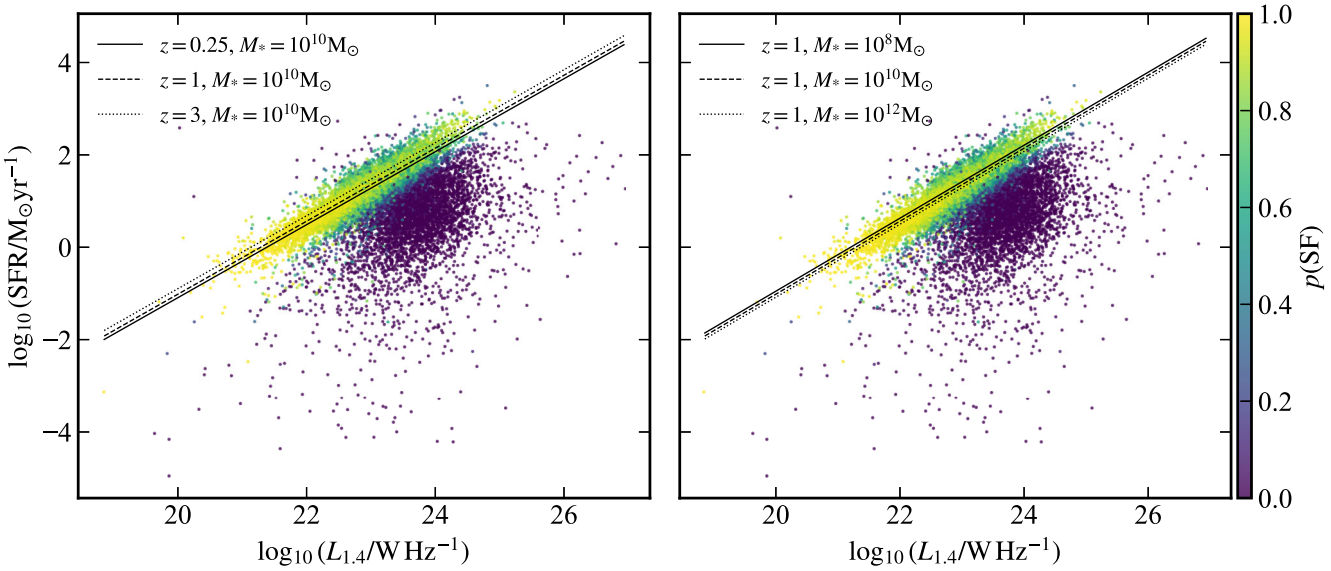


Figure 3. The distribution of sources in $\log_{10}(L_{1.4})$–$\log_{10}(\mathrm{SFR})$ space, as determined by the overall best performing model. Points are coloured by $p(\mathrm{SF})$, which is the probability of a source belonging to the SF-dominated component, as given by equation (11). The various lines correspond to the best-fitting $\mu_{\rm SF}$, $\mu_{\rm SF} = \log_{10}(\mathrm{SFR}) = 0.790L' + 1.244(1+z)^{0.122} - 0.033M'$, with differing redshift and stellar mass values. The left panel shows the impact of varying redshift between $z = 0.25$ and $z = 3$ for a fixed $M_* = 10^{10}$ $\mathrm{M}_\odot$, which results in an ∼0.2 dex difference in $\log_{10}(\mathrm{SFR})$ for a given radio luminosity. The right panel shows the impact of varying M_* over four orders of magnitude at a fixed $z = 1$. For a fixed radio luminosity, the difference in $\log_{10}(\mathrm{SFR})$ between $M_* = 10^8$ and 10^{12} $\mathrm{M}_\odot$ is ∼ 0.13 dex.

dominated sources decreases with increasing radio luminosity, yet increases with increasing stellar mass. The trend with radio luminosity corresponds to the highest radio luminosity sources being associated with powerful AGN activity, much more luminous than what can be attributed to star formation alone. The increasing fraction of SF-dominated sources at higher M_* can be understood as our evolving fraction capturing the mass bias imposed by a flux limited sample. That is, our evolution of f should not be strictly interpreted as though highest mass galaxies are physically preferentially SF-dominated, but instead that only those SF-dominated sources with relatively high M_* can produce enough radio emission to meet the detection threshold. That being said, there is also the physical effect of the main sequence of star formation, where more massive galaxies have higher SFRs compared to lower mass galaxies (K. G. Noeske et al. 2007; K. E. Whitaker et al. 2014; R. Johnston et al. 2015). Clearly, the trends in f should be carefully interpreted as a convolution of genuine underlying population trends and selection effects that the evolving background formulation acts to capture and prevent from impacting the overall SFR–radio correlation. It should equally be noted that the absolute values derived for our population fraction evolution (equation 15) are highly affected by the survey flux limit, and will certainly vary if the analysis is repeated on a survey of a differing depth, by design.

4.4 Redshift evolution: exploring the possible impact of spectral index evolution

There has been much debate in the literature about the existence of a redshift dependence in the SFR–radio relation, though most work has focussed on potential evolution in the IRRC, which is assumed to propagate through from the underlying SFR–radio relation – though see D. C. Molnár et al. (2021) for evidence that non-linearities in the IRRC drive observed dependences. B. Magnelli et al. (2015), J. Delhaize et al. (2017), and E. F. Ocran et al. (2020) all find a mild, negative evolution in the (F)IRRC, whilst H. S. B. Algera et al. (2020), G. De Zotti et al. (2024), and F. Tabatabaei et al. (2025) find no redshift evolution in the range $1.5 < z < 3.5$, for example. In this work, we account for mass and redshift dependant selection effects through the use of our evolving SFG–AGN fraction, and recover a mild but statistically non-negligible evolution in the SFR–radio relation with redshift, though note that the number density of our sample is significantly reduced at $z > 4$ compared to $0.5 < z < 2$, and therefore the conclusions we are able to draw are most robust at low to intermediate redshifts (e.g. $z \lesssim 4$).

There are multiple potential explanations for this evolution. Amongst those important to consider is the impact of inverse Compton (IC) scattering and synchrotron losses on the radio spectral index. Our sample spans a large redshift range and whilst IC losses from interactions with the cosmic microwave back-

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

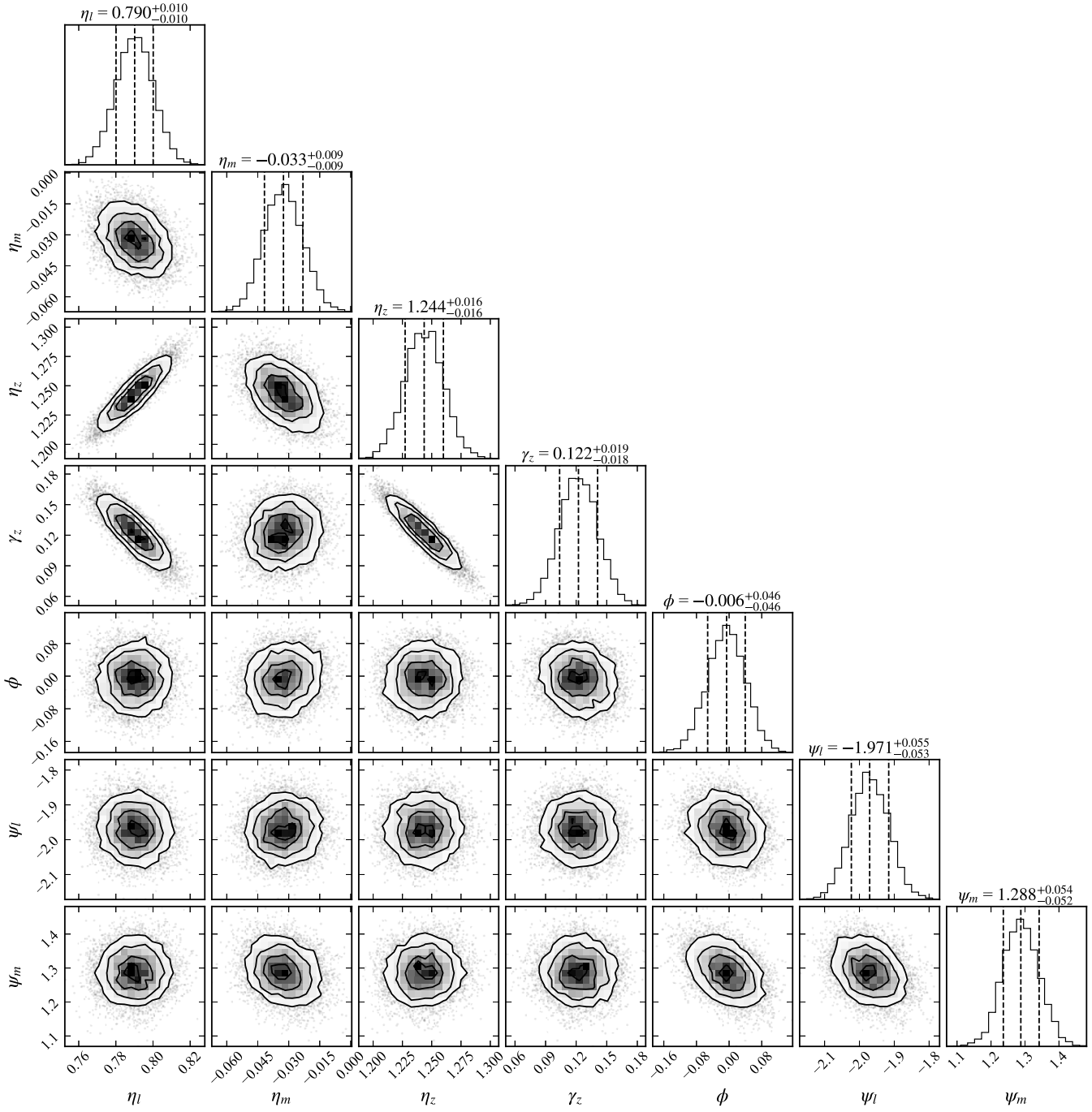


Figure 4. The posterior probability distributions of the free parameters in the optimal SFR–radio model, $\mu_{\rm SF}[M' + z]$; f $[L' + M']$, as described by equations (14) and (15), ascertained by NUTS sampling. The median, 16 per cent and 84 per cent confidence intervals are stated.

ground (CMB) are minimal locally, they scale with the energy density of the CMB, which goes as $(1+z)^4$. E. J. Murphy (2009) demonstrates, for example, that this should become significant by $z \approx 3$, and I. H. Whittam et al. (2025) finds compelling evidence for a redshift evolution in 1.4 GHz flux density and luminosity at $3 < z < 5$ that is consistent with IC scattering, in line with previous findings by G. De Zotti et al. (2024). In our primary analysis, we choose to take the standard literature approach of assuming the radio emission we observe in our sample is dominated at 1.4 GHz by synchrotron emission, and can be treated with a constant $\alpha = -0.7$. However, any increase in electron cooling effects with redshift could quite reasonably lead to a radio spectral steepening, meaning our flux to luminosity conversion is underpredicting radio luminosities in sources at higher redshift. Therefore, a moderate positive redshift evolution would be needed to account for this in the data: much like we recover in our model. Motivated by this, we perform a simple test to ascertain the strength of spectral index evolution required to explain the derived redshift term in our SFR–radio relationship.

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

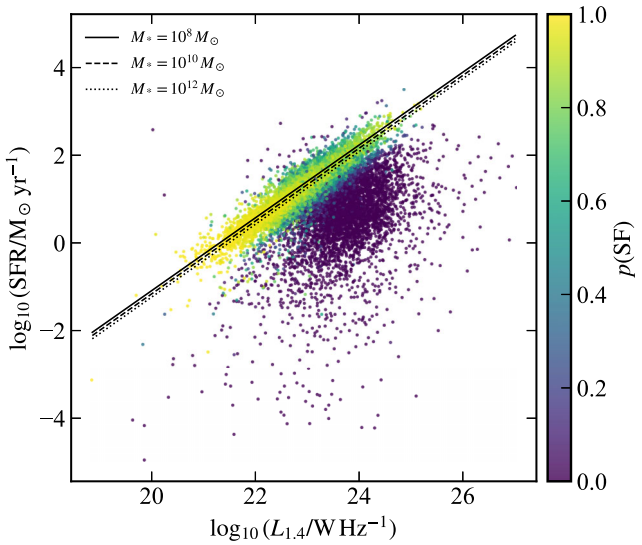


Figure 5. The distribution of sources in $\log_{10}(L_{1.4})$–$\log_{10}(\mathrm{SFR})$ space, with an evolving α factor included in the calculation of $\log_{10}(L_{1.4})$, of the form $\alpha = -0.7 - 0.068z$. Points are coloured by $p(\mathrm{SF})$, which is the probability of a source belonging to the SF-dominated component, as given by equation (11). The lines correspond to the SFR–radio relation, $\log_{10}(\mathrm{SFR}) = 0.830L'[\alpha(z)] - 0.037M' + 1.329$, evaluated at different stellar masses. The four orders of magnitude change in stellar mass corresponds to a change of 0.12 dex in SFR.

We move the calculation of radio luminosity to within the sampling procedure, parametrizing the spectral index as $\alpha(z) = \alpha_0 + \eta_\alpha z$, where $\alpha_0 = -0.7$. The non-evolving case can therefore be recovered if the best-fitting parameter of η_α is zero. The flux to luminosity conversion is therefore the same as equation (2), but with $\alpha \equiv \alpha(z)$. In our fitting of the SFR–radio relation, we use $\log_{10}(L_{1.4})$. Taking the base-10 logarithm of equation (1), it can shown that the evolving spectral index case described above is mathematically equivalent to a linear additive redshift term, i.e.

$$\begin{aligned}\log_{10}(L_{1.4})[\alpha(z)] &= \log_{10}(4\pi d_L^2 S_{1.4}) - (1 + \alpha_0 + \eta_\alpha z)\log_{10}(1+z) \\ &= \log_{10}(L_{1.4})[\alpha_0] - \eta_\alpha z \log_{10}(1+z). \end{aligned} \quad (16)$$

The sole difference between the implementation of an evolving spectral term within the radio luminosity calculation and a simple redshift term of the form $\eta_\alpha z \log_{10}(1+z)$ in μ_{SF} is whether the new calculated luminosity is included in the background fraction or not. We rerun the fitting procedure using both formulations, both of which require a constant intercept for normalization purposes. The value of η_α in both cases is found to be $-0.068^{+0.025}_{-0.024}$. The other model parameters change slightly in comparison to the best-fitting model from the primary analysis, and the optimized μ_{SF} is found to be

$$\begin{aligned}\log_{10}(\mathrm{SFR}) &= 0.830L'[\alpha(z)] - 0.037M' + 1.329 \\ \alpha[z] &= -0.7 - 0.068z \end{aligned} \quad (17)$$

or equivalently,

$$\log_{10}(\mathrm{SFR}) = 0.830L'[\alpha_0] - 0.037M' + 0.056z\log_{10}(1+z) + 1.329. \quad (18)$$

These results are presented in Fig. 5, with the values and uncertainties presented in Table 5 and the posterior distributions shown in Fig. 6. The WAIC and PSIS-LOO-CV would place these evolving spectral index models as the fourth and 5th best overall, with a best case difference of $\Delta\mathrm{ELPD}_{\mathrm{WAIC}} = 15.88 \pm 5.11$ and $\Delta\mathrm{ELPD}_{\mathrm{LOO}} = 15.90 \pm 5.11$, compared to the highest ranking model. We do not set a_0 as a free parameter in our test, as our current single frequency data is insufficient to constrain it, and note that setting it to a different (reasonable) value causes a small change to the inferred evolution, as might be expected. For example, the case of $\alpha_0 = -0.75$ gives the evolution $\alpha[z] = -0.7 - 0.057^{+0.022}_{-0.023}z$, which is weaker but consistent within errors to the $\alpha_0 = -0.7$ case.

Recent work by V. Heesen et al. (2026) defined a unified 'fundamental plane' of the SFR–radio relation for both individual galaxies and kiloparsec-scale spatially resolved regions by introducing an additional linear spectral index term to the form of equation (3). They attribute the dependence to a steepening of measured spectral index in sources that undergo more significant synchrotron losses in their CRs, alongside reduced radio luminosity in sources with more efficiently escaping CRs. In our case, we choose not to speculate as to whether or not a spectral index evolution is the genuine cause of the observed redshift dependence, but note that our value of $\eta_\alpha = -0.068$ corresponds to a spectral index of $\alpha(z=0) = -0.7$, $\alpha(z=1) = -0.768$ and $\alpha(z=2) = -0.836$, values which would not be especially extreme. An alternative explanation beyond a physical spectral index evolution would be if the population of radio detected galaxies had a curved spectrum, rather than one that could be described by a single spectral index, or even just an average spectral index greater than the assumed $\alpha = -0.7$. At higher redshift, the observed 1.4 GHz radiation has its origins at increasingly high frequencies, scaling as $\propto 1.4(1+z)$. Therefore, using a single $\alpha = -0.7$ value would increasingly underestimate the rest frame luminosity as redshift increased – inducing an observed redshift evolution in the sample. Furthermore, even if a single α value were to adequately describe the average radio spectrum of our sources, as a higher rest-frame frequency is probed, the fundamental relationship between SFR and $L_{\nu,\mathrm{emit}}$ may not be constant. An example of this in the local Universe is shown in F. S. Tabatabaei et al. (2017). Equally, it is possible that there is a breakdown of the assumptions underpinning the SFR–radio relation at higher redshifts – for example, the reduction in radio luminosity with increasing redshift for a fixed SFR and M_* could point towards a potential disconnect in the calorimetry models that link the two properties. This could occur due to galaxies generally being more compact at earlier epochs, resulting in a greater fraction of CR electrons being able to escape from the source without interacting with magnetic field lines, thus producing weaker observable synchrotron radiation.

Fundamentally, it is difficult to disentangle which of these possible causes of an observed redshift evolution is ultimately responsible for the relationship we find; however, there is clearly good reason to expect that the SFR–radio correlation should vary with redshift in one manner or another. Whether the apparent redshift evolution of our SFR–radio relationship is a result of an evolving spectral index; an innate link between spectral index, CR escape efficiency and radio luminosity; curved radio spectra, or perhaps some alternative cause will rapidly enter the realm of testability with the MIGHTEE *S*-band data (Hale et al. in preparation) and (Thykkathu et al. in preparation) over the MIGHTEE-DR1 fields.

4.5 Testing for dependence on selection effects

The main selection effect that may impact our results if improperly treated is the Malmquist bias – for example, is our measured correlation between $L_{1.4}$ and SFR simply a result of comparing

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

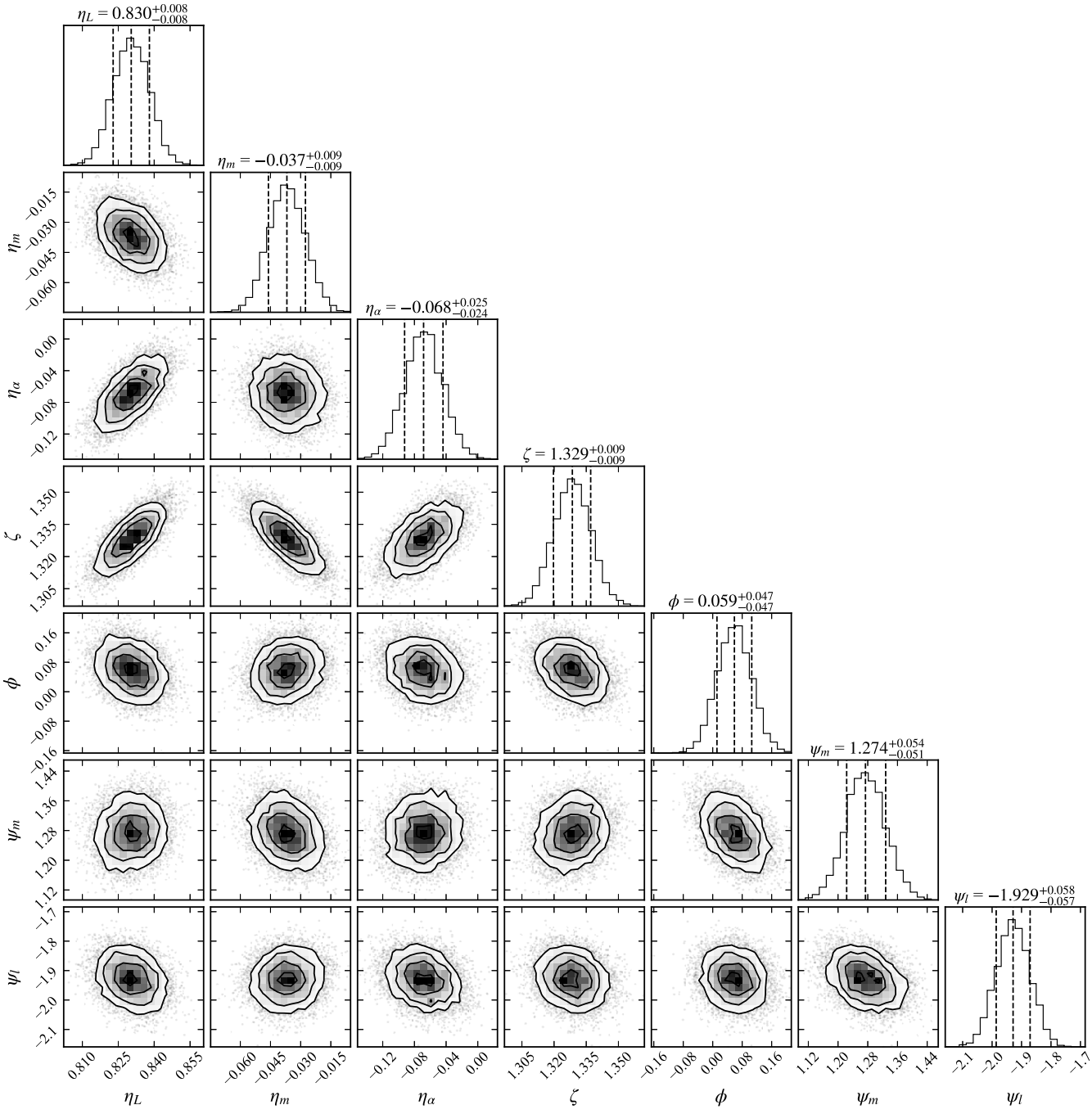


Figure 6. The posterior probability distributions of the best-fitting combination of parameters for the evolving spectral index model, as ascertained by the NUTS sampling. The median, 16 per cent and 84 per cent confidence intervals are stated.

luminosities from two flux limited samples? The fact that we fit for redshift, mass, and luminosity dependences in the correlation between SFR and radio luminosity, somewhat mitigates the impact of Malmquist bias. Furthermore, our parent sample of optical and near-infrared galaxies is stellar mass limited, which although correlated with SFR through the star-formation main sequence (K. G. Noeske et al. 2007; K. E. Whitaker et al. 2014; R. Johnston et al. 2015), is not directly impacted by the optical/near-infrared flux limit. We also have a relatively large baseline in both radio luminosity and redshift, allowing any SFR–radio relation to be constrained by the data at low redshift, whereas the redshift evolution constraints are obviously more influenced by the high-luminosity SFGs that we detect across a broad redshift range. Similarly, any dependence of the SFR–$L_{1.4}$ relation on stellar mass is largely constrained by the same lower-redshift populations where we have the largest baseline in stellar mass from $\sim 10^8$ to 10^{11} $M_\odot$ (Fig. 1).

To test the impact of Malmquist bias we limit our sample to a region in the $L_{1.4}$–z plane where we can expect that observations are volume limited, i.e. complete to the lower $L_{1.4}$ limit, such that if the same inferred SFR–radio correlation holds then it can be safely separated from Malmquist bias. We restrict our catalogue

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

to sources in a defined region of $L_{1.4}$–z space, indicated in Fig. 1, and draw a bootstrap sample of 500 sources from this volume-limited catalogue. The choice is made to bootstrap to a relatively small number in order to minimize the impact of sample size on our eventual population statistics, but with many realizations to ensure robustness. For each source, we randomly sample from their SFR posterior distributions from the GRAHSP SED fitting to produce an observed benchmark. A single full posterior sample of the fitted model parameters is randomly selected from the chain, and is then combined with the observed $L_{1.4}$, z, and M_* values of the sources to calculate a probability of belonging to either component, given by f, as in equation (15). A Bernoulli draw based on this probability then dictates whether an individual source will have its SFR calculated using equation (14) (i.e. it belongs to the SFR–radio correlation) or if it is assigned to the Gaussian background with mean $\mu_{\rm bkg}$ and $\sigma_{\rm bkg}$ (values given in Table 5). The simulated and observed catalogues are compared with a Kolmogorov–Smirnov (KS) test from SCIPY (P. Virtanen et al. 2020). We choose a KS test over the commonly used Anderson–Darling (AD) test, as AD tests are more sensitive to differences in the tail of distributions compared to a KS test, and we are primarily concerned with how well the SF-dominated relationship is recovered, rather than the background. This procedure is then repeated 5000 times, across many different posterior realizations, and thus yields a distribution of KS statistics and p-values through which we can ascertain how well the equations (14) and (15) reproduce the observed population within a volume limited sample.

We perform the test above on both the case of $\log_{10}(L_{1.4}) > 23, z < 1$ and $\log_{10}(L_{1.4}) > 22.5, z < 0.75$, which are both indicated in Fig. 1. The results are shown in Fig. 7, and the median KS test statistics are 0.078 for $\log_{10}(L_{1.4}) > 23, z < 1$ and 0.068 for $\log_{10}(L_{1.4}) > 22.5, z < 0.75$, and the median p-values are 0.096 and 0.198, respectively. Therefore, in both cases, the simulated samples can be taken as statistically indistinguishable from the observed samples. Furthermore, as can be seen from the alignment of $\mu_{\rm SF}$ as described in equation (14) and in Fig. 7, our derived relation holds equally well in a complete, volume limited sample, and we can conclude that it is unlikely to be a result of flux-related selection bias.

Another potential issue that could affect our results is an overweighting of high redshift sources, both as a consequence of $(1+z)$ not being proportional to comoving volume, and as the uncertainty on the intrinsic assumed spectral index increases. We verify that our results are insensitive to an imposed lower redshift limit ($z < 2$) by rerunning the fit for that subsample, finding parameter values $\eta_L = 0.794 \pm 0.010$, $\eta_m = -0.035 \pm 0.000$, $\eta_z = 1.258 \pm 0.017$, $\gamma_z = 0.103 \pm 0.021$, $\phi = 0.221^{+0.050}_{-0.051}$, $\psi_m = 1.244 \pm 0.061$, $\psi_L = -2.083^{+0.055}_{-0.059}$, and $\sigma_{\rm int} = 0.181^{+0.04}_{-0.03}$. All free parameters are reproduced within the confidence interval of the value derived from the full sample, importantly including those determining the redshift evolution, and the conclusions of this paper are unchanged regardless of which redshift range is included in the analysis.

5 CONCLUSIONS

We have presented a new, improved method for determining the star formation rate–$L_{1.4}$ correlation in star-forming galaxies. The principal strengths of this method include the flexible treatment of contaminating AGNs (rather than needing to rely on binary cuts), a full inclusion of SFR posterior probabilities in the calculation, and a rigorous testing of the optimal functional form for the relationship. By applying this method to a combination of deep MIGHTEE radio flux densities and SFR estimates derived from state-of-the-art SED fitting with the new code GRAHSP, we obtain the relationship

$$\mu_{\rm SF} = \log_{10}({\rm SFR}) = 0.79L' + 1.244(1+z)^{0.122} - 0.033M',$$

where $L' = \log_{10}(L_{1.4}/{\rm W\,Hz^{-1}}) - 23$ and $M' = \log_{10}(M_*/{\rm M_\odot}) - 10$. A key result is that we find a mild but statistically significant positive redshift evolution, such that at a higher redshift there is less radio luminosity associated with the same SFR. We find a minimal stellar mass evolution, much weaker than previous works based on the IRRC.

The redshift evolution that we recover can reasonably be explained by a mild spectral index evolution in redshift of the form $\alpha = -0.7 - 0.068z$, and an associated SFR–radio correlation of $\log_{10}({\rm SFR}) = 0.830L'[\alpha(z)] - 0.037M' + 1.329$. We do not know if this scenario represents physical reality, but it will be directly testable with matched-resolution imaging of the MIGHTEE fields in both L band and more recent S band that will become available

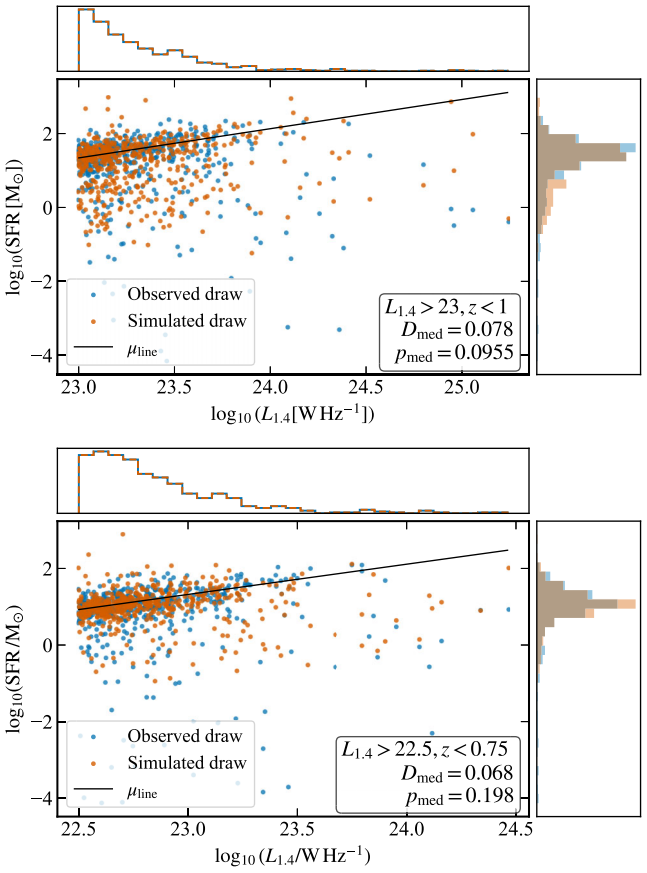


Figure 7. In SFR–$L_{1.4}$ space, a realization of distribution of observed sources (blue) – where their SFR value has been randomly selected from their posterior – and simulated sources (orange), calculated from observed $L_{1.4}$, z, and M_* values combined with the posterior probability distributions of the fitted parameters in our derived SFR–radio correlation. The line on each plot shows the median best-fitting values of the SFR–radio correlation, i.e. equation (14). A KS test has been performed on the SFR distributions and the median KS statistic ($D_{\rm med}$) and p-value ($p_{\rm med}$) are quoted. Both panels represent a different volume limited sample: The top has been restricted to $L_{1.4} > 23, z < 1$ and the bottom $L_{1.4} > 22.5, z < 0.75$.

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

in the near future (Hale et al. in preparation; Thykkathu et al. in preparation). We also show that our derived correlation is unlikely to be a direct result of Malmquist bias.

Previous studies have found a much stronger dependence of the SFR–radio correlation on the stellar mass (e.g. G. Gürkan et al. 2018; I. Delvecchio et al. 2021; D. J. B. Smith et al. 2021) than found in this work. Although it is difficult to assess the root cause of this difference, it is likely due to divergent approaches to the problem, specifically how different studies deal with AGN contamination is likely to be a critical factor. The mass dependences found previously clearly show that the radio luminosity increases with the mass of the galaxy, and the main constraints on this are from galaxies with $M_* > 10^{10}\,\mathrm{M}_\odot$. As highlighted by D. J. B. Smith et al. (2021), it is difficult to rule out AGN contamination in this regime with methods that apply cuts based on distance from an (assumed) underlying correlation prior to measuring the correlation itself. Furthermore, works that obtain IR luminosities from SED fitting have rarely included appropriate AGN components in their modelling, further adding to the issue of AGN contamination in their results. On the other hand, our method is purely driven by the data itself, using a statistical approach that marginalizes over the full per-source SFR posterior distribution derived from GRAHSP SED fitting, where AGN contributions are flexibly accounted for, coupled with a hierarchical Bayesian model that accounts for an evolving background population (AGN). We therefore suggest that the strong mass dependence is at least partially an artefact of the initial sample selection of SFGs (or removal of AGN from the sample). Furthermore, we directly derive the SFR–radio relationship, without using an intermediate scaling relation such as the IRRC. If there is a stellar mass dependence in the way that the IR and radio emission links in relation to SF activity, then using the IRRC could artificially induce a stellar mass evolution in the resulting SFR–radio relationship.

The relation from this work can be widely applied to studies investigating the evolution of the star formation history of the Universe using radio data. However, we caution that applying it naively to the general radio population where traditional cuts have been made to remove AGN will continue to result in uncertain levels of AGN contamination in any derived quantities. As such, we may need to move towards a new framework, such as implemented in this work, to obtain more accurate measurements of the star formation in galaxies traced via radio emission. Furthermore, given that AGN emission can contaminate across the whole of the electromagnetic spectrum, such an approach may also be required for other tracers of star formation.

ACKNOWLEDGEMENTS

CLJ and JHM acknowledge funding from a Royal Society University Research Fellowship (URF\R1\221062). CLJ gratefully acknowledges support from the Balliol College Dervorguilla Scholarship. MJJ, IHW, and CLH acknowledge support from the Hintze Family Charitable Foundation through the Oxford Hintze Centre for Astrophysical Surveys. MJJ, NS, RGV, and SLJ acknowledge funding from a UKRI Frontiers Research Grant (EP/X026639/1), and MJJ acknowledges the support of the Science and Technology Facilities Council (STFC) consolidated grant (ST/W000903/1). JH acknowledges funding from STFC (grant code ST/Y509474/1). CLH acknowledges support from STFC through grant ST/Y000951/1. DJBS acknowledges support from STFC under grant ST/Y001028/1, and from the Leverhulme Trust via Research Project Grant RPG-2025-078. We thank Nick Seymour, Bohan Yue, and Fatemeh Tabatabaei for their helpful comments and suggestions, and the anonymous reviewer for their useful feedback.

The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. We acknowledge the use of the ilifu cloud computing facility—www.ilifu.ac.za, a partnership between the University of Cape Town, the University of the Western Cape, Stellenbosch University, Sol Plaatje University, and the Cape Peninsula University of Technology. The Ilifu facility is supported by contributions from the Inter-University Institute for Data Intensive Astronomy (IDIA – a partnership between the University of Cape Town, the University of Pretoria, and the University of the Western Cape, the Computational Biology division at UCT and the Data Intensive Research Initiative of South Africa (DIRISA). The authors acknowledge the Centre for High Performance Computing (CHPC), South Africa, for providing computational resources to this research project. This work is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 (Ultra-VISTA) and ID 179.A-2006 (VIDEO) and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the Ultra-VISTA and VIDEO consortia. The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University. This work is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT), which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada–France–Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

DATA AVAILABILITY

All observational data used in this paper are publicly available. See C. L. Hale et al. (2025) for the MIGHTEE radio data, and C. L. Hale et al. (2026) for the cross-matched catalogue. The GRAHSP SED fits and derived SFRs used in this paper will be released in a subsequent paper (Jackson et al. in preparation). Additional derived data products are available on reasonable request.

REFERENCES

Abril-Pla O. et al., 2023, PeerJ Comput. Sci., 9, e1516
Aihara H. et al., 2022, PASJ, 74, 247

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026

Akaike H., 1987, Psychometrika, 52, 317
Alarcon A. et al., 2021, MNRAS, 501, 6103
Algera H. S. B. et al., 2020, ApJ, 903, 138
An F. et al., 2021, MNRAS, 507, 2643
An F. et al., 2024, MNRAS, 528, 5346
Appleton P. N. et al., 2004, ApJS, 154, 147
Arnaudova M. I. et al., 2025, MNRAS, 542, 2245
Bait O., Schaerer D., Izotov Y. I., Sebastian B., 2026, MNRAS, 548, stag492
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Bonfield D. G. et al., 2011, MNRAS, 416, 13
Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, A&A, 622, A103
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buchner J. et al., 2024, A&A, 692, A161
Calistro Rivera G., Lusso E., Hennawi J. F., Hogg D. W., 2016, ApJ, 833, 98
Calistro Rivera G. et al., 2017, MNRAS, 469, 3468
Chabrier G., 2003, PASP, 115, 763
Condon J. J., 1992, ARA&A, 30, 575
Cook R. H. W. et al., 2024, MNRAS, 531, 708
Cram L., Hopkins A., Mobasher B., Rowan-Robinson M., 1998, ApJ, 507, 155
Cuillandre J.-C. J. et al., 2012, in Peck A. B., Seaman R. L., Comeron F., eds, Proc. SPIE Conf. Ser. Vol. 8448, Observatory Operations: Strategies, Processes, and Systems IV. SPIE, Bellingham, p. 84480M
Dale D. A., Helou G., Magdis G. E., Armus L., Díaz-Santos T., Shi Y., 2014, ApJ, 784, 83
Davies L. J. M. et al., 2025, MNRAS, 544, 3005
De Zotti G., Bonato M., Giulietti M., Massardi M., Negrello M., Algera H. S. B., Delhaize J., 2024, A&A, 689, A272
Delhaize J. et al., 2017, A&A, 602, A4
Delvecchio I. et al., 2021, A&A, 647, A123
Donley J. L. et al., 2012, ApJ, 748, 142
Draine B. T., 2011, Physics of the interstellar and intergalactic medium (Vol. 41). Princeton University Press, Princeton
Drake A. B. et al., 2024, MNRAS, 534, 1107
Duncan K. J., Shivaei I., Shapley A. E., Reddy N. A., Mobasher B., Coil A. L., Kriek M., Siana B., 2020, MNRAS, 498, 3648
Fawcett V. A., Alexander D. M., Rosario D. J., Klindt L., Fotopoulou S., Lusso E., Morabito L. K., Calistro Rivera G., 2020, MNRAS, 494, 4802
Garn T., Green D. A., Riley J. M., Alexander P., 2009, MNRAS, 397, 1101
Gupta Y. et al., 2017, Curr. Sci., 113, 707
Gürkan G. et al., 2018, MNRAS, 475, 3010
Hale C. L. et al., 2025, MNRAS, 536, 2187
Hale C. L. et al., 2026, MNRAS, 552
Hardcastle M. J., Looney L. W., 2008, MNRAS, 388, 176
Hatfield P. W., Jarvis M. J., Adams N., Bowler R. A. A., Häußler B., Duncan K. J., 2022, MNRAS, 513, 3719
Heesen V. et al., 2022, A&A, 664, A83
Heesen V. et al., 2026, A&A, 712, A242
Helou G., Soifer B. T., Rowan-Robinson M., 1985, ApJ, 298, L7
Herrera Ruiz N. et al., 2017, A&A, 607, A132
Hodge J. A., Becker R. H., White R. L., de Vries W. H., 2008, AJ, 136, 1097
Hoffman M. D., Gelman A. et al., 2014, J. Mach. Learn. Res., 15, 1593
Ibar E. et al., 2008, MNRAS, 386, 953
Ivison R. J. et al., 2010, A&A, 518, L31
Jackson C. L., Matthews J. H., Whittam I. H., Jarvis M. J., Temple M. J., Rankine A. L., Hewett P. C., 2026, MNRAS, 546, stag065
Jarvis M. J. et al., 2010, MNRAS, 409, 92
Jarvis M. J. et al., 2016, in MeerKAT Science: On the Pathway to the SKA. p. 6, preprint (arXiv:1709.01901)
Johnston R., Vaccari M., Jarvis M., Smith M., Giovannoli E., Häußler B., Prescott M., 2015, MNRAS, 453, 2540
Jonas J., 2018, PoS, MeerKAT2016, 001
Jonas J. L., 2009, Proc. IEEE, 97, 1522
de Jong T., Klein U., Wielebinski R., Wunderlich E., 1985, A&A, 147, L6
Khostovan A. A. et al., 2026, ApJS, 282, 6
Klessen R. S., Glover S. C. O., 2016, Saas-Fee Adv. Course, 43, 85
Kondapally R. et al., 2022, MNRAS, 513, 3742
Kumar R., Carroll C., Hartikainen A., Martin O., 2019, J. Open Source Software, 4, 1143
Maddox N. et al., 2017, MNRAS, 470, 2314
Magnelli B. et al., 2015, A&A, 573, A45
McCracken H. J. et al., 2013, The Messenger, 154, 29
Merloni A., Heinz S., 2007, MNRAS, 381, 589
Molnár D. C. et al., 2018, MNRAS, 475, 827
Molnár D. C. et al., 2021, MNRAS, 504, 118
Mountrichas G., Carrera F. J., Georgantopoulos I., Mateos S., Ruiz A., Corral A., 2025, A&A, 700, A234
Murphy E. J., 2009, ApJ, 706, 482
Murphy E. J. et al., 2011, ApJ, 737, 67
Noeske K. G. et al., 2007, ApJ, 660, L43
Novak M. et al., 2017, A&A, 602, A5
Ocran E. F., Taylor A. R., Vaccari M., Ishwara-Chandra C. H., Prandoni I., Prescott M., Mancuso C., 2020, MNRAS, 491, 5911
Panessa F., Baldi R. D., Laor A., Padovani P., Behar E., McHardy I., 2019, Nat. Astron., 3, 387
Peluso G. et al., 2026, A&A, 706, A111
Planck Collaboration VI,, 2020, A&A, 641, A6
Rankine A. L., Matthews J. H., Hewett P. C., Banerji M., Morabito L. K., Richards G. T., 2021, MNRAS, 502, 4154
Seymour N. et al., 2026, ApJ, 998, 306
Shirley R. et al., 2021, MNRAS, 507, 129
Smith D. J. B. et al., 2021, A&A, 648, A6
Stylianou N., Jarvis Matt J., Varadaraj R. G., Hatfield P. W., Hamlett J., Hale C. L., Yasin T. , 2026, MNRAS
Smolčić V. et al., 2017, A&A, 602, A6
Sokol A. D., Yun M., Pope A., Kirkpatrick A., Cooke K., 2023, MNRAS, 521, 818
Swarup G., 1991, in Cornwell T. J., Perley R. A., eds, ASP Conf. Ser. Vol. 19, IAU Colloquium 131: Radio Interferometry. Theory, Techniques, and Applications. Astron. Soc. Pac., San Francisco, p. 376
Symeonidis M., Giblin B. M., Page M. J., Pearson C., Bendo G., Seymour N., Oliver S. J., 2016, MNRAS, 459, 257
Symeonidis M. et al., 2022, MNRAS, 514, 4450
Szokoly G. P. et al., 2004, ApJS, 155, 271
Tabatabaei F. S. et al., 2017, ApJ, 836, 185
Tabatabaei F. et al., 2025, ApJ, 989, 44
Vaccari M., 2015, in The Many Facets of Extragalactic Radio Surveys: Towards New Scientific Challenges. p. 27, preprint (arXiv:1604.02353)
Vaccari M., 2026, Res. Notes Am. Astron. Soc., 10, 118
van Haarlem M. P. et al., 2013, A&A, 556, A2
Vehtari A., Gelman A., Gabry J., 2017, Stat. Comput., 27, 1413
Vehtari A., Simpson D., Gelman A., Yao Y., Gabry J., 2024, J. Mach. Learn. Res., 25, 1
Villa-Vélez J. A., Buat V., Theulé P., Boquien M., Burgarella D., 2021, A&A, 654, A153
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Watanabe S., Opper M., 2010, Journal of machine learning research, 11
Weaver J. R. et al., 2022, ApJS, 258, 11
Whitaker K. E. et al., 2014, ApJ, 795, 104
Whitaker K. E., Pope A., Cybulski R., Casey C. M., Popping G., Yun M. S., 2017, ApJ, 850, 208
White S. V., Jarvis M. J., Kalfountzou E., Hardcastle M. J., Verma A., Cao Orjales J. M., Stevens J., 2017, MNRAS, 468, 217
Whittam I. H. et al., 2022, MNRAS, 516, 245
Whittam I. H. et al., 2025, MNRAS, 543, 507
Yang G. et al., 2022, ApJ, 927, 192
Yun M. S., Reddy N. A., Condon J. J., 2001, ApJ, 554, 803
Zakamska N. L., Greene J. E., 2014, MNRAS, 442, 784
Zhu C., Byrd R. H., Lu P., Nocedal J., 1997, ACM Trans. Math. Softw., 23, 550

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

© The Author(s) 2026.
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from academic.oup.com/mnras/article/553/1/stag1947/8943798 by University of Hertfordshire user on 24 November 2026