Universiteit Leiden
The Netherlands

Matching LOFAR sources across radio bands

Böhme, L.; Schwarz, D.J.; Gasperin, F. de; Röttgering, H.J.A.; Williams, W.L.

Matching LOFAR sources across radio bands[*]

L. Böhme[1] , D. J. Schwarz[1], F. de Gasperin[2,3] , H. J. A. Röttgering[4], and W. L. Williams[4,5]

[1] Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
  e-mail: Lboehme@physik.uni-bielefeld.de
[2] Istituto di Radioastronomia, Istituto Nazionale di Astrofisica, Via Gobetti 101, 40127 Bologna, Italy
[3] Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany
[4] Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
[5] SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield SK11 9FT, UK

ABSTRACT

Aims. The recent preliminary release of the LOFAR LBA Sky Survey (LoLSS) is the first wide-area, ultra-low frequency observation published from the Low Frequency Array (LOFAR). Our aim is to combine this data set with other surveys at higher frequencies to study the spectral properties of a large sample of radio sources.
Methods. We present a new cross-matching algorithm that takes into account the sizes of the radio sources. We applied it to the LoLSS-PR, LoTSS-DR1, LoTSS-DR2 (all LOFAR), TGSS-ADR1 (GMRT), WENSS (WSRT), and NVSS (VLA) catalogues. We then studied the number of matched counterparts for LoLSS radio sources and their spectral properties.
Results. We found counterparts for 22 607 (89.5%) LoLSS sources. The remaining 2640 sources (10.5%) were identified either as an artefact in the LoLSS survey (3.6%) or flagged due to their closeness to bright sources (6.9%). We found an average spectral index of $\alpha = -0.77 \pm 0.18$ between LoLSS and NVSS. Between LoLSS and LoTSS-DR2, we found $\alpha = -0.71 \pm 0.31$. The average spectral index is independent of the flux density above $S_{54} = 181$ mJy. A comparison of the spectral slopes from LoLSS–LoTSS-DR2 with LoTSS-DR2–NVSS indicates that the probed population of radio sources shows evidence of a negative spectral curvature.

Key words. radio continuum: general – galaxies: general – methods: data analysis – catalogs

1. Introduction

Radio continuum surveys map areas of the sky at a range of depths, resolution, and frequency. The cross-identification of sources from surveys at different radio frequencies and other wavelengths is a key element in the astrophysical interpretation of radio sources.

For the simplest cross-matching task, a single emission mechanism is dominant, such as synchrotron (radio) or thermal dust emission (infrared). In that case, the general properties of the spectral energy distribution (SED) are known, namely, whether it has a power-law or black-body spectrum. Then, the main task is to measure the flux density at different frequencies and to determine a small number of parameters that describe the SED (e.g. the spectral index or physical temperature). In contrast, cross-matching of sources involving multiple emission mechanisms is more complicated. It requires full multi-wavelength (and sometimes multi-messenger) astronomy. This is, for example, the case when matching radio with optical and infrared data. Here, the general form of the SED is unknown and the task is to identify sources across different wavelengths.

In this work, we focus on the synchrotron regime by cross-matching continuum radio surveys from the lowest radio frequencies as observed with the Low Frequency Array (LOFAR; van Haarlem et al. 2013), up to 1.4 GHz. We use a new algorithm for radio-radio cross-matching of catalogues, which also

incorporates the source extension of resolved sources. We apply it to radio catalogues from LOFAR, the Giant Metrewave Radio Telescope (GMRT), the Westerbork Synthesis Radio Telescope (WSRT), and the Very Large Array (VLA). The corresponding radio surveys are the LOFAR LBA Sky Survey (LoLSS; de Gasperin et al. 2021), the LOFAR Two-meter Sky Survey (LoTSS; Shimwell et al. 2019, 2022), the TIFR GMRT Sky Survey alternative data release 1[1] (TGSS ADR1; Intema et al. 2017), the Westerbork Northern Sky Survey (WENSS; Rengelink et al. 1997), and the NRAO VLA Sky Survey (NVSS; Condon et al. 1998). These radio surveys were chosen, as they all cover the *Hobby-Eberly* Telescope Dark Energy Experiment (HETDEX; Hill et al. 2008) Spring Field and have a broad range in frequencies from 54 to 1400 MHz. The angular resolutions of the surveys are between 6″ and 54″. The final cross-matching catalogue includes 22 624 LoLSS sources and is publicly available.

Except for the deepest observations, most of the observed objects in the radio sky are active galactic nuclei (AGNs) for which synchrotron radiation from relativistic electrons (CREs) forms the origin of the radio emission (Best et al. 2023; de Gasperin et al. 2018). To gain insight into the physical phenomena that produce the emission, the characteristics of the SED in the radio band is an important tool. Generally, the radio SEDs are smooth over a large frequency range and often a single power law can be used to describe the SEDs (see e.g. Kellermann et al. 1969).

Radio-selected AGNs sample a small fraction (<10%) of all AGNs, with a bias towards bright and jetted AGNs

---

[*] Full catalogue as described in Table 2 and the sub-catalogues described in Sect. 3.5 are available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/674/A189

[1] We will use TGSS as an abbreviation for TGSS ADR1.

(Padovani et al. 2017). In jetted AGNs, there are four types of regions: a core, jet(s), hot spot(s), and lobe(s). Often a binary classification based on the morphology of emission is possible (Fanaroff & Riley 1974). For instance, Fanaroff-Riley class I (FR I) describes centre-brightened sources with jets disrupting on the kpc scales to subsonic plumes. For edge-brightened AGNs (FR II), the jets remain relativistic throughout, escaping the galactic nucleus with little deceleration and eventually each ending up in a hot spot (Mingo et al. 2019). FR II sources are generally brighter, although a low-luminosity population (FR II-Low) was recently found, with lower luminosities by three orders of magnitude (Mingo et al. 2019). Both classes can also be used to describe the different jet interactions with the environment (Blandford et al. 2019). Relativistic jets (FR II) create hot cocoons that protect them from destructive instability, while FR I sources directly heat their galactic surroundings. These FR I plumes can still propagate several Mpc away from their host and the heating can stimulate star formation and may therefore play a major role in the evolution of galaxies (Blandford et al. 2019). For both classes, a flat spectrum core has been found, which may result from synchrotron self-absorption.

The spectral index α is defined as the slope of the radio SED in a log-log plot,

$$\alpha(\nu) = \frac{d \log S_\nu}{d \log \nu}, \qquad (1)$$

with the flux density S_ν. The spatial variation of the spectral index across extended sources is fundamental in understanding the different mechanisms of radio emission in galaxies (de Gasperin et al. 2018). SED studies of radio galaxy lobes, which generally show a negative curvature, with a steeper spectrum at high frequencies and with older age, can offer insights into the source age, energetics, and plasma composition (de Gasperin et al. 2018). At low frequencies, the lobes are the dominant flux contributing component. When integrating the flux into a single point (as for unresolved sources), the lobes therefore play an important role in the measured curvature of the spectral index. Thus, even the measurement of curvature in unresolved sources should give insight into the morphology of these. To estimate the spectral index or SED of a radio source, measurements at different frequencies are needed. Since intensity and resolution, as well as astrometric precision, cannot be expected to agree at different frequencies, the cross-matching of radio sources at different frequencies must ensure that components (or sources) are matched properly.

A fast and easy approach to measure spectral indices in the sychrotron emission regime from source catalogues at different radio frequencies is positional cross-matching (de Gasperin et al. 2018). With a fixed search radius around each source in a catalogue, counterparts in other catalogues at different frequencies can be identified. A caveat of this approach is that resolution differs from survey to survey. Multiple sources in a high angular resolution survey can appear blended into a single source at lower resolution. For surveys at different frequencies, the sources might also appear different, as some source components may not be visible at the lower or higher frequency.

Several improved algorithms have been presented in the literature. In the following, a brief summary is given: the general use of a Likelihood Ratio (LR) was discussed by Sutherland & Saunders (1992), summarising previous works and deriving the reliability of radio with optical associations. A general approach was developed by Budavári & Szalay (2008) for cross-identifying point sources in multiple observations based on a

Likelihood Ratio (LR) using a Bayesian approach. These authors assumed astrometric precision to be the limiting factor. This method gives a more refined cross-matching than the simple positional search based on the assumption that sources have a single component and are point-like (or unresolved); however, it is not suitable for extended and asymmetric sources. An updated version of this LR matching was presented by Fan et al. (2020), which also takes into account the lobe morphology and different angles between them. This solves several drawbacks of the previous model, but is still only used for cross-matching extended radio sources with optical/IR point-sources. It can be used on the results received from the approach of this paper, to cross-identify the optical/IR counterparts for the highest-resolution radio survey and then propagate these results to the other matched radio surveys. A more specific approach is the positional update and matching algorithm (PUMA; Line et al. 2017) which was developed for cross-matching low-frequency radio catalogues and combines the general approach by Budavári & Szalay (2008) with a spectral matching criterion and the possibility for multiple counterparts. It first carries out a positional match between a base catalogue and all others, then isolated matches (having a 1:1 assignment) are accepted if either their assigned matching probability by the algorithm is above some chosen threshold or if a power-law spectral model shows a fit. For non-isolated matches (i.e. for several possible candidate matches in one survey for one source in the base survey), if one of them has a much better power-law fit, it is accepted. If this is not the case, a combination of all single sources with combined flux density is tested for a power-law fit. Overall, the PUMA algorithm gives a better positional match, especially when the source density is high and multiple possible counterparts are nearby. The advantage of LR algorithms is the handling of multiple potential matches (Sutherland & Saunders 1992). In the method presented in this work, every match above some threshold ends up in the catalogue, while LR algorithms account for the possibility that the match sits at a completely different redshift. As only radio surveys are considered in this work and LoLSS sources are dominantly AGNs, this is not an important issue, as the source density is still low enough. However, for deeper fields like the LOFAR Deep Fields (Best et al. 2023), future radio surveys and matching with optical surveys, this will be an issue.

The new algorithm that is presented here presents an improvement over a simple positional cross-matching in a fixed search radius and attempts to take the position, shape, and orientation of radio sources into account. It currently provides a binary answer ('match' or 'no match'), which may be generalised to produce a posterior probability distribution.

In Sect. 2 of this work, we introduce the radio surveys used for this study. Section 3 describes the cross-matching algorithm, how the resulting catalogue was cleaned, and its properties. Section 4 is about the dependency between the spectral index and flux density, redshift, and source compactness. In Sect. 5, we conclude with suggested future updates to the algorithm and summarise our findings.

2. Data

The spectral index catalogue is created from the following five radio surveys. An overview of the most important characteristics of the surveys is also given in Table 1 and their sky coverage is shown in Fig. 1.

LoLSS. preliminary release (PR; de Gasperin et al. 2021) covers 740 deg^2 in the northern sky around the HETDEX Spring

Table 1. Overview of the surveys used in this study.

Survey name	Frequency (MHz)	Resolution (arcsec)	rms noise (mJy beam^{-1})	Sky area (deg^2)	Source density (deg^{-2})	Release	Reference
LoLSS-PR	54	47	5	740	34.1	2021	de Gasperin et al. (2021)
LoTSS-DR1	144	6	0.071	424	765.6	2019	Shimwell et al. (2019)
LoTSS-DR2	144	6	0.083	5635	780.2	2022	Shimwell et al. (2022)
TGSS-ADR1	147	25	3.5	36 900	16.9	2016	Intema et al. (2017)
WENSS	325	$54 \times 54 / \cos(\delta)$	3.9	10 177	22.5	2000	Rengelink et al. (1997)
NVSS	1400	45	0.45	33 000	53.7	1998	Condon et al. (1998)
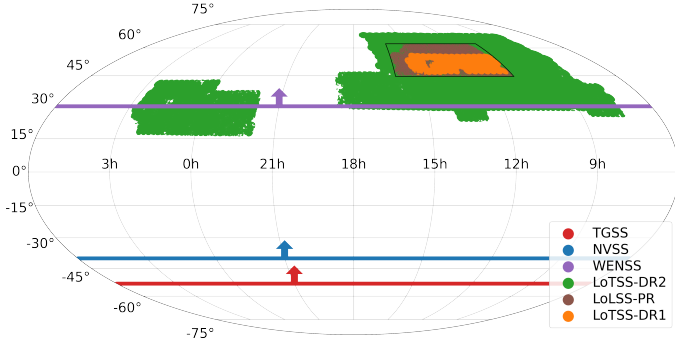
Fig. 1. Coverage map of the radio surveys used in this work. A line indicates the (southern) survey coverage boundary, where the area of observation is in the direction of the arrow pointing north. The black quadrangle around the survey footprint of LoLSS-PR marks the cross-matching region of this study, which is the HETDEX Spring Field.

Field. The observations for the survey were taken with the low band antennas (LBA) using the Dutch stations of LOFAR in the frequency range of 42–66 MHz. For these observations, the 'LBA OUTER' mode was used, whose primary beam's full width at half maximum (FWHM) is similar to the HBA counterpart at 144 MHz. The resolution of LoLSS-PR is 47″, due to the restrictions of direction-independent calibration, with a root mean square (rms) noise of about 5 mJy beam^{-1} and an astrometric accuracy of 2.5″.

LoTSS. data release 1 and 2 (DR1 / DR2; Shimwell et al. 2019, 2022), with the first data release covering a 424 deg^2 region in the northern sky around the HETDEX Spring Field. The second data release covers a significantly larger portion of the northern sky – 5635 deg^2. Both surveys were taken with the high band antennas (HBA) using the Dutch stations of LOFAR. Both releases have a resolution of 6″, with a median sensitivity of 71 and 83 μJy beam^{-1} for DR1 and DR2, respectively. The astrometric accuracy for both releases is 0.2″. For the first data release, the LOFAR Radio Galaxy Zoo results were used to combine multiple components as jets, lobes, etc. present as separated entries in the source catalogue into single entries in the value-added catalogue (Williams et al. 2019). In addition, artefacts were removed. When available, redshift estimates (most of them photometric), as well as infrared and optical identifications (Duncan et al. 2019, 2021) from Pan-STARRS and WISE were added.

NVSS (Condon et al. 1998) sky coverage includes the whole sky north of declination −40°, thereby covering 82% of the whole sky at a central frequency of 1400 MHz. It was observed with the VLA in New Mexico in D and DnC configurations. The resolution is 45″ with an average rms noise of 0.45 mJy beam^{-1}.

The VLA configuration and snapshot duration changed with declination, although this happens below the sky coverage used in this study and does not affect our results.

TGSS. alternative data release 1 (ADR1; Intema et al. 2017) covers the full sky north of Dec −53° (90% of the whole sky) at the central frequency of 147 MHz. It was observed with the Giant Meterwave Radio Telescope (GMRT) in India with a final resolution of 25″ and a rms noise of 3.5 mJy beam^{-1}.

WENSS (Rengelink et al. 1997) covers the full sky north of 30° declination at a frequency of 325 MHz. It was observed with the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands. The resolution is declination dependent and is given as $54'' \times 54'' / \cos(\delta)$. The rms noise is 3.9 mJy beam^{-1}.

3. Cross-matching

3.1. Cross-matching method

The most important consideration when cross-matching these six catalogues is their different resolutions of 6″ (for two surveys), 25″, 45″, 47″, and 54″. Sources that may be resolved into multiple sources or multi-component sources at a high resolution, may appear as a single resolved or even a single unresolved source at lower resolution. Therefore, the source size is an important parameter to consider. In order to include this parameter, a geometrically based search algorithm was developed. It takes one source in the base catalogue – in this work, that is LoLSS, as it has the lowest frequency, a constant low angular resolution and should contain many lobes – and compares it with all sources within the chosen search radius in the other catalogues. As a result, these are either accepted as matches or discarded. The multi-frequency-catalogue is then formed from the accepted matchings[2].

As a preparation, 2D Gaussian functions $G(x, y)$ were created from the major axis (MA), minor axis (MI) and positional angle (PA) for all sources from the different catalogues. These functions use the fitted Gaussian distribution from the source extraction in PyBDSF (Mohan & Rafferty 2015) and describe the FWHM ellipse at the 0.5 (half maximum) level. As a first step, the peak-normalised 2D Gaussian (maximum equal to 1) of a LoLSS-source is taken and its value $v = G(x_{cp}, y_{cp})$ calculated at the centres of the potential counterparts. If this value is above some threshold value v_{thr} in the interval (0, 1), the pair is accepted. For this analysis, v_{thr} was chosen as 0.5, as it describes the FWHM ellipse. When looking at the distribution of the centre matching values between all potential matches (shown in Fig. 2 for the LoLSS and LoTSS-DR2 catalogues), most values

---

[2] An implementation of the algorithm can be found at https://github.com/lboehme/cross-matching-lofar
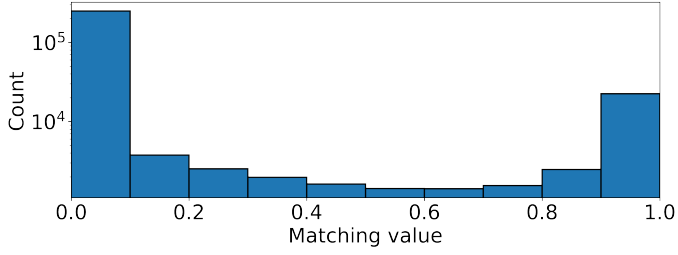
Fig. 2. Distribution of the Gaussian matching values v for all potential matches calculated at their centre.
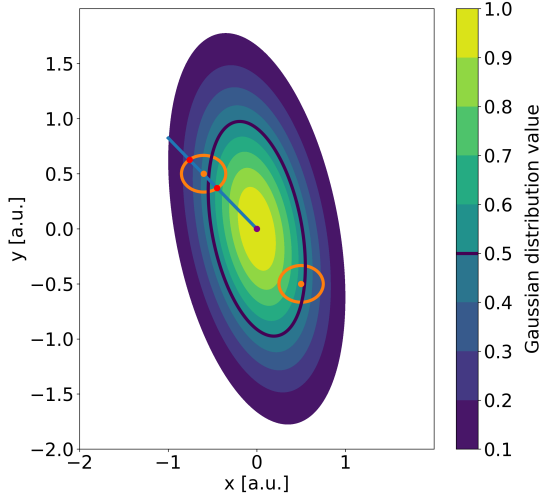

Fig. 3. Sketch of three sources matching. The original (LoLSS) source in the middle is displayed as filled contours with its Gaussian distribution in 0.1 steps from 0 (minimum) to 1 (maximum). The potential counterpart sources are shown with their FWHM in orange. The blue line connects both centres and the red points are where the Gaussian value, v, is additionally calculated.

are either close to 0 or 1. Therefore, the impact of the precise choice of the threshold at values between 0.5 and 0.7 is small. An example for an immediate match is given in Fig. 3 and shown as the lower right orange contour. Here, the centre is inside the FWHM ellipse and therefore $v > v_{thr}$.

If the potential counterpart was not accepted, in a second step, another value was calculated. For this, both source centres are connected with a line and v is calculated at the intersection with the FWHM ellipse of the counterpart source (red dots in Fig. 3). If v at one of these coordinates is above $1.1v_{thr} = 0.55$, the pair is also accepted. The value of 1.1 was chosen from testing the algorithm in several example situations and lead to more consistent matching results. If it were chosen as 1.0, it would be enough for the two FWHM to only touch at one point, but as some overlap is demanded for physical matches, this case is excluded. If a higher value was chosen, that is, at 1.2, the condition would be too strict and matches would be missing. This value may offer some room for improvement in further iterations of the method, but this issue is beyond the scope for this first release. This additional property ensures the correct identification of elongated sources, while the slightly higher threshold keeps misclassifications low. A sketch of this second step is given in Fig. 3 (the upper left orange contour), while a real example from the LoLSS-LoTSS-DR2 cross-match is given in Fig. 4 (the upper left match). Here, 205″ was chosen as the search radius, as it is the lowest radius which includes all accepted matches found up to 1000″. This is demonstrated in Fig. 5.
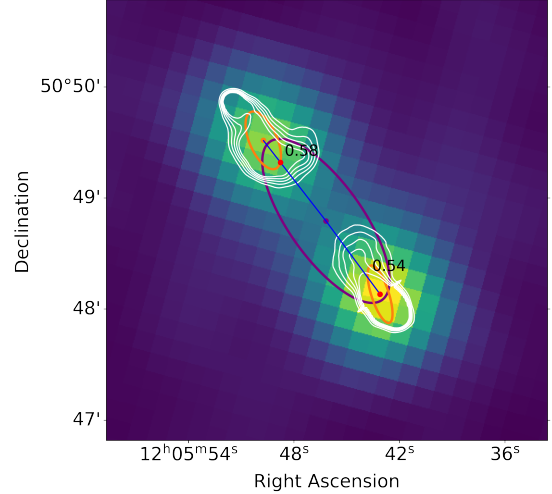
Fig. 4. Example of two sources in LoTSS-DR2 (orange and white contours) matching with a LoLSS (purple outline and colour background) source. The orange and purple ellipses represent the FWHM, the blue line connects the centres, while the red dot is the point where the value exceeds the matching condition and is accepted. Both matches are accepted, because either the second matching condition (upper left source) or the first (lower right source) is fulfilled, as indicated by the values.
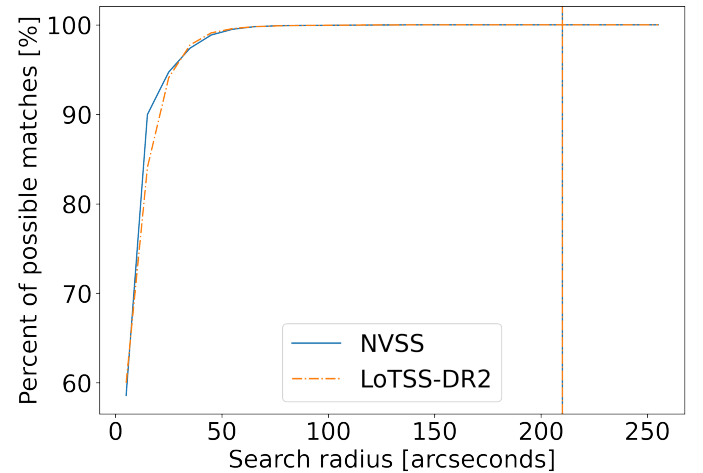

Fig. 5. Relation between search radius and included percentage of accepted matches. The dotted vertical lines give the radius of 100% completeness. The 100% completeness was calculated by finding all accepted matches in a 1000″ search radius.

3.2. Artefacts and cleaning

To refine these results, the unmatched LoLSS sources are checked for their distance to bright sources, as the noise is especially high in the vicinity of such sources. This noise may have been picked up by the source finder PyBDSF and imaging artefacts were thereby added to the source catalogue. In Fig. 6 the distance of matched (in any survey) and unmatched (in all surveys) sources to the next bright source (>0.5 Jy in LoLSS) is displayed. For the matched sources, a nearly constant count above 400″ is apparent. For the unmatched sources, around 75% of the sources are closer than 200″ to a bright source and 85% closer than 300″. Since a constant relation can be seen above 300″, all unmatched sources (no match in any survey) up to 300″ close to a bright source were removed. This labels about 819 LoLSS sources as artefacts.
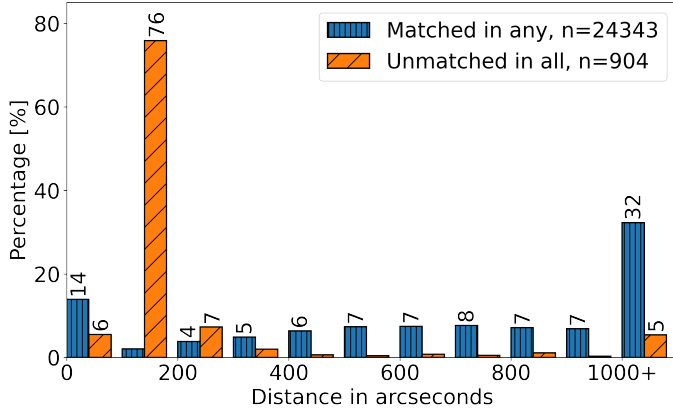
Fig. 6. Percentage of sources for a given distance from the nearest bright sources – defined here as brighter than 0.5 Jy in LoLSS. The 1000+ bin includes all sources with a distance greater than 1000″ from the nearest bright source and therefore has a different area than the other bins.

Fig. 7. Percentage of sources in three spectral index bins for a given distance from the nearest bright sources – defined here as brighter than 0.5 Jy in LoLSS. The 1000+ bin includes all sources with a distance greater than 1000″ from the nearest bright source and therefore has a different area than the other bins.

Fig. 8. Empirical probability for the number of source matches for LoTSS-DR1 value-added, LoTSS-DR2, NVSS, TGSS, and WENSS per non-artefact LoLSS source.

For the remaining unmatched sources, a visual inspection was done which led to the conclusion that at least 90% of these are artefacts and not actual sources. This is apparent due to their closeness to sources with calibration errors and the comparison with the higher resolution survey LoTSS-DR1 or -DR2. Therefore, as a conservative choice, all unmatched sources are removed. Overall, 956 unmatched LoLSS sources have been flagged as artefacts, which corresponds to a total of 3.79% of the LoLSS catalogue.

In the publication accompanying the release of the data set (de Gasperin et al. 2021), 1055 sources (4%) were estimated as being false positives. These were found to be mostly concentrated at the edges and around bright sources. This is in good agreement with the findings presented in this work.

In addition to the anomalously high number of unmatched sources close to bright sources, we also find that a high number ($\approx$40%) of steep ($\alpha < -3$) sources are located up to 300″ around bright sources. In Fig. 7 the distance distribution of sources in three spectral index bins is shown. Combined with the previous removal of unmatched sources below 300″ around bright sources, we decided to make a general cut and remove all sources closer than 300″ to any bright source. These are labelled as artefacts in the catalogue although their origin is still unclear. Despite labelling them as artefacts, some of those sources might nonetheless prove to be real, but may have been recovered with largely offset flux density, especially in LoLSS-PR due to the missing direction-dependent calibration. The offset could be due to the presence of a bright source in the neighbourhood leading to imaging artefacts, while a higher LoLSS flux density leads to a steeper spectral index. With this additional 6%, an overall 2640 (10.5%) sources are labelled as artefacts.

3.3. Properties of the catalogue

The cross-matching was done pairwise and the base catalogue is LoLSS in all cases. It was cross-matched with LoTSS-DR1, LoTSS-DR2, TGSS, WENSS, and NVSS. The final catalogue is described in Table 2. The number of matches per LoLSS source for each counterpart survey is shown in Fig. 8. There it is apparent that most sources have exactly one counterpart and LoTSS-DR2 has the most matches. Only very few sources (<150) do not have a counterpart in LoTSS-DR2, but in other surveys. These are mostly due to problems with the source finding in LoLSS-PR, where one ellipse is ill fitted to two lobes
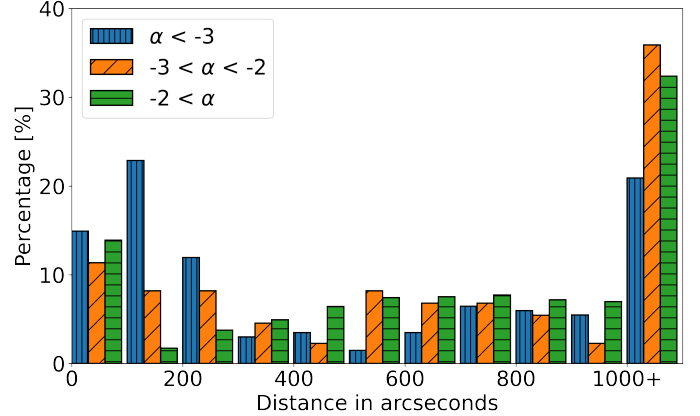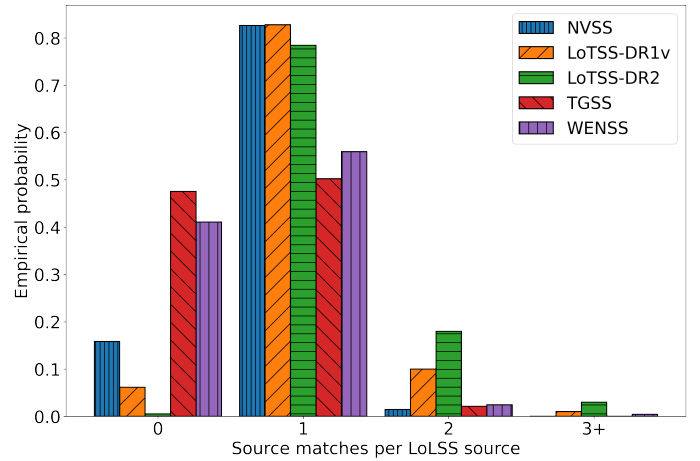
without a core. This should be fixed in further iterations of the LoLSS survey. Few LoLSS sources (less than 20%) have two counterparts in LoTSS-DR2. In the other catalogues, very few sources (<10%) have more than one counterpart. For NVSS with a low source density and similar resolution to LoLSS, less than 300 multiple counterparts (2+) were found, while for 80% of the LoLSS sources, a one-to-one match was found. In TGSS and WENSS, no counterpart was found for around 45% of the LoLSS sources. Our analysis does not further investigate whether multiple associations are exclusively physical or contain chance associations due to blending.

The cross-matching catalogue has 25 247 entries and is available at CDS, of these 2640 have no counterpart and are labelled as 'artefacts' (A). Of the remaining, 16 266 are labelled as 'single' (S) when all matches are singular and 6341 as 'multiple' (M), if in any survey multiple counterparts were matched.

For all matches, the spectral index was calculated by bootstrapping 1000 times the flux density values for each survey from a Gaussian distribution. For S-type sources, it is centred on the measured flux density value with a standard deviation according to the flux density error given in the catalogue. For M-type sources, the Gaussian distribution is centred on the sum of the individual flux densities, while the error is given by the root

Table 2. Description of catalogue columns, available at CDS.

ID	Column name	Unit	Notes
0	Source_id_L	–	LoLSS source identification number
1	RA	degree	LoLSS position
2	Dec	degree	LoLSS position
3	Flux_LoLSS	mJy	Sum of detection fluxes
4	E_Flux_LoLSS	mJy	Sum of detection flux errors
5	Peak_Flux_LoLSS	mJy beam^{-1}	Maximum of detection peak fluxes
6	E_Peak_Flux_LoLSS	mJy beam^{-1}	Maximum of detection peak flux errors
7	Rms_LoLSS	mJy beam^{-1}	Local rms noise
8–33			3–7 Repeated for all other surveys
33	Spidx_LT1	–	Spectral index between LoLSS and LoTSS-DR1
34	Spidx_LT2	–	Spectral index between LoLSS and LoTSS-DR2
35	Spidx_LG	–	Spectral index between LoLSS and TGSS
36	Spidx_LW	–	Spectral index between LoLSS and WENSS
37	Spidx_LN	–	Spectral index between LoLSS and NVSS
38	Scode	S,M,A	S = Single, M = Multiple, A = Artefact/Close to bright source
39	Num_match_T1	–	Number of LoTSS-DR1 counterparts
40	Source_id_T1	–	Array of source IDs with Num_match_T1 entries
41	Num_match_T2	–	Number of LoTSS-DR2 counterparts
42	Source_id_T2	–	Array with Num_match_T2 entries
43	Num_match_G	–	Number of TGSS counterparts
44	Source_id_G	–	Array with Num_match_G entries
45	Num_match_W	–	Number of WENSS counterparts
46	Source_id_W	–	Array with Num_match_W entries
47	Num_match_N	–	Number of NVSS counterparts
48	Source_id_N	–	Array with Num_match_N entries

mean square of the individual errors. If no flux density error is given, 10% of the flux density is used. The resulting mean is used as the spectral index and the standard deviation as the corresponding error. A chance association to any given association normally has a very small impact on the spectral index, as in most of these cases, the real association will be significantly brighter. This method is the same as de Gasperin et al. (2018) used to calculate the spectral index of TGSS–NVSS. The normalised spectral index distribution is then retrieved by stacking normalised Gaussians with the values retrieved from the bootstrap and dividing by the amount of spectral indices. The spectral index distributions for LoLSS–NVSS and LoLSS–LoTSS-DR2 are shown in Figs. 9 and 10, respectively. For the LoLSS–NVSS spectral indices a narrow peak at -0.77 ± 0.18 is found after fitting a Gaussian to the central distribution. It is also apparent that the distribution is skewed towards positive values. For the LoLSS–LoTSS-DR2, a broader peak centred at -0.71 ± 0.31 is found, with significant counts down to -3 and up to 1. To check whether the multiple matches are physical matches or just chance alignments, we also compared the spectral index of the single and multiple matches. Both distributions only differ by $\Delta\alpha = 0.002$ and $\Delta\sigma = 0.03$. The spectral index distribution of LoLSS–TGSS and LoLSS–WENSS is shown in Figs. A.1 and A.2. These have a fitted mean spectral index of -0.82 ± 0.36 and -0.77 ± 0.24 for TGSS and WENSS, respectively.

We compared the spectral index values from this study with a similar study from de Gasperin et al. (2018). In that work, a spectral index catalogue was retrieved from NVSS and TGSS by convolving TGSS to 45″, re-gridding the mosaic images, and running PyBDSF on these. The comparison was done by a quick positional cross-match of LoLSS source positions in a
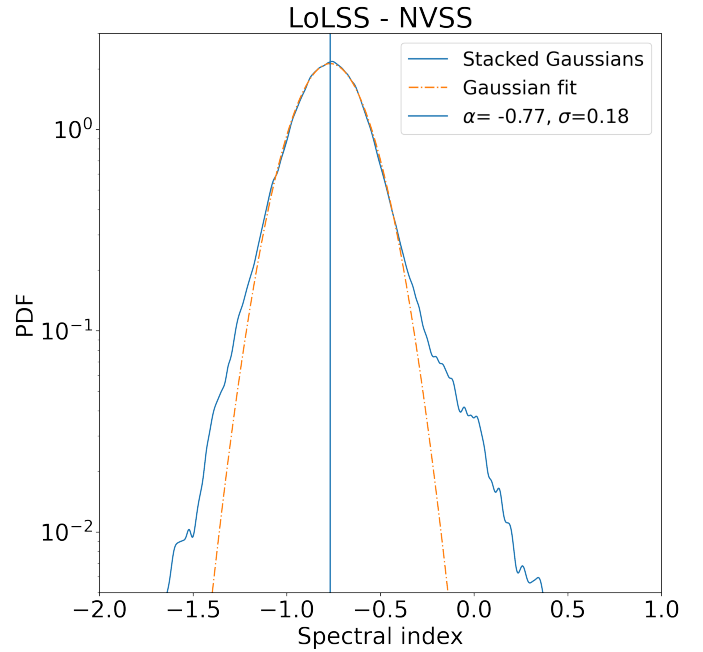
Fig. 9. Spectral index distribution for sources matched in LoLSS and NVSS. The orange line is a Gaussian fit with the parameters noted in the legend.

standard search radius of 5″. For each match, the spectral index difference $\Delta_\alpha = \alpha^{\text{NVSS or LoTSS}}_{\text{LoLSS}} - \alpha^{\text{NVSS}}_{\text{TGSS}}$ as the difference of both spectral indices was calculated. We did this for both the LoLSS–NVSS and the LoLSS–LoTSS-DR2 matching. For NVSS, we
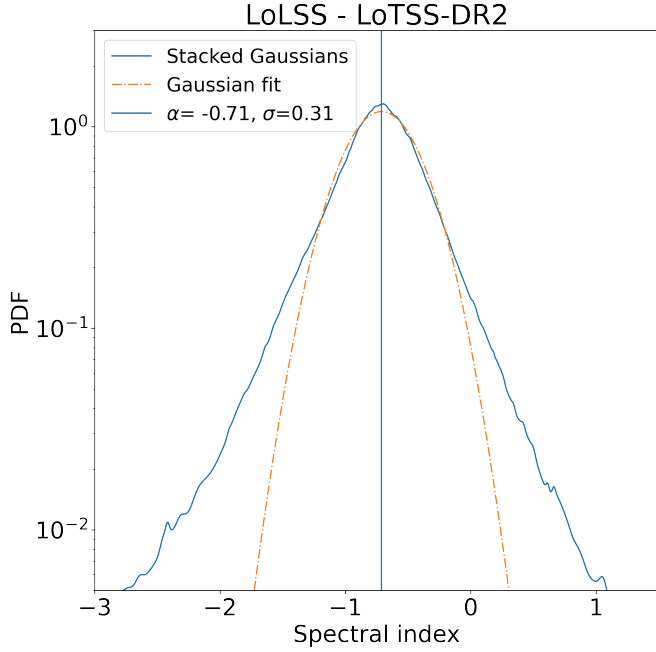
Fig. 10. Spectral index distribution for sources matched in LoLSS and LoTSS-DR2. The orange line is a Gaussian fit with the parameters noted in the legend.
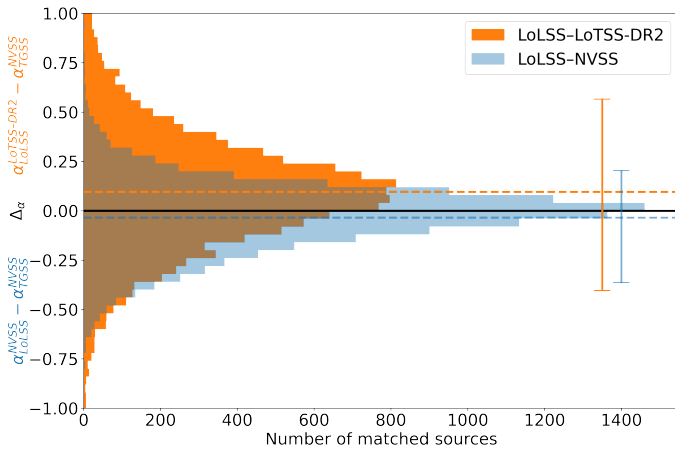

Fig. 11. Spectral index difference between two spectral indices calculated in this study (LoLSS–LoTSS-DR2 and LoLSS–NVSS) and the TGSS–NVSS spectral index catalogue of de Gasperin et al. (2018). The dashed lines indicate the mean value and the bars on the right depict the 90 per cent interval of Δ_α.

find an average $\Delta_\alpha = -0.034 \pm 0.178$ and for LoTSS-DR2 $\Delta_\alpha = 0.097 \pm 0.333$. This indicates that the LoLSS–LoTSS-DR2 spectral indices are on average flatter. In Fig. 11, the 90% errors of the Δ_α distribution are plotted.

3.4. Cross-matching incompleteness

The cross-matching incompleteness can be estimated for each survey matched with LoLSS by the number of unmatched, non-artefact LoLSS sources. For this purpose, LoLSS is divided into logarithmic flux density bins and in each bin the fraction of unmatched to the total number of LoLSS sources in that bin is calculated. This incompleteness is shown in the right panel of Fig. 12 for LoTSS-DR2. At low LoLSS flux densities, the incompleteness is very low at the 10^{-3} level, indicating that virtually all sources have a counterpart in LoTSS-DR2. Overall,
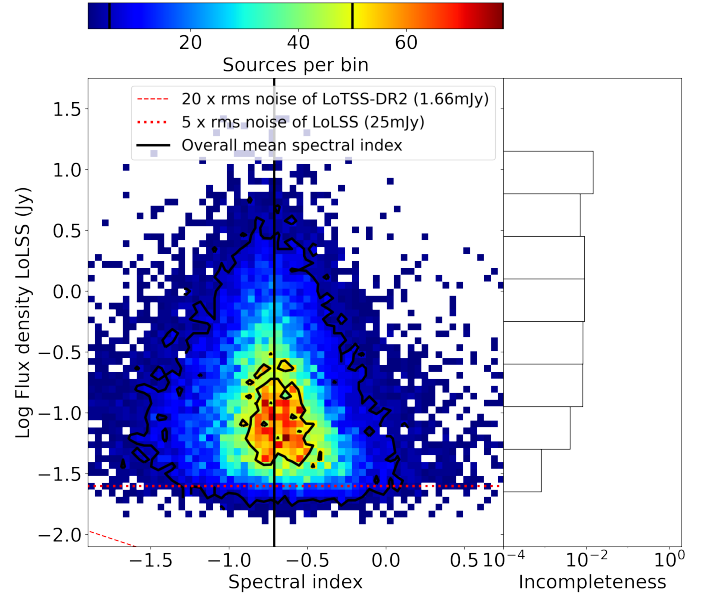
Fig. 12. Spectral index and flux density 2D histogram. Right plot gives the incompleteness as the ratio of unmatched to all sources in a given flux density bin. The red dashed lines indicate the flux density limits of the surveys, while the black vertical line indicates the average spectral index. The black lines in the color bar define the contours.

the incompleteness stays around the 1% level. Therefore, this cross-matching is always ≥99% complete in terms of the cross-matching . The incompleteness of the cross-matching for the other surveys with LoLSS is shown in Fig. A.3. A strong similarity between NVSS and TGSS can be seen. Both are close to an incompleteness of one at very faint flux densities and only reach around 2×10^{-2} as the lowest incompleteness. At high flux densities, both become more incomplete, although this is biased due to less populated bins at high flux densities. At low- and mid-flux densities, the WENSS match behaves similarly, reaching values below 10^{-2}. It is even 100% complete above $0.5 \log_{10}$ Jy.

3.5. Specific sub-catalogues

As the whole catalogue is rather large and sources for further scientific research may not be immediately visible, we produced three additional, smaller sub-catalogues (also available at CDS), which highlight sources with different characteristics, although all are candidates for high-redshift radio galaxies (HzRGs). Many HzRGs are hosted by some of the most luminous and massive galaxies, which may be the predecessors for massive ellipticals at low redshifts (Best et al. 1998; Carilli et al. 2002; Reuland et al. 2004; Afonso et al. 2011; Saxena et al. 2019). One successful tracer for HzRGs is the relation between radio spectra and redshift, as a large fraction of the steepest spectra sources also have a high-redshift (Miley & De Breuck 2008).

These sub-catalogues are all based on the LoLSS–LoTSS-DR2 cross-matching. The first subcatalogue is based solely on ultra steep spectrum (USS) sources and includes 292 sources with a spectral index below minus two and may still include artefacts. The second one includes 9 high-redshift ($z > 2$) and steep spectrum ($\alpha < -1$) sources, which were visually inspected for artefacts or matching errors. For the third subcatalogue, a stronger selection criteria based on Saxena et al. (2019) is chosen: compact (MAJ < 10″ and Peak/Total flux density >0.8),

single component, and steep spectrum sources ($\alpha_{54}^{144} < -1.3$) – resulting in 46 sources. All conditions were evaluated in LoTSS-DR2 and the sources visually inspected. All 46 sources have a flux density below $S_{144} = 50$ mJy, which was the lowest flux density in the sample of Saxena et al. (2019); therefore no overlap is expected.

4. Discussion

4.1. Cross-matching method

We compared our cross-matching method also with the simple positional cross-matching. For this purpose, we tried two different matching radii, 50″ (close to the resolution of LoLSS) and 12″, which is two times the resolution of LoTSS. For a radius of 50″, it leads to more multiple matches while also resulting in a more enhanced tail of steep spectral sources, therefore indicating that these are chance associations (random matches lead to steep indices). When using a radius of 12″ for the positional cross-matching, significantly fewer counterparts are matched and most doubles in LoTSS-DR2 are missing. Therefore, we conclude that our algorithm provides better results than a simple positional cross-matching.

The other introduced cross-matching methods from Sect. 1, which were developed for cross-identifying radio sources in optical surveys, are also expected to show worse performance than our developed method. That is because they can only take into account the source extensions from one survey, while the method introduced in this paper uses the source extensions from both surveys. This method should also be able to do the optical cross-identification between extended (radio) and point-like (optical) sources, however, it will perform worse than, for example Fan et al. (2020), which also cross-matches the lobes and core at the same time. However, this is not the problem for which our method was developed, our goal was to match extended sources with extended sources from different radio surveys. In such cases, it can also be applied to all other radio surveys – or even in other cases where a matching of extended or resolved sources is needed.

4.2. Spectral index and flux density

Previous studies have found conflicting results as to how the median spectral index changes with flux density. Some found a flatter spectral index at low flux densities (Prandoni et al. 2006; Ishwara-Chandra et al. 2010; Intema et al. 2011; Williams et al. 2021), albeit with inconsistencies in the rate of change between them. In other studies, a constant spectral index was observed (Kapahi & Kulkarni 1986; Sirothia et al. 2009).

Figure 12 shows the spectral indices between LoLSS and LoTSS-DR2, along with the LoLSS flux densities in the main panel and the above-mentioned matching incompleteness in the right panel. The two red dashed lines indicate two flux density values: 5× rms noise of LoLSS (25 mJy) and 20× rms noise of LoTSS-DR2 (1.66 mJy). The LoTSS-DR2 value was chosen with a factor of 20, as it would not show up otherwise due to being too low. The vertical black line indicates the average spectral index as found before in Fig. 10. Above 5× the rms noise of LoLSS, no bias due to incomplete data impacts the results. Overall, the majority of sources can be found at the fainter flux density bins. The whole shape appears nearly symmetric around the mean spectral index. For the right panel, the cross-matching incompleteness is calculated as explained before bin-wise for ten bins between –2 and 1.5 in units of ($\log_{10}$ Jy). The first and last
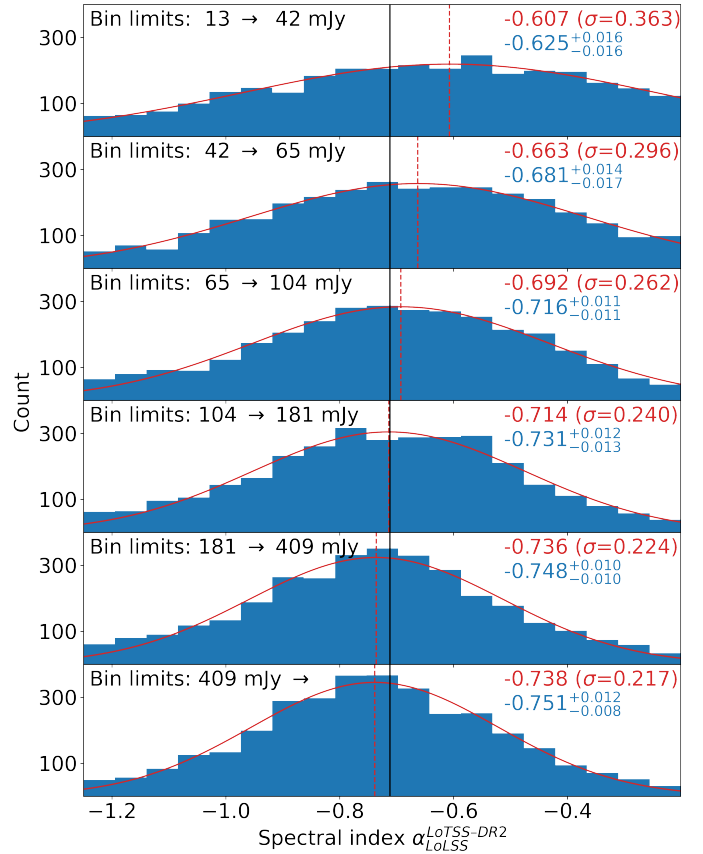

Fig. 13. Spectral index $\alpha_{LoLSS}^{LoTSS-DR2}$ distribution in six equally-numbered flux density bins at 54 MHz. In each bin, a Gaussian fit and its centre is marked and the fit details (mean and standard deviation) are in red in the top right corner. Below that, in blue, is the median, along with the 95% confidence interval ranges. In each top-left corner, the flux density limits of the bin are noted, while the black line indicates the overall mean as previously found in Fig. 10.

bin do not show up on the histogram, as their incompleteness is zero, and they can also not be trusted, given that their counts are too low.

For a closer inspection of the flux density dependence of the spectral index, the spectral index was binned in six equally-populated flux density bins and the histogram is shown in Fig. 13. For each bin, a Gaussian was fit to the data to retrieve the Gaussian mean and standard deviation; additionally, the sample median was calculated. Since LoTSS-DR2 has much better sensitivity than LoLSS, no limitation on the spectral index was expected, even the faintest flux density bin should be complete above $\alpha = -3.51$. A flatter spectral index can be seen in the three faintest bins, where the distributions are also more widespread ($\sigma = 0.36$ in the first bin, $\sigma = 0.26$ in the third bin). Above a flux density of $S_{54} = 181$ mJy, the mean spectral index is found to be independent of flux density.

In the following, the median spectral index is compared between the different surveys used in this study. In Fig. 14, the median spectral index between LoLSS and each survey is shown. For this, as before, the median spectral index is calculated in six equally-populated flux density bins. Since no flux density cuts are applied to any survey catalogue, the first bin may feature incomplete data and is therefore not suitable for comparison and greyed out. Overall, for each survey pair a flattening in at least the lowest flux density bin is apparent, although it is sometimes within the error bars. The LoLSS-LoTSS–DR1 spectral
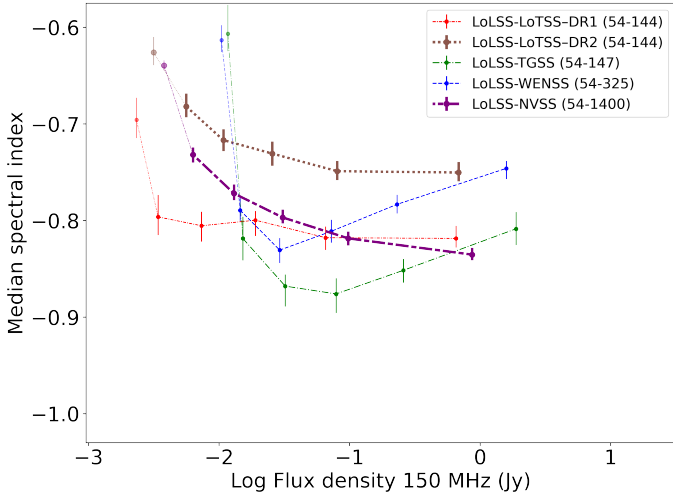
Fig. 14. Comparison of median spectral index in flux density bins for the surveys introduced in Sect. 2. On the x-axis the LoLSS flux density, scaled to 150 MHz according to the spectral index of each bin, is shown. The 95% confidence intervals are shown by the error bars. The first bin is grayed out due to completeness concerns.

Fig. 15. Comparison of median spectral index in flux density bins. Data is taken from Intema et al. (2011), Williams et al. (2013), Mahony et al. (2016), de Gasperin et al. (2018) and Fig. 13 of this work. Data from this work is scaled to 150 MHz according to the spectral index of each bin. The 95% confidence intervals are shown by the error bars for this work. The first bin for each line from this work is greyed out due to completeness concerns.

Fig. 16. LoLSS flux density plotted against peak-to-total flux ratio from LoLSS with the spectral index to LoTSS-DR2 as colour on the *z*-axis. The contours indicate the number counts at the 5, 10, 20, 40, and 80 levels.

index is the only one for which all bins but the first (which may be incomplete) show a near-constant value. It is important to note that due to an improved flux density calibration scheme for LoTSS–DR2, which is aligned with NVSS and 6C, the values differ on the order of up to 10% to LoTSS–DR1. Therefore, the spectral indices involving these two are different. For the TGSS and WENSS match a similar shape is found, where both spectral indices become flatter again with higher flux density, but the TGSS shows more negative values in all except the first bin. Any remaining discrepancies between the various results may either be due to systematic differences in the flux density scale on the order of up to 10% in the different surveys (TGSS, WENSS, LoTSS-DR1 compared to LoTSS-DR2 and NVSS), which can lead to a difference in the spectral index of up to 0.1, but should not influence trends within each result. However, for LoLSS–LoTSS-DR2 and LoLSS–NVSS, which are on the same flux density scale, the differences may be due to spectral curvature (see Sect. 4.3).

The median spectral index is also compared to similar studies in the frequency range of 150 to 1400 MHz by Intema et al. (2011), Williams et al. (2013), Mahony et al. (2016), de Gasperin et al. (2018) in Fig. 15. We compare these to our calculation of the spectral index between LoLSS and LoTSS-DR2 and LoLSS and NVSS. In all studies, including our results, we find the same trend of a flattening of the spectral index at the lower flux densities. Any remaining discrepancies in Fig. 15 may be due to the usage of different generations of LoTSS data, each with a different flux density calibration, and TGSS. However, we note that the trends within each result should be consistent.

4.2.1. Compact and extended sources

The flux density dependence is further observed based on the compactness of sources. The compactness is calculated as the peak-to-total flux ratio of the matched sources. In Fig. 16, the spectral index as a function of LoLSS flux density and compactness is plotted. The plot shows that generally compact and faint sources have a flatter spectrum than extended or bright sources. A possible reason for this division lies in the different parts of radio galaxy emission. Young and core-dominated radio galaxies
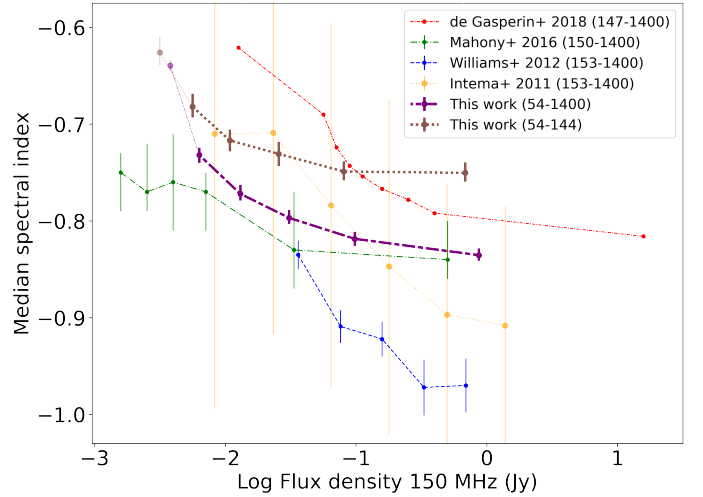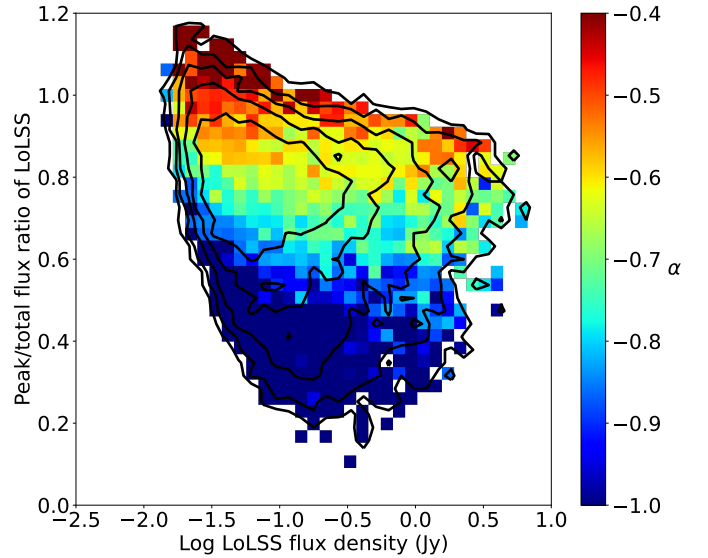
show compact emission with a flat radio spectrum ($\alpha \approx -0.5$). In comparison, in older radio galaxies, the emission of large (thus: less compact) lobes, which is steeper (as discussed in the introduction), dominates over the core emission, moving it more to the right (and slightly lower) in the plot. When the AGN jets shut down, the lobes become fainter, the expansion continues and the emission steepens, thereby moving more to the bottom left of the plot (de Gasperin et al. 2018).

In Fig. 17, the compactness and flux density from LoTSS-DR2 is used instead. The difference between both plots arises from the different resolutions of 6 and 47″ affecting the peak flux densities and the flux density of each bin which scales (differently), depending on its spectral index. For the LoLSS plot
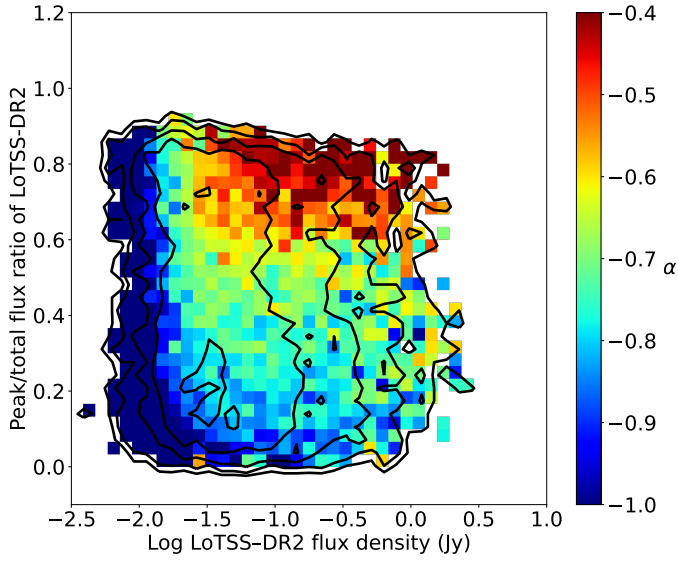
Fig. 17. LoTSS–DR2 flux density plotted against peak-to-total flux ratio from LoTSS–DR2 with the spectral index to LoLSS as colour on the z-axis. The contours indicate the number counts at the 5, 10, 20, 40, and 80 levels.
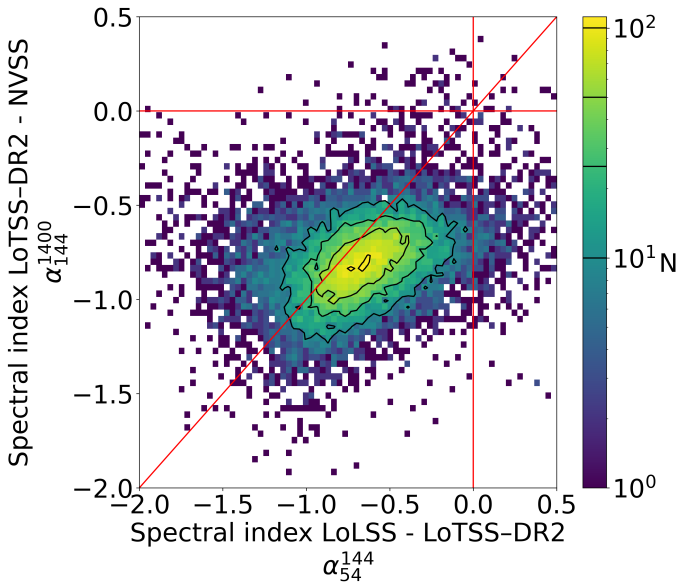

Fig. 18. Spectral curvature between LoLSS – LoTSS-DR2 – NVSS. The colour indicates the number of sources per pixel, while the contours indicate the source counts as marked on the colour scale.

Fig. 16, diffuse sources are missing due to sensitivity limitations of the LBA antennas.

4.3. Spectral curvature

Looking at the difference of the spectral index between LoLSS–LoTSS-DR2 and LoTSS-DR2–NVSS in Fig. 18, it is apparent that most sources exhibit a negatively curved spectrum, namely, the concave shape of the SED. The mean spectral index for LoLSS–LoTSS-DR2 is $\alpha_{54}^{144} = -0.71 \pm 0.31$, while for LoTSS-DR2–NVSS it is found as $\alpha_{144}^{1400} = -0.80 \pm 0.17$. The peak of the 2D histogram in Fig. 18 is offset from the diagonal to flatter values at the lower frequency range and steeper values for 144–1400 MHz. The difference between the spectral indices is

found as $\Delta_\alpha = 0.089$. The same trend was already visible in Fig. 14, where the spectral index between LoLSS and NVSS is flatter than LoLSS–LoTSS-DR2. This is expected because when the turnover happens at, for instance, around 100 MHz, α_{54}^{144} becomes flatter and α_{144}^{1400} steeper than without flattening, while α_{54}^{1400} is in-between. Due to the change to a more reliable flux density calibration for DR2, this effect became apparent. Using the LoTSS-DR1 data, no spectral curvature is evident. The observed field is big enough to rule out the effect of cosmic variance, as a test with two different RA divided samples showed.

A more detailed study of spectral curvature is possible with the direction-dependent calibration and release of LoLSS-DR1 (de Gasperin et al. 2023), which also include in-band spectra. In a first assessment in de Gasperin et al. (2023), a similar trend in spectral curvature was found, which is even more apparent when the in-band spectra are included.

4.4. Spectral index and redshift

As mentioned before, the value-added LoTSS-DR1 catalogue provides (mostly photometric) redshift information for LoTSS-DR1 sources in the catalogue (Duncan et al. 2019). For our subsample of LoTSS-DR2 sources matched to LoLSS, redshift information is available for 5662 of the 12 491 sources.

For each survey matched with LoLSS, the spectral index above and below the mean redshift of $z_{mean} = 0.736$ is almost the same, as shown in Fig. A.4. For LoLSS–LoTSS-DR2, the difference in α above and below z_{mean} is only 0.011, while the biggest difference is in LoLSS–TGSS with 0.035. This result is insignificant compared to the spread of the spectral index of 0.31. Therefore, no change in the spectral index with redshift is observed. The relation was also tested for high-power radio sources ($L_{54} > 10^{27}$ W Hz^{-1} with $z > 0.3$) with a split at $z = 1$, and no difference between the two populations was found.

5. Conclusions

The goal of this work was to construct a suitable cross-matching method and to apply it to radio catalogues at various wavebands and angular resolutions. These catalogues are: the newly released low-frequency catalogue from the LoLSS pre-release and the higher frequency catalogues from the two LoTSS data releases, TGSS, WENSS, and NVSS. We successfully developed a cross-matching method that also accounts for the different resolutions of these surveys by utilising the fitted source sizes.

In the second part of this work, a cross-matched catalogue was obtained and analysed. Concerning artefacts from LoLSS, de Gasperin et al. (2021) estimated 1055 sources (4%) to be false positives. These were found to be mostly concentrated at the edges and around bright sources. In our study we found most artefacts around bright sources and therefore used this as a criteria for artefact selection. Overall 2640 (10.5%) sources are labelled as artefacts, also including sources that may be real but are likely to have unreliable flux density measurements.

Through cross-matching, the majority (16 266 sources, 64.4%) of LoLSS sources were found to be single component, while 6341 – 25.1% – appear to be double sources. For the sources that were bright enough to be observed in NVSS, an average spectral index of $\alpha = -0.77 \pm 0.18$ was found. In contrast, for sources matched in LoLSS and LoTSS-DR2, a flatter average spectral index of $\alpha = -0.71 \pm 0.31$ was found. Comparison of the spectral slopes from LoLSS–LoTSS-DR2 with

LoTSS-DR2–NVSS indicates that the probed population of radio sources exhibits evidence of a negative spectral curvature. This could also be used to extract source information such as the age or the spectral index of the core and lobes from the single-point spectral index measurement.

For sources matched in LoLSS and LoTSS-DR2, no flux density dependence of the spectral index above the flux density $S_{54} = 181$ mJy is found. For fainter sources, typically a flatter spectrum is found. No evolution of the spectral index with redshift is observed.

Finally, along with the whole cross-matching catalogue, separate smaller catalogues are made available. These include different categories of sources that may be worth a deeper investigation and yield scientifically interesting results. These catalogues include 292 very steep spectrum sources, 9 high-redshift, steep spectrum sources, and 46 potential candidate high-redshift radio galaxies (HzRGs).

References

Afonso, J., Bizzocchi, L., Ibar, E., et al. 2011, ApJ, 743, 122
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Best, P. N., Röttgering, H. J. A., Bremer, M. N., et al. 1998, MNRAS, 301, L15
Best, P. N., Kondapally, R., Williams, W. L., et al. 2023, MNRAS, 523, 1729
Blandford, R., Meier, D., & Readhead, A. 2019, ARA&A, 57, 467
Budavári, T., & Szalay, A. S. 2008, ApJ, 679, 301
Carilli, C. L., Harris, D. E., Pentericci, L., et al. 2002, ApJ, 567, 781
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
de Gasperin, F., Intema, H. T., & Frail, D. A. 2018, MNRAS, 474, 5008
de Gasperin, F., Williams, W. L., Best, P., et al. 2021, A&A, 648, A104
de Gasperin, F., Edler, H. W., Williams, W. L., et al. 2023, A&A, 673, A165
Duncan, K. J., Sabater, J., Röttgering, H. J. A., et al. 2019, A&A, 622, A3
Duncan, K. J., Kondapally, R., Brown, M. J. I., et al. 2021, A&A, 648, A4
Fan, D., Budavári, T., Norris, R. P., & Basu, A. 2020, MNRAS, 498, 565
Fanaroff, B. L., & Riley, J. M. 1974, MNRAS, 167, 31P
Hill, G. J., Gebhardt, K., Komatsu, E., et al. 2008, in Panoramic Views of Galaxy Formation and Evolution, 399, eds. T. Kodama, T. Yamada, & K. Aoki, 115
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Intema, H. T., van Weeren, R. J., Röttgering, H. J. A., & Lal, D. V. 2011, A&A, 535, A38
Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, A&A, 598, A78
Ishwara-Chandra, C. H., Sirothia, S. K., Wadadekar, Y., Pal, S., & Windhorst, R. 2010, MNRAS, 405, 436
Kapahi, V. K., & Kulkarni, V. K. 1986, A&A, 165, 39
Kellermann, K. I., Pauliny-Toth, I. I. K., & Williams, P. J. S. 1969, ApJ, 157, 1
Line, J. L. B., Webster, R. L., Pindor, B., Mitchell, D. A., & Trott, C. M. 2017, PASA, 34, e003
Mahony, E. K., Morganti, R., Prandoni, I., et al. 2016, MNRAS, 463, 2997
Miley, G., & De Breuck, C. 2008, A&ARv, 15, 67
Mingo, B., Croston, J. H., Hardcastle, M. J., et al. 2019, MNRAS, 488, 2701
Mohan, N., & Rafferty, D. 2015, Astrophysics Source Code Library [record ascl:1502.007]
Newville, M., Stensitzki, T., Allen, D. B., et al. 2016, Astrophysics Source Code Library [record ascl:1606.014]
Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, A&ARv, 25, 2
Prandoni, I., Parma, P., Wieringa, M. H., et al. 2006, A&A, 457, 517
Rengelink, R. B., Tang, Y., de Bruyn, A. G., et al. 1997, A&AS, 124, 259
Reuland, M., Röttgering, H., van Breugel, W., & De Breuck, C. 2004, MNRAS, 353, 377
Saxena, A., Röttgering, H. J. A., Duncan, K. J., et al. 2019, MNRAS, 489, 5053
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
Sirothia, S. K., Dennefeld, M., Saikia, D. J., et al. 2009, MNRAS, 395, 269
Sutherland, W., & Saunders, W. 1992, MNRAS, 259, 413
Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, 347, Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert, 29
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Comput. Sci. Eng., 13, 22
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
Williams, W. L., Intema, H. T., & Röttgering, H. J. A. 2013, A&A, 549, A55
Williams, W. L., Hardcastle, M. J., Best, P. N., et al. 2019, A&A, 622, A2
Williams, W. L., de Gasperin, F., Hardcastle, M. J. H., et al. 2021, A&A, 655, A40
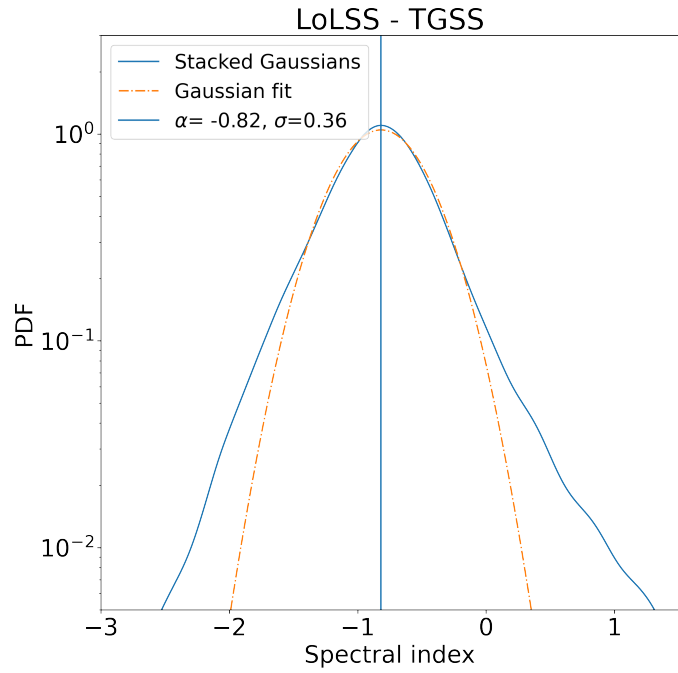
Appendix A: Additional Figures

Fig. A.1. Spectral index distribution for sources matched in LoLSS and TGSS. The orange line is a Gaussian fit with the parameters noted in the legend.
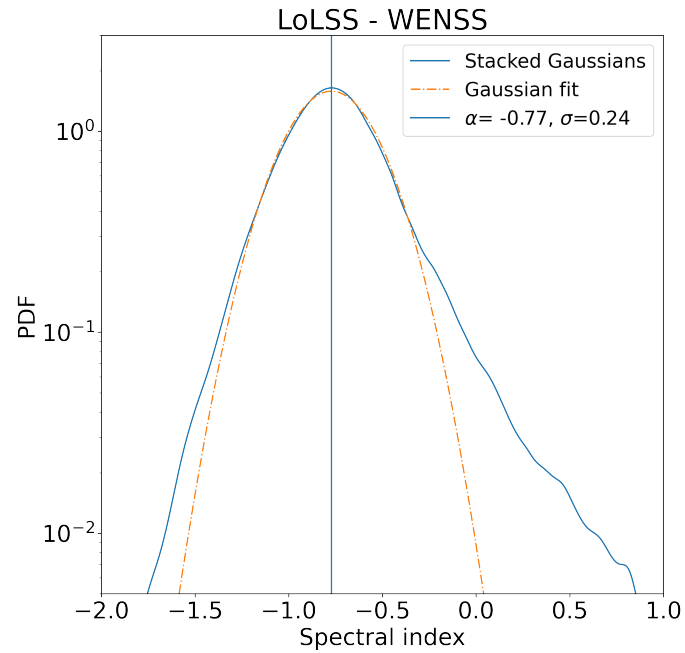

Fig. A.2. Spectral index distribution for sources matched in LoLSS and WENSS. The orange line is a Gaussian fit with the parameters noted in the legend.
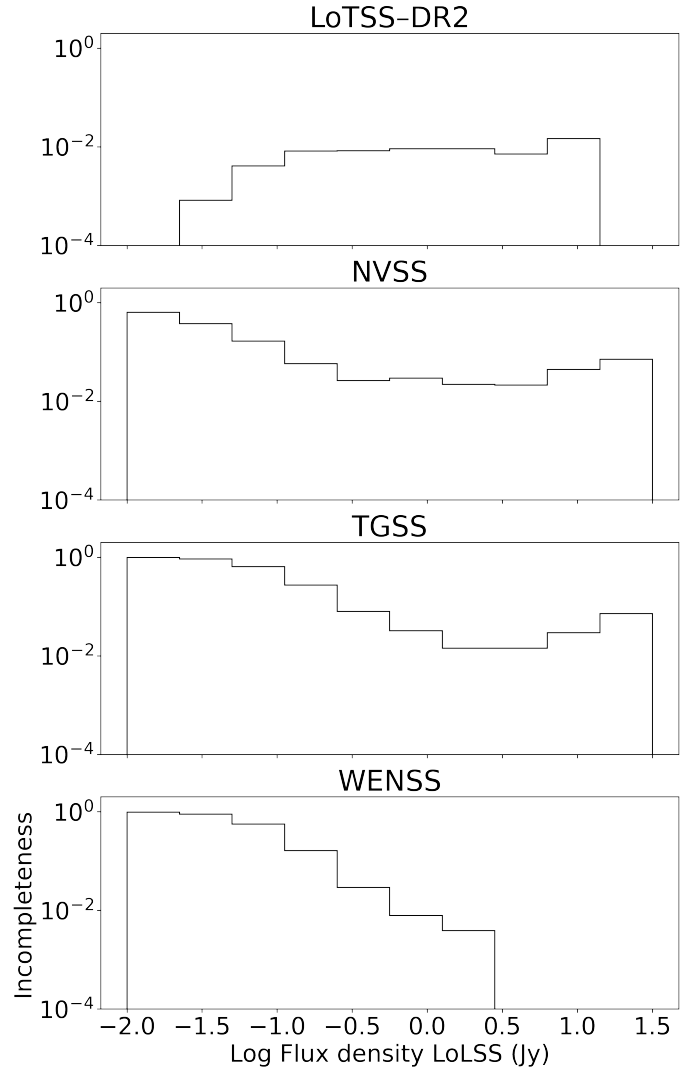

Fig. A.3. Cross-matching incompleteness as a function of flux density for the different surveys (named above each plot) matched with LoLSS.
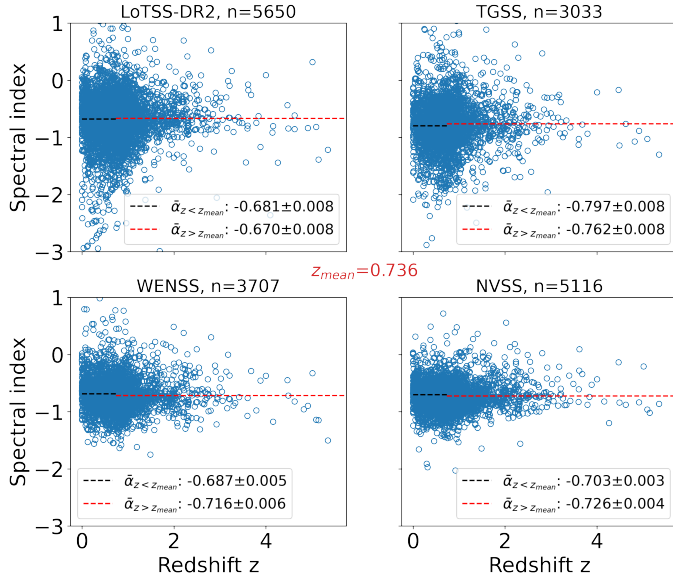
Fig. A.4. Spectral index – redshift relation. In each lower right corner is the mean spectral index in the two redshift bins above and below z$_{mean}$ (given in the centre). The black and red dotted lines show the two mean spectral indices.