

# The LOFAR LBA Sky Survey at 54 MHz

## II. First data release

F. de Gasperin<sup>1,2</sup>, H. W. Edler<sup>2</sup>, W. L. Williams<sup>3</sup>, B. Asabere<sup>4</sup>, M. Brüggen<sup>2</sup>, G. Brunetti<sup>1</sup>, J. R. Callingham<sup>3,4</sup>, C. Tasse, T. J. Dijkema<sup>4</sup>, M. J. Hardcastle<sup>5</sup>, M. Iacobelli<sup>4</sup>, A. Offringa<sup>4</sup>, M. J. Norden<sup>4</sup>, H. J. A. Röttgering<sup>3</sup>, T. Shimwell<sup>3,4</sup>, R. van Weeren<sup>3</sup>, D. J. Bomans, A. Botteon<sup>3</sup>, K. T. Chyzy, G. Miley<sup>3</sup>, B. Mingo, D. Schwarz<sup>6</sup>, and et al.

<sup>1</sup> INAF - Istituto di Radioastronomia, via P. Gobetti 101, 40129, Bologna, Italy; e-mail: fdg@ira.inaf.it

<sup>2</sup> Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112, D-21029, Hamburg, Germany

<sup>3</sup> Leiden Observatory, Leiden University, P.O.Box 9513, NL-2300 RA, Leiden, The Netherlands

<sup>4</sup> ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, NL-7990 AA, Dwingeloo, The Netherlands

<sup>5</sup> Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

<sup>6</sup> Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany

Preprint online version: October 12, 2022

### Abstract

**Context.** The Low Frequency Array (LOFAR) is the only existing radio interferometer able to observe at ultra-low frequencies (< 100 MHz) with high resolution (< 15'') and high sensitivity (< 1 mJy beam<sup>-1</sup>). To exploit these capabilities, the LOFAR Surveys Key Science Project is carrying out sensitive wide-area surveys using the two LOFAR antenna types: the LOFAR Two-metre Sky Survey (LoTSS) at 120 – 168 MHz using the LOFAR High Band Antenna and the LOFAR LBA Sky Survey (LoLSS) at 42 – 66 MHz using the LOFAR Low Band Antenna.

**Aims.** LoLSS is covering the whole northern sky above declination > 24° with a resolution of 15'' and a sensitivity of 1–2 mJy beam<sup>-1</sup> ( $1\sigma$ ) depending on declination, field properties, and observing conditions. Here we present the first data release, including a discussion of the calibration strategy and the properties of the released images and catalogues.

**Methods.** A fully automated pipeline was used to reduce the 95 fields included in this release. The data reduction procedures developed for this project have general application and are currently being used to process almost all LOFAR LBA interferometric observations. Compared to the preliminary release, direction-dependent errors have been derived and corrected for during the calibration process. This results in a typical sensitivity of 1.55 mJy beam<sup>-1</sup>, four times better than for the preliminary release, at the target resolution of 15''.

**Results.** The first release of the LOFAR LBA Sky Survey covers 650 deg<sup>2</sup> in the HETDEX spring field. The resultant data products released to the community include mosaic images (I and V Stokes) of the region, and a catalogue of 42 463 detected sources and related Gaussian components used to describe sources' morphologies. Separate catalogues for the 6 in-band frequencies of 44, 48, 52, 56, 60, and 64 MHz are also released.

**Conclusions.** The first data release of LoLSS shows that, despite the influences of the ionosphere and radio frequency interference, LOFAR can conduct large scale surveys in the frequency window 42 – 66 MHz with unprecedentedly high sensitivity and resolution. The data can be used to derive unique information on the low-frequency spectral properties of many thousands of sources that will impact several topics in extragalactic and galactic astronomy.

**Key words.** surveys – catalogs – radio continuum: general – techniques: image processing

## 1. Introduction

The LOw Frequency ARray (LOFAR; van Haarlem et al. 2013) is the world's largest and most sensitive radio telescope in the radio window covering 10 – 240 MHz (wavelength  $\lambda = 1$  – 30 metres). Unlike higher frequency "dish" radio interferometric arrays, the basic antenna elements of LOFAR are wide-field dipoles. There are two separate dipole types at each LOFAR station, covering the regions above and below 100 MHz. The Low Band Antenna (LBA) operates at 10–90 MHz and the High Band Antenna (HBA) covers the 120 – 240 MHz band. The intermediate region is dominated by radio frequency interference (RFI) due to FM radio stations and is not accessible for observations.

The LOFAR LBA can reach frequencies close to the ionospheric plasma cutoff frequency, the ultimate boundary for ground-based radio astronomy. The history of astronomy shows that one of the most effective ways of making fundamental dis-

coveries is to open up new regions of the electromagnetic spectrum to observation. As the only array capable of operating below 100 MHz over baselines of tens to thousands of kilometers, the International LOFAR Telescope (ILT) is a unique instrument for exploring the low frequency radio Universe at high angular resolution.

The relativistic electrons which the LBA observes are less energetic and in general older than those observed at higher frequencies. Studying these "fossil", ultra-steep spectrum sources at the high resolution of the ILT can provide new information about fundamental topics, ranging from the formation and evolution of the first massive galaxies and protoclusters to the detection of radio emission from Jupiter-like exoplanets (see de Gasperin et al. 2021, for a summary of the science cases). Moreover, with the unprecedented sensitivity of LOFAR in this hitherto neglected spectral region, there is the possibility of serendipitous discoveries based on the large number of faint

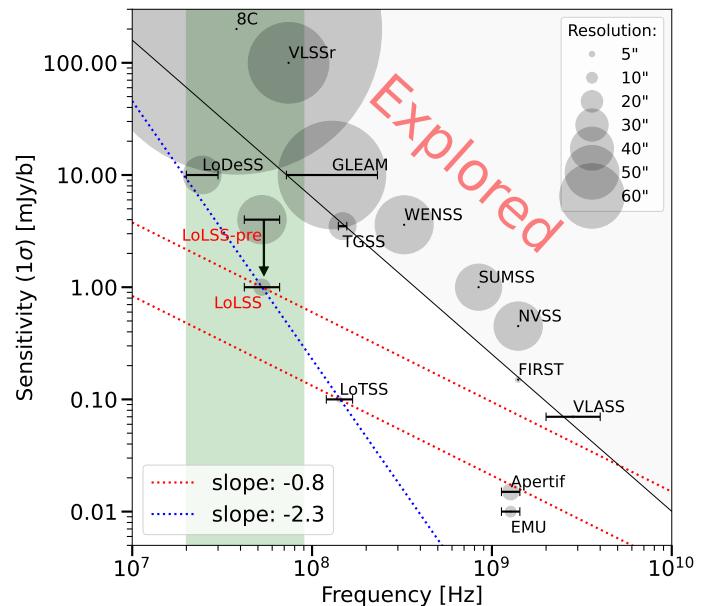
ultra-steep spectrum sources that will be detected in the LOFAR large-sky low-frequency surveys.

The high survey speed of modern radio telescopes, combined with increased complexity in dealing with extremely large data rates, shifted the observing strategy of several radio observatories in favour of large scale surveys (Norris 2017, see Fig. 1)<sup>1</sup>. The LOFAR survey team responded to this challenge by setting up a multi-tier strategy aimed at covering a large portion of the northern sky at both HBA and LBA frequencies:

- The LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al. 2019, 2022) is a survey of the entire northern sky in the frequency range 120 – 168 MHz. The survey reaches a sensitivity of about  $80 \mu\text{Jy beam}^{-1}$  at a resolution of  $\approx 6''$ . Several deeper fields with much longer exposures were also included in the LoTSS survey within the frame of the LoTSS deep field program (see e.g. Tasse et al. 2021; Williams et al. 2021; Sabater et al. 2021). Although LoTSS does not use them, the survey data include international stations that allow imaging at  $0.3''$  resolution (Morabito et al. 2022). A re-imaged version of the LoTSS survey data including these stations is planned and the first high-resolution pilot study has recently been released (Sweijen et al. 2022).
- The LOFAR LBA Sky Survey (LoLSS; de Gasperin et al. 2021) was designed as the LBA counterpart of LoTSS. It aims at covering the northern sky at  $\text{Dec} > 24^\circ$ , with a sensitivity of  $\approx 1 \text{ mJy beam}^{-1}$  and a resolution of  $\approx 15''$ . Several selected deep fields were also observed within the survey framework, to achieve increased sensitivity and reaching lower frequencies (see e.g. de Gasperin et al. 2020b; Williams et al. 2021).
- The LOFAR Decameter Sky Survey (LoDeSS) is an experimental survey designed to cover the northern sky at the lowest frequency range of the LBA system (15 – 30 MHz).

Due to the low operating frequency and wide field of view, the main challenge that needs to be overcome to produce reliable LOFAR images is the removal of ionospheric-induced systematic effects (Intema et al. 2009; Mevius et al. 2016; de Gasperin et al. 2018). Because they vary across the field of view, these are direction-dependent effects. Therefore, self-calibration strategies previously adopted for correcting small-field interferometric data cannot be used to simply calibrate LOFAR interferometric data sets. For this reason more sophisticated tools needed to be developed or improved to reduce LOFAR data, such as WSClean (Offringa et al. 2014) and DDFacet (Tasse et al. 2018) for imaging, DP3 (van Diepen et al. 2018) and KillMS (Tasse 2014) for calibration, and LoSoTo for solution manipulation (de Gasperin et al. 2019). The availability of these new tools required the adoption of new and specialised analysis strategies. The high sensitivity of HBA observations allowed for the first development of a working strategy based on facet calibration and imaging (van Weeren et al. 2016), although LoTSS ended up using a faster approach based on the simultaneous derivation of solutions using KillMS and DDFacet (Tasse et al. 2021). In the LBA low signal-to-noise ratio regime, the facet calibration approach delivers stable results, so variations of this serial-

<sup>1</sup> Examples of these are: the Evolutionary Map of the Universe (EMU; Norris et al. 2011, 2021), the Polarization Sky Survey of the Universe's Magnetism (POSSUM; Gaensler et al. 2010), the APERture Tile In Focus surveys (APERTIF; Hessin in prep.), the GaLactic and Extragalactic All-Sky MWA-eXtended survey (GLEAM-X Hurley-Walker et al. 2017, 2022), and the Karl G. Jansky Very Large Array Sky Survey (VLASS; Lacy et al. 2020).



**Figure 1.** Comparison of sensitivity for a number of completed and ongoing wide-area radio surveys. The diameters of the grey circles are proportional to the survey resolution as shown in the top right corner. The green region shows the frequency range covered by the LOFAR LBA system. Data presented in this paper are labelled as ‘LoLSS’, whilst the preliminary release of the LoLSS survey is labelled as ‘LoLSS-pre’. For sources with a very steep spectral index ( $\alpha \lesssim -2.3$ ), LoLSS is the most sensitive survey currently available. References: 8C (Rees 1990); GLEAM (GaLactic and Extragalactic All-sky Murchison Widefield Array survey; Hurley-Walker et al. 2017); TGSS ADR1 (TIFR GMRT Sky Survey - Alternative Data Release 1; Intema et al. 2017); VLSSr (VLA Low-frequency Sky Survey redux; Lane et al. 2014); FIRST (Faint Images of the Radio Sky at Twenty Centimetres; Becker et al. 1995); NVSS (1.4 GHz NRAO VLA Sky Survey; Condon et al. 1998); WENSS (The Westerbork Northern Sky Survey; Rengelink et al. 1997); SUMSS (Sydney University Molonglo Sky Survey; Bock et al. 1999); Apertif (Adams et al. in prep.); EMU (Evolutionary Map of the Universe Norris et al. 2011); VLASS (VLA Sky Survey; Lacy et al. 2020); LoTSS (LOFAR Two-metre Sky Survey; Shimwell et al. 2017).

solving approach were implemented in producing LoLSS (e.g. Edler et al. 2022, ; see also Sec. 3).

The paper is organised as follows: in Sec. 2 we describe the main observational parameters of the survey and its current status. The information included in that section is an updated version of what was initially presented in Sec. 3 of de Gasperin et al. (2021). In Sec. 3 we give an update on the data reduction procedure initially presented in de Gasperin et al. (2019, 2020a, 2021). In Sec. 4 we describe the LOFAR LBA Sky Survey First Release and in Sec. 5 we give quantitative estimation of the survey properties. A summary of the paper is in Sec. 7. The spectral index  $\alpha$  is defined as  $S_\nu \propto \nu^\alpha$ , with  $S_\nu$  the flux density.

## 2. The LOFAR LBA Sky Survey

To cover the entire sky north of declination  $> 24^\circ$ , the survey was divided into 1 889 pointings that are centered on the same coordinates to those of the corresponding LoTSS pointings to simplify future combined analysis. Each observation used the

Number of pointings	1 889
Separation of pointings	2.58°
Integration time (per pointing)	3 h (8 h for HETDEX)
Frequency range	42 – 66 MHz
Array configuration	LBA OUTER / LBA SPARSE
Angular resolution	~ 15''
Noise level	~ 1 – 2 mJy beam <sup>-1</sup>
Time resolution	1 s
– after averaging	4 s
Frequency resolution	3.052 kHz
– after averaging	24.414 or 48.828 kHz

**Table 1.** LoLSS observational setup.

LOFAR LBA multi-beam capability to observe simultaneously one calibrator chosen among 3C 196, 3C 295, and 3C 380 and three target fields. All observations were carried out in exposure blocks of 1 h, so that each field was observed multiple times on different days to reduce the probability that a field could not be used due to bad ionospheric conditions. The various observations of the same field were scheduled to maximise the LST coverage when the elevation of the target was above 60°. We set such a limit on the elevation in order to minimise both the suppression of the signal due to the dipole beam and the path length through the ionospheric layer that is pierced by the incoming radio waves.

Multi-beam capability comes at the cost of a reduced bandwidth per beam. The total bandwidth that LOFAR can process is 96 MHz. In our case, each beam has allocated 24 MHz, i.e. 122 sub-bands (SB) of 0.195 MHz each, in order to cover the frequency range 42 – 66 MHz. This is the most sensitive frequency range of the LBA system once the sky temperature and the dipole bandpass are taken into consideration (van Haarlem et al. 2013). At the station level, a 30-MHz high-pass filter is applied to the signal path to suppress radio frequency interference (RFI) reflected by low ionospheric layers at < 20 MHz.

There are three operating modes for LOFAR LBA:

- LBA.INNER, for which the innermost 48 dipoles of each station are used. This mode gives the largest beam size at the cost of a reduced sensitivity. The calibration of the individual station dipoles in this mode is less effective than for the other modes due to mutual coupling and increased response to Galactic emission during the station calibration procedure. The effective size of the station in this mode is 32 m, which corresponds to a primary beam full width at half maximum (FWHM) of 10° at 54 MHz.
- LBA.OUTER, where only the outermost 48 dipoles of each station are used. This mode minimises the coupling between dipoles and reduces the beam size. The effective size of the station in this mode is 84 m with a primary beam FWHM of 3.8° at 54 MHz.
- LBA.SPARSE (ODD or EVEN), where half of the dipoles, distributed across the station, are used. This mode results in an intermediate performance between LBA.INNER and LBA.OUTER, with a suppression of the magnitude of the side-lobes compared to the latter. The effective size of the station in this mode is about 65 m, which results in a primary beam FWHM of 5.5° at 54 MHz.

Planned updates of LOFAR will enable the use of all LBA dipoles simultaneously (LBA.ALL mode), effectively doubling the station collecting area with a FWHM similar to LBA.SPARSE. Until the end of 2020, LoLSS observations were

carried out using the LBA\_OUTER mode. This includes all observations of the first release discussed in this paper. After commissioning LBA\_SPARSE, we switched to that mode to increase the FoV and therefore the survey speed while reducing the problem of unmodelled flux spilling into the side-lobes of the primary beam.

The pointing schemes used for LoTSS and LoLSS were the same and followed a spiral pattern starting from the north celestial pole, with pointing positions determined using the Saff & Kuijlaars (1997) algorithm and separated by 2.58°. With such a pointing scheme, LoLSS required 1889 pointings. Assuming circular beams and LBA\_OUTER, this separation provides a distance between pointing centres of FWHM /  $\sqrt{2}$  at the highest survey frequencies and better than FWHM / 2 at the lowest. With LBA\_SPARSE, the overlap of the fields is even greater.

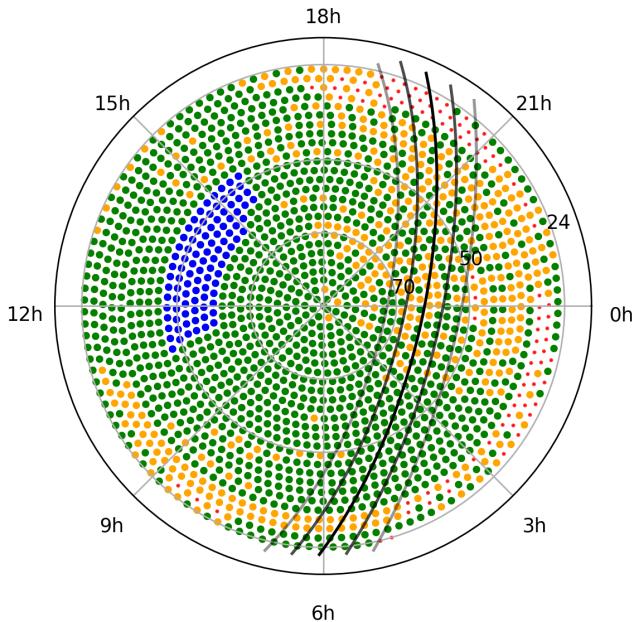
At present, the International LOFAR Telescope has 24 core stations (CS), 14 remote stations (RS), 14 international stations (IS) and two more under construction. The CSs are spread across a region of radius ~ 2 km and provide 276 short baselines. The RS are located within 70 km from the core and their longest baseline provides a resolution of ~ 15'' at 54 MHz. LoLSS makes use of CS and RS, whilst IS data were not recorded to keep the size of the data set manageable.

The aim of LoLSS is to reach a sensitivity of ~ 1 mJy beam<sup>-1</sup>. With the LBA system, this requires ≈ 8 hrs of integration time at the optimum declination. However, because of the strong overlap of LBA\_SPARSE fields, this integration time can be relaxed to about 3 hours per field. However, the final noise is limited by ionospheric conditions and dynamic range. Experiments with comparable total integration times indicate that the rms noise ranges between 1 and 2 mJy beam<sup>-1</sup> (e.g. de Gasperin et al. 2020a). In the preliminary release, in which the direction-dependent errors were not corrected, the noise ranged between 4 and 5 mJy beam<sup>-1</sup> (de Gasperin et al. 2021), compared with a median noise of 1.6 mJy beam<sup>-1</sup> for this release (see Sec. 5.2).

The time and frequency resolution were chosen to balance the data size and the effect of time/frequency smearing at the edges of the field of view. A time resolution of a few seconds is also necessary to track fast evolving ionospheric variations. Data were initially recorded at 1 s and 3.052 kHz resolution to properly identify fast and narrow band RFI (Offringa et al. 2010). The high resolution is also needed to remove bright sources (Cygnus A and Cassiopeia A) from the far side lobes (de Gasperin et al. 2019). Data were then averaged to 2 s and 48.828 kHz for LBA\_OUTER observations and to 4 s and 24.414 kHz for LBA\_SPARSE. The decrease in frequency averaging was applied so as to keep the frequency smearing at to less than 5% at the edge of the Field of View (FoV). Time smearing is less of a problem (< 1%) and an averaging time of 4 s was selected to keep the data volume low. A summary of the observational setup is given in Table 1.

## 2.1. Survey status

The initial observations of LoLSS, those covering the region of the Hobby-Eberly Telescope Dark Energy Experiment (HETDEX Hill et al. 2008) spring field, were obtained in 2017 and 2019. After proving the feasibility of the survey, the bulk of the observations started in 2020 and are expected to be completed by the end of 2022. In Fig. 2, we give an overview of the observing status at of September 2022. At the time of writing 1 396 (73.9%) pointings have been observed with at least 3



**Figure 2.** Current and planned sky coverage of LoLSS. Each dot represents a pointing of the full survey. The pointings of the region presented in this paper are coloured in blue. Green and yellow dots represent pointings with archived observations (3 hrs for green, 1 or 2 hrs for yellow). Small red dots will be observed after 2022. Solid lines show the position of the Galactic plane with Galactic latitude:  $-10^\circ, -5^\circ, 0^\circ, 5^\circ, +10^\circ$ .

hrs of data, 272 (14.4%) have been observed for 2 hrs and 132 (7%) for 1 hrs, while 89 (4.7%) pointings still have no observations. Pointings that were unusable due to hardware issues or that were taken during particularly bad ionospheric conditions (about 15%) were identified. They are not included as valid data in the above summary and re-observations are on-going.

### 3. Data reduction

The images presented in the first release of LoLSS were obtained through a full re-run of the calibration process, neglecting the partially calibrated data of the preliminary release. The data reduction was carried out at the Hamburg Observatory using five computing nodes with 32 cores each, 256 GB of memory and 20 TB of storage in local drives. A centralised database kept track of the various jobs and coordinated the nodes. The computation was carried out in an ad-hoc environment based on Ubuntu 20.4 for which a Docker container is available<sup>2</sup>.

For the data reduction, an automated Pipeline for LOFAR LBA (PiLL) that is described in de Gasperin et al. (2019), for the calibrator part, and de Gasperin et al. (2020a) for the target part, was employed. The direction-dependent portion was completely rewritten compared with that used to produce the first publication of a thermal-noise limited LOFAR LBA image presented in de Gasperin et al. (2020a). The main differences are described below.

<sup>2</sup> See <https://github.com/revoltek/LiLF>.

### 3.1. Pre-processing pipeline

Immediately following the observations, the following steps were carried out by the LOFAR Observatory at ASTRON: 1. Flagging of RFI with AOflagger (Offringa et al. 2012), 2. Subtraction (“demixing”) of the ultra-bright sources Cygnus A and Cassiopeia A using the algorithm described in van der Tol et al. (2007) and 3. Averaging of the data to the relevant time and frequency resolution. As described above, these were 4 s and 24.414 or 48.828 kHz per channel depending on the observing mode, with the former used with LBA\_SPARSE and the latter with LBA\_OUTER. The data were then ingested into the LOFAR Long Term Archive (Renting & Holties 2012).

### 3.2. Calibrator pipeline

The calibrator pipeline was then applied to data from the beam pointing at the calibrator source. The calibrator was chosen from the bright compact sources 3C 196, 3C 295, and 3C 380, depending on which was at higher elevation during the observation. Since our observations did not include data from international stations, the resolution is limited to  $15''$ . We therefore could use simple models for the calibrators. 3C 196 was parameterised by four point sources and 3C 295 by two point sources. However, 3C 380 is well-resolved even at  $15''$  and a more detailed model for this source was obtained after self-calibrating it using LOFAR LBA data including international stations (Groeneveld et al. 2022). All models were rescaled so their total flux densities matched the Scaife & Heald (2012) flux density scale.

The calibrator pipeline then isolated the polarisation misalignment introduced by the station calibration table. This is an artificial delay between the two polarisation components that is constant in time and varies with station. The Faraday rotation was removed by converting the data set to circular polarisation and measuring the separation between right and left circularly polarised polarisation, which has a frequency dependency  $\propto 1/\nu^2$ . This approach has the advantage that it does not depend on the sky model. The Faraday rotation was derived solely by measuring the relative misalignment between the two polarisations (Groeneveld et al. in prep). Finally, a time-independent bandpass was derived together with a time and frequency dependent scalar phase that encodes differential station delays due to misaligned clocks and differential ionospheric delays. The two effects are theoretically separable using their different frequency dependency (Mevius et al. 2016). However, in the low signal-to-noise ratio regime of LOFAR LBA strong cross-contamination was found to frequently persist.

### 3.3. Target pipeline: direction-independent

Data from each target beam were corrected using solutions derived from the simultaneous calibrator observation. The solutions applied were: polarisation alignment, bandpass and scalar phase calibration. Here the polarisation alignment, bandpass and the clock part of the phases are instrumental systematic effects that are direction-independent, and are therefore fully corrected for using the calibrator solutions. However, the ionospheric part of the scalar phases corrupts the data with the effect of the ionosphere in the calibrator direction. Finally, the theoretical element beam of LOFAR LBA dipoles was applied and the data were combined into a single measurement set.

At this stage the direction-independent self-calibration can begin. The data were initially calibrated using a source model

derived from the combination of data from TGSS (Intema et al. 2017), NVSS (Condon et al. 1998), WENSS (Rengelink et al. 1997), and VLSS (Lane et al. 2014). This enabled the spectral index to be estimated and the flux density of each source to be extrapolated to the LBA frequency range. Sources with flux densities smaller than 1 Jy at 60 MHz were discarded from the model to speed up the prediction of the visibilities.

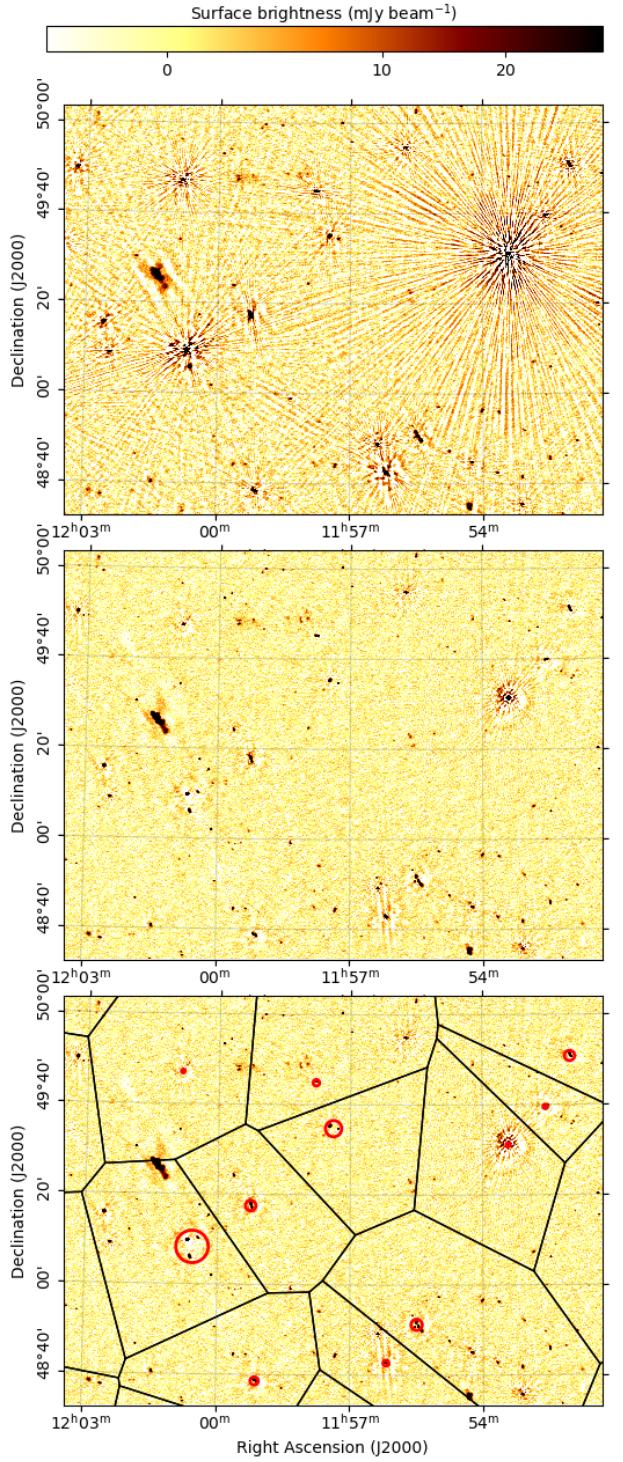
During self-calibration, corrections for three effects were derived. The first is a direction-independent Faraday rotation, derived using the same procedure described for the calibrator field. Secondly, the direction-independent total electron content (TEC) of the ionosphere was estimated for each station. This was done using slow-varying (1 minute) solutions for the stations within 10 km from the central core of LOFAR (the ‘Superterp’) and fast-varying (4 seconds) solutions for the other stations. Thirdly, station-independent amplitude solutions were obtained on all core stations in an attempt to derive the second-order correction on the analytical beam model. These solutions were then applied to data from the core and remote stations.

During the process, a wide-field, low-resolution image was produced. This image was used to identify strong sources present outside the first null. These are particularly problematic in LBA\_OUTER mode for which the first side-lobe is prominent. Sources found outside the main lobe were then subtracted from the visibilities. An image covering the entire field of view was finally produced. The images produced at this stage of the process were the ones presented in the preliminary release (de Gasperin et al. 2021).

### 3.4. Target pipeline: direction-dependent

At this point, differential (w.r.t. the average TEC) direction-dependent ionospheric errors caused by local TEC variations across the field and second-order direction-dependent beam errors need to be addressed. The pipeline identifies suitable direction-dependent calibrators using the image produced at the end of the previous step. These calibrators are grouped using machine learning techniques based on mean shift clustering (Cheng 1995), and finally selected based on a minimum apparent flux density, with  $S_\nu > 1.0$  Jy at  $\nu = 60$  MHz assuming a spectral index  $\alpha = -0.8$  and avoiding clusters of sources for which at least one source that contributes with more than 20% to the cluster flux density is extended with an integrated to peak flux density ratio  $S_I/S_P > 4$  (see Fig. 3 for examples).

The data set was averaged to a time resolution of 8 s to decrease the data volume. In order of flux density, for each calibrator all other sources were subtracted from the data set corrupting the model with the in-hand direction-independent error estimations. The selected direction-dependent calibrator is then added back to the visibilities and the data set is phase-shifted in its direction and averaged to channels of  $2 \times 192$  kHz and to 16 or 32 seconds if the calibrator flux density is lower than 10 or 4 Jy, respectively. This averaging is possible as most of the ionospheric effects are already removed at this stage and frequency/time smearing is negligible in the small FoV surrounding the calibrators. The differential beam effect is applied to the averaged data set and solutions are derived in several cycles of self-calibration on the calibrator sources. In each cycle, fast scalar phases are derived for remote stations, with solutions forced to be smooth in frequency through the solver. The loop continues as long as the image noise and dynamic range improve. For sources with an apparent flux density  $> 5$  Jy at 54 MHz, from the fourth cycle amplitude solutions are derived in two steps. Firstly, slow amplitude solutions for six frequency chunks are derived forcing



**Figure 3.** A region of the LBA survey with only direction-independent error corrections applied (top) and with direction-dependent error corrections added (middle and bottom). In the bottom panel we identify the direction-dependent calibrators used (sources within red circles) and the relative facets. Inclusion of direction-dependent error corrections results in a sensitivity improvement by a factor  $\sim 2$ .

all stations to have the same solution. These amplitude variations reflect imperfect element beam modelling that are expected to be the same for each station. Secondly, even slower solutions are derived for each antenna forcing the solver to keep the solutions

smooth in frequency. All amplitude solutions are normalised so as not to affect the overall flux scale. During the self-cal cycles the calibration steps are carried out using DP3 (van Diepen et al. 2018), the imaging using WSClean (Offringa et al. 2014; Offringa & Smirnov 2017), and the handling of the solutions using LoSoTo (de Gasperin et al. 2019).

Solutions from direction-dependent calibrators that resulted in a lowering of the local noise and increase in the dynamic range are preserved and combined in an imaging call using DDFacet (Tasse et al. 2018). The imager applies the solutions to each facet without attempting a smooth transition between them. Together with the solutions, a time dependent model of the beam is also applied to each facet. A mask is also generated from the best available image and passed to DDFacet to include all extended emission in the deconvolution process, providing a good fidelity (see Fig. 4). An example of the final result of the imaging step is visible in Fig. 3. Finally, the entire direction-independent calibration process is repeated but starting from the improved model and the direction-dependent solutions for the initial subtraction. At the end of the second main cycle, DDFacet produces I- and V-Stokes images as well as a 6-channel cubes which will be used to derive in-band information.

## 4. Public data release

Here we describe the data products released to the community. They consist of images and source catalogues. The present products cover a region around the HETDEX spring field (RA: 11 h to 16 h and Dec: 45° to 62°; area: 650 deg<sup>2</sup>), which is 5% of the sky area to be covered by the LoLSS survey.

### 4.1. Mosaic images

We produced one mosaic of 3.3° × 3.3° around each of the 95 pointing centres of the HETDEX region. Mosaics were made by taking all neighbouring pointing images with centre distances within 6°, and reprojecting them into the frame of the central pointing. The images were then corrected for any global astrometric shift derived by cross matching isolated and compact sources against the FIRST catalogue. The distribution of the corrections applied has a standard deviation of 2'' for both RA and Dec. Finally, the reprojected images' data were averaged to create a single output image using weights obtained by combining the beam attenuation and the average rms noise for each input image.

The same process was repeated for each channel of the final image cube, producing mosaics at 44, 48, 52, 56, 60, and 64 MHz for which only the catalogue will be provided. For the combined image at 54 MHz we also produced and released Stokes V mosaics. Together with the 95 mosaics at 54 MHz, we also prepared two large mosaics covering the east and west part of the HETDEX region: these are shown in Fig. 5 and 6. Within each figure, a small panel shows a blowup of a region of the mosaic to illustrate the data quality.

Finally, for each pointing a source-subtracted low-resolution image was produced starting from direction-independent error corrected data. In this case, the direction-dependent error correction is less relevant as only short baselines were preserved (max 5 km). The data set weights were Gaussian tapered to reach a resolution of 180''. The mosaic image is presented in Fig. 7. The image shows the presence of large angular scale structures that were independently detected also in LoTSS data and are likely of Galactic origin (Oei et al. in prep.). We caution

Catalogue	Central frequency	Number of source	Number of gaussian components
Combined	54 MHz	42 463	53 377
Chan 00	44 MHz	22 689	27 067
Chan 01	48 MHz	24 511	29 213
Chan 02	52 MHz	25 488	30 443
Chan 03	56 MHz	25 456	30 363
Chan 04	60 MHz	23 894	28 380
Chan 05	64 MHz	21 659	25 636

**Table 2.** Catalogue frequencies and numbers of entries.

the reader that the surface brightness of such structures might be strongly biased due to their large extension combined with missing short baselines. Because of the presence of the extended emission in all regions of the map, it is difficult to estimate the background rms noise. Measuring it in a few regions where the emission appears less dominant we find an rms noise of about 10 – 15 mJy beam<sup>-1</sup>. Images described in this section are available at [www.lofar-surveys.org/lolss.html](http://www.lofar-surveys.org/lolss.html).

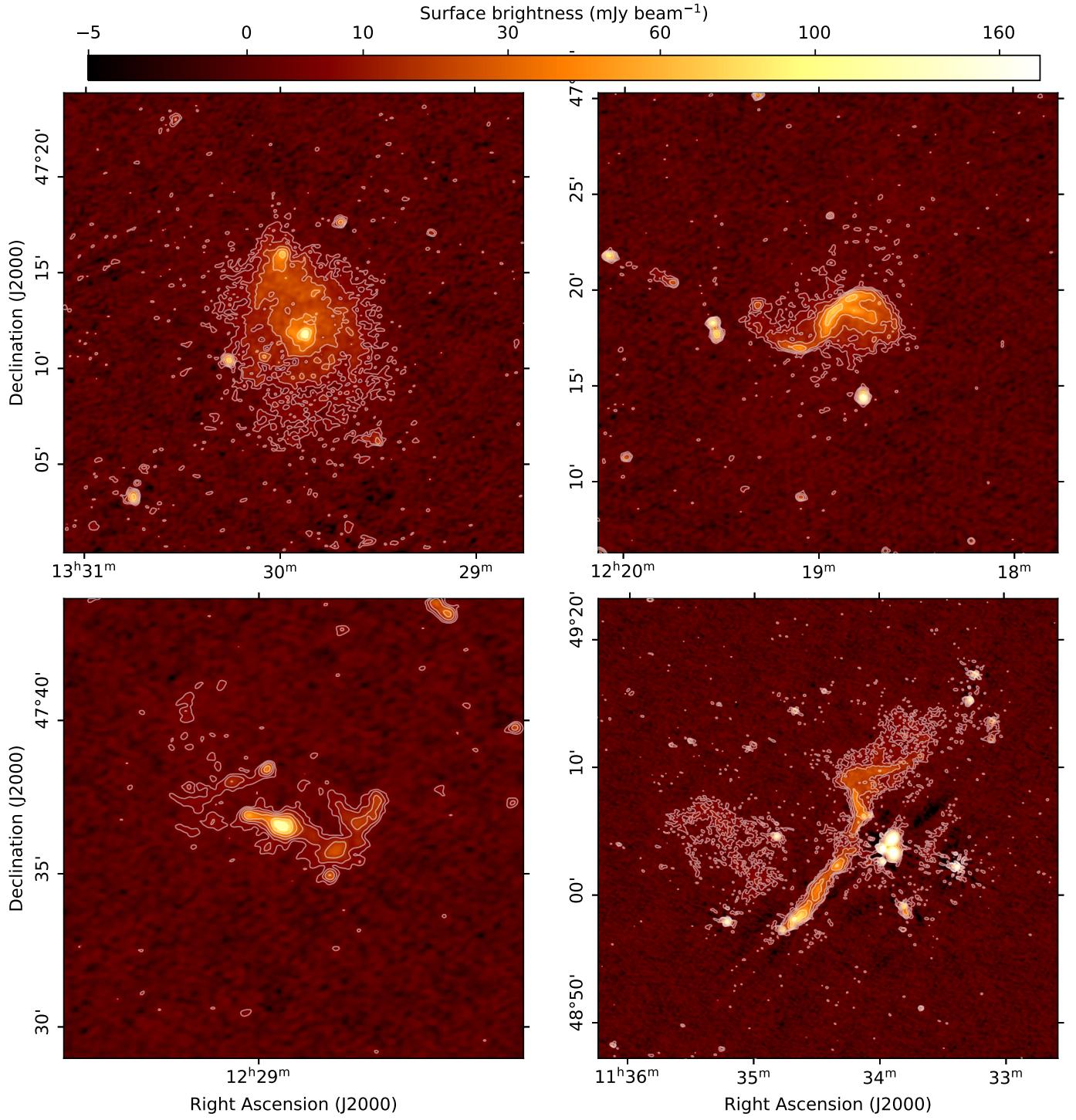
### 4.2. Source catalogues

Source catalogues were generated with PyBDSF (Mohan & Rafferty 2015). The source extraction procedure used a 4σ detection threshold to find islands of emission with a 5σ threshold on the brightest pixel. The significance of the emission in each pixel is measured against the local rms noise. To reduce the number of false positive detections, we use an adaptive rms box size that shrinks around sources with pixels above 50 times the local rms. We saved both the Gaussian components used to fit the source shape and the source catalogue. Rms images were also saved to be used in subsequent analysis. Each run of the source finder was performed on a 3.3° × 3.3° mosaic; this produced 95 Gaussian components and 95 source catalogues with significant overlap. The combined source catalogue was then produced by merging the 95 separate source catalogues and for each retaining only the sources closest to the centre of that particular mosaic image. The combined Gaussian components catalogue retained only Gaussian components associated with sources in the combined source catalogue. The combined source catalogue contains 42 463 entries and the combined Gaussian components catalogue contains 53 377 components. The catalogues are available at [www.lofar-surveys.org/lolss.html](http://www.lofar-surveys.org/lolss.html).

The same process described above was repeated for each channel image derived from the final image cubes, therefore providing 6 additional source and Gaussian components catalogues. They have the same structure as the combined catalogue but because they are derived from images made with a sixth of the bandwidth, they contain fewer sources (see Table 2).

#### 4.2.1. Completeness and false positives

To evaluate the catalogue completeness, we adopted the procedure outlined by Heald et al. (2015). For this purpose, we used the residual mosaic images created by PyBDSF after subtracting the Gaussian components derived during the source detection. These images include information about the distribution of the rms noise and artifacts, and can therefore be used to inject artificial sources into the mosaics in order to assess to what level they can be retrieved by the source finder. To study this, we injected a population of 10 000 point sources into each empty image, randomly distributed, with flux densities ranging between 3 mJy

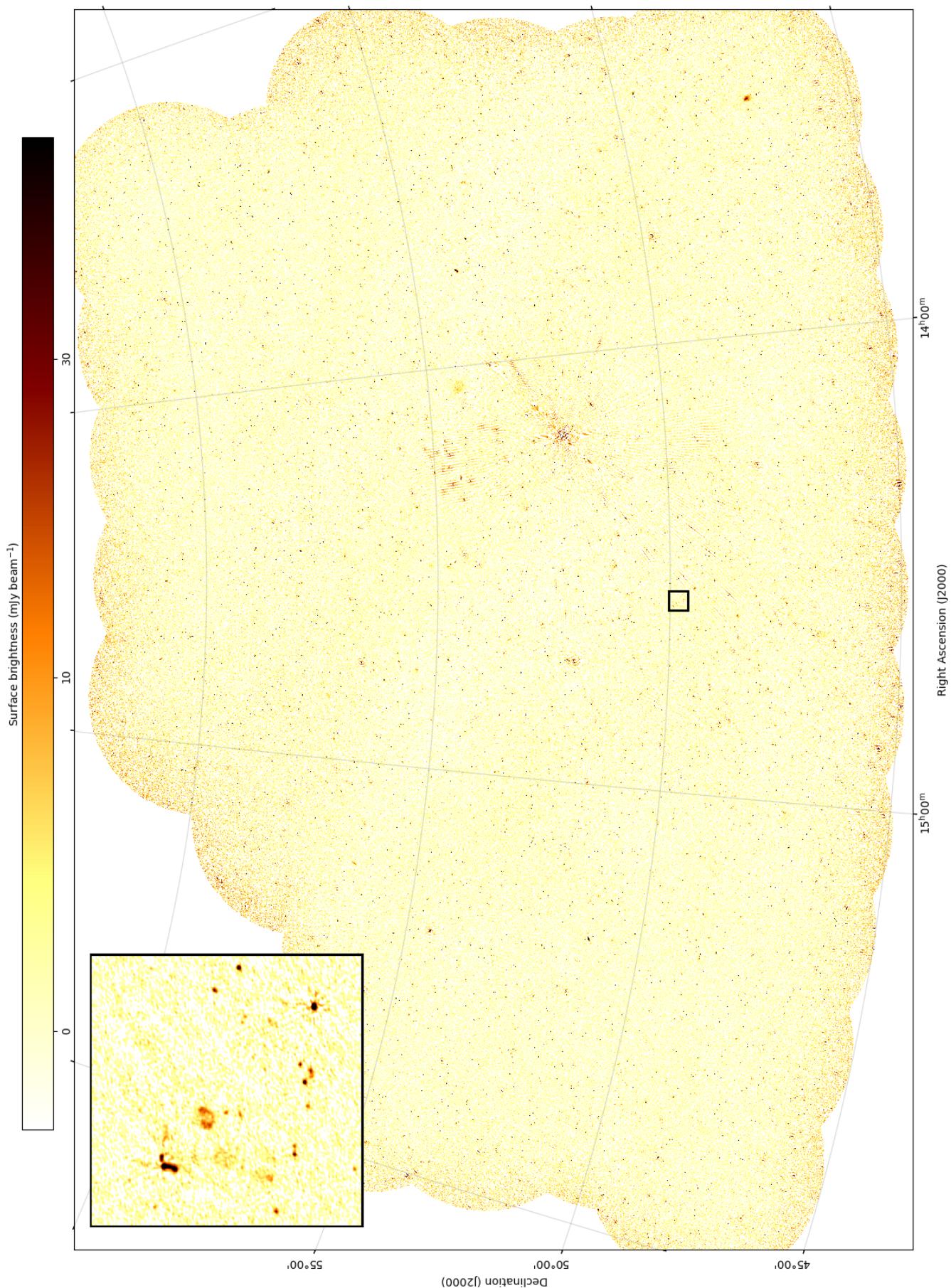


**Figure 4.** Some examples of extended sources in the data release published with this paper. From top-left to bottom-right M 51, M 106, the galaxy cluster Abell 1550, and the complex blend of emission coming from both AGN activity and diffuse sources in the intra-cluster medium of Abell 1314. Contours are at  $3\sigma, 8, 15, 30, 50, 100$  mJy beam $^{-1}$ , with  $\sigma = 1.5$  mJy beam $^{-1}$ .

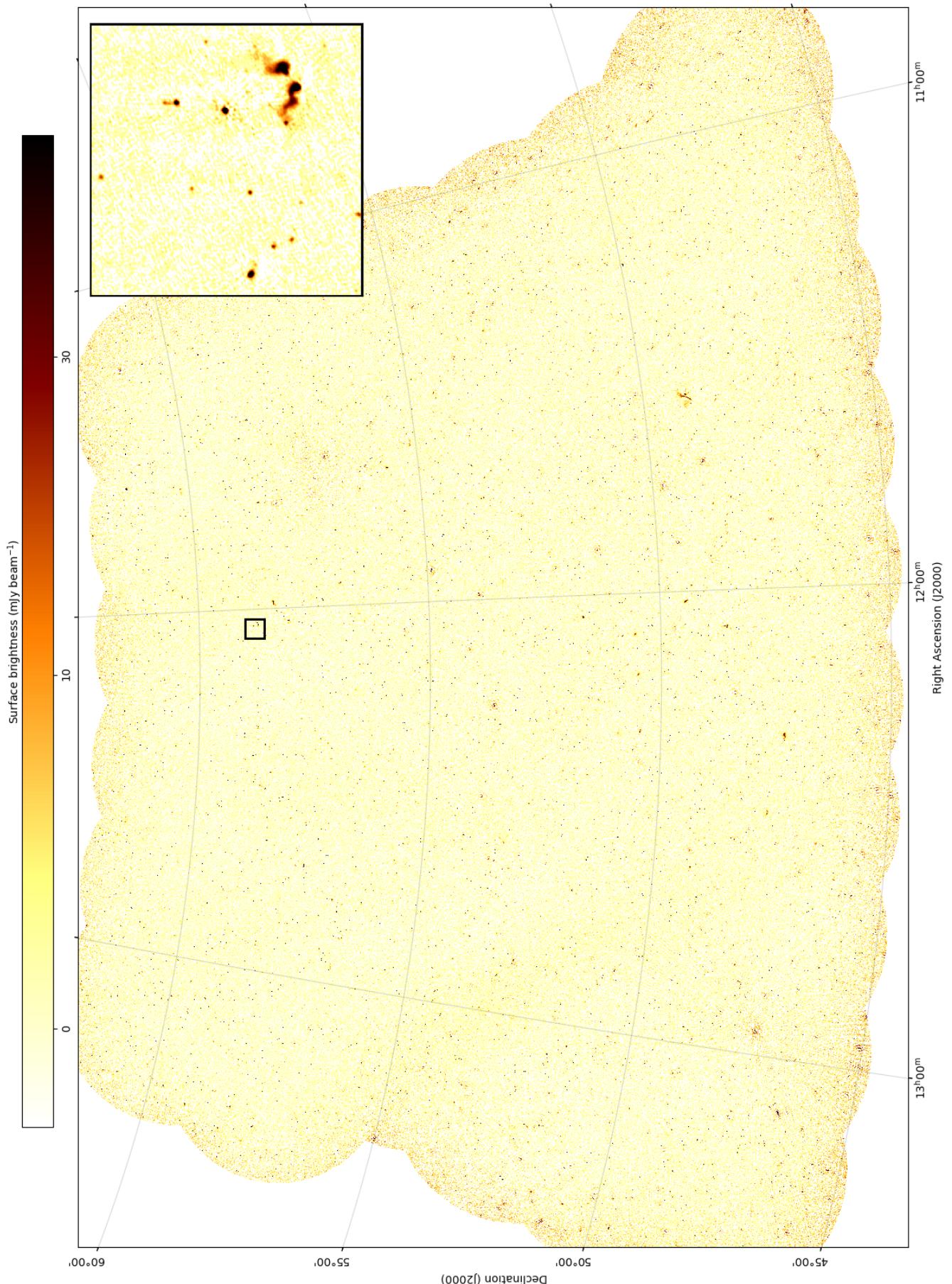
and 10 Jy, and following a number count power-law distribution of  $dN/dS \propto S^{-1.6}$  (Wilman et al. 2008). We then attempted to detect these sources using PyBDSF with the same parameters used in compiling the real catalogue. The process was then repeated 20 times to decrease sample noise.

A source was accepted as detected if it was found to be within  $45''$  (3 times the synthesised FWHM beam) of its input position and whose measured flux density was within ten times

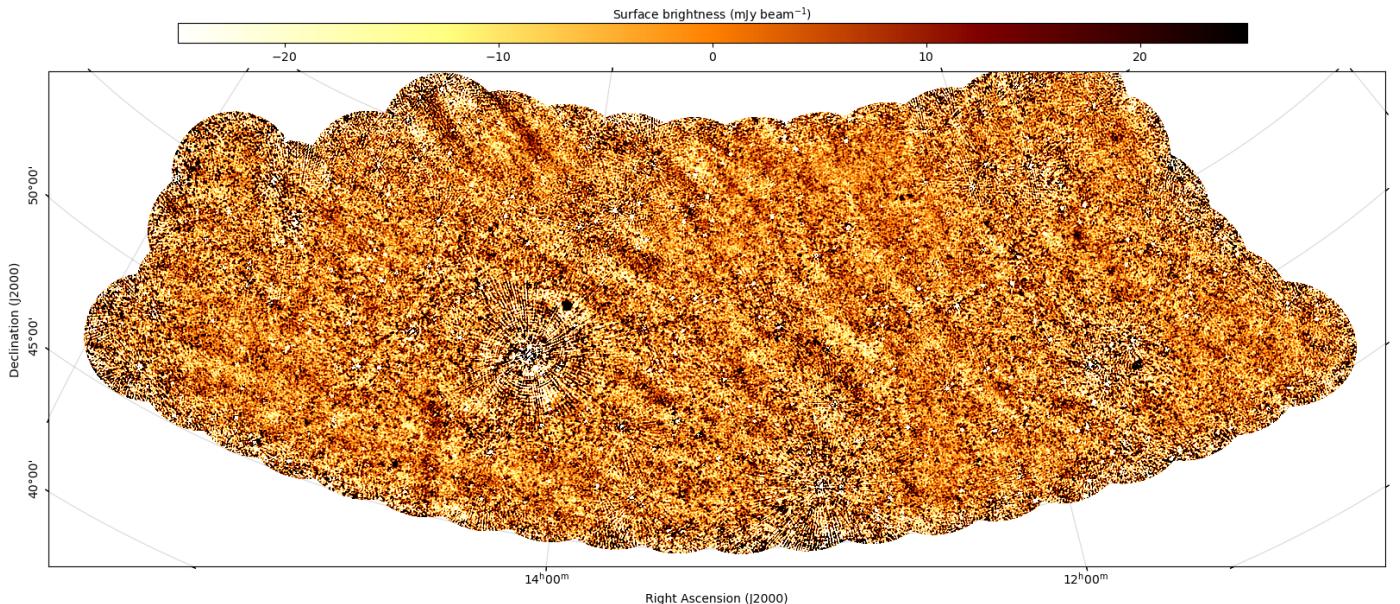
the error on the recovered flux density from the simulated value. In Fig. 8 we show the results for all 95 mosaics combined. We found a 50 percent probability of detecting sources at 8.5 mJy and 95 percent probability of detecting sources at 15 mJy. The same image also reports the fraction of recovered sources above a certain flux density. This shows that our catalogue is 50 percent complete over 5.2 mJy and 95 percent complete over 11 mJy, al-



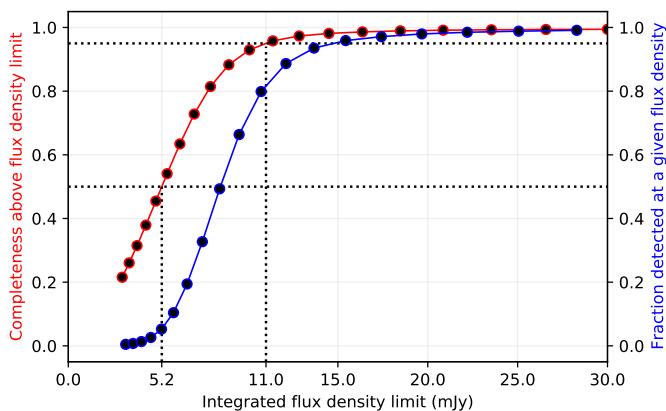
**Figure 5.** Mosaic image of the total intensity emission in the eastern half of survey area. The small panel is a zoom in the marked region.



**Figure 6.** Mosaic image of the total intensity emission in the western half of survey area. The small panel is a zoom in the marked region.



**Figure 7.** Low-resolution ( $3'$ ), source-subtracted map of the survey area. Large stripes of emission are visible crossing the region. The same structures were identified in the corresponding LoTSS images, which were calibrated independently with a different calibration code, demonstrating that these features are not artefacts.



**Figure 8.** In red the completeness function above a certain flux density limit and in blue the fraction of detected sources at that flux density. The dotted lines show the 50% and the 95% completeness at 5.2 and 11 mJy respectively.

though we note that these values for cumulative completeness depend on the assumed slope of the input source counts.

To assess the number of false positive detections, we started from the residual maps produced by the imager. These were mosaiced as described above and their pixel values are inverted. This resulted in negative pixels due to noise and artefacts becoming positive. PyBDSF was then run with the nominal parameters used to generate the source catalogue. However, since real sources were not present in the images, the rms map would have been different from the original run, therefore the code was forced to use the rms map generated from the original images. On this basis, we estimated that 1.4% of the sources in the LoLSS catalogue released here are due to artifacts.

## 5. Results

In this section we will present the properties of the first data release of the LOFAR LBA Sky Survey, including source extensions, sensitivity, astrometric accuracy and precision, and flux density uncertainties.

### 5.1. Source extension

Separating point-like from extended sources is notoriously difficult in regimes where phase errors are a relevant systematic effect. A perfectly point-like source, in the absence of calibration and deconvolution errors, is defined to have a ratio of integrated flux density ( $S_I$ ) to peak flux density ( $S_P$ ) equal to unity, with a source size equal to that of the restoring beam. As discussed in Sec. 3.1 of Shimwell et al. (2022), in optimal conditions the natural logarithm of the ratio of the two quantities,  $R = \ln(S_I/S_P)$ , is expected to have a Gaussian distribution. However, (i) the non-Gaussian rms noise of the map, (ii) the correlation between the error of the two quantities, and (iii) the increase of errors and general overestimation of the source sizes as  $S_I$  decreases, all combine to skew the distribution of  $R$ . The actual distribution is further complicated by the quantisation of the pixels in our images and the residual time- and bandwidth-smearing effects. Most importantly, any uncorrected variation in the ionospheric-induced phase shift adds to the source smearing, therefore artificially increasing the ratio  $R$ .

Deciding whether a source is extended in LoLSS is therefore not simple. We adopted a strategy similar to the one used for LoTSS, but taking advantage of the information obtained by combining the two surveys. Firstly, we located all isolated sources in the overlapping region of LoTSS DR2 and LoLSS, where “isolated” was taken to mean that no other detected source in the same survey was present within  $30''$ . Isolated sources of LoTSS and LoLSS are then cross-matched using a maximum distance of  $15''$ . The signal-to-noise ratio (SNR; defined as  $S_I/\sigma_{S_I}$ ) of LoLSS sources is plotted on the top panel of Fig. 9 against  $R = \ln(S_I/S_P)$ . It is evident that all sources are affected

by smearing, most likely dominated by residual ionospheric-induced phase errors combined with the other aforementioned effects.

In the second panel of Fig. 9, we restricted the catalogue to those sources classified as point sources in LoTSS following the definition of Sec. 3.1 of Shimwell et al. (2022). This reduces the number of sources by about a factor of two. Given the higher sensitivity and resolution of LoTSS, this procedure should ensure that the selected sources are point-like in LoLSS. We then binned the sources in ten SNR ranges starting from  $\text{SNR} = 4$ . The position of the 95th percentile was calculated in each bin and a sigmoid function was used to fit the envelope, finding:

$$R_{95} = 0.15 + \left( \frac{0.87}{1 + \left( \frac{\text{SNR}}{41.78} \right)^{1.23}} \right) \quad (1)$$

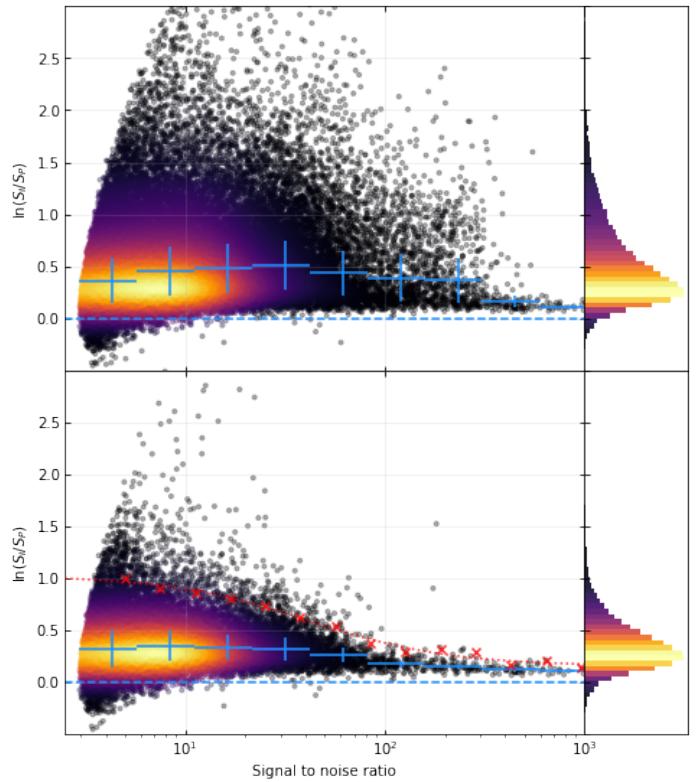
This function can be used to decide if a source in the LoLSS catalogue is resolved. Whilst this is merely one possible approach, it is more accurate than the simple  $S_I/S_P$  ratio, because it takes into account the effect of SNR. Other approaches for tackling the issue with LOFAR data were presented by e.g. Williams et al. (2016); Mahony et al. (2016). By applying our abovementioned criterion, we find that 9420 of LoLSS sources (22%) are classified as extended.

## 5.2. Sensitivity

In Fig. 10 we present the spatial distribution of the local rms noise value across the survey area. This was derived using the PyBDSF rms noise estimation. Rms noise images from each  $3.3^\circ \times 3.3^\circ$  mosaic were then reprojected and combined into a single large mosaic using montage<sup>3</sup>. The main source of local variation in the rms noise map are dynamic range limitations close to bright sources. The clearest example is seen around 3C 295. that creates the large region of low sensitivity visible in Fig. 10 around its position, marked with an “x”. Observations for the two northern strips were taken in 2019, while the rest of the field was covered in 2017 when the solar activity was higher due to the phase of the 11-yr solar cycle. A high solar activity induces strong ionospheric disturbances that results into more severe systematic errors in the data. While most of the fields include 7 to 8 hours of good data, P218+55 (marked in blue in Fig. 10) was observed for 16 hrs, and therefore shows a higher local sensitivity. On the contrary, field P174+57 (marked in red in Fig. 10) was not observed and the region was covered only by neighbouring fields resulting in a reduced local sensitivity.

In Fig. 11 we show the histogram of the pixel values of Fig. 10, to illustrate the rms noise distribution. The median value is  $1.63 \text{ mJy beam}^{-1}$ , however the distribution is not symmetric due to a long tail of high rms noise pixels caused by dynamic range limitations around bright sources and edge effects. For the majority of the survey area that is not at the edge of the coverage or particularly close to bright sources, the peak of the distribution at an rms noise of  $1.55 \text{ mJy beam}^{-1}$  is more representative for the data quality.

In order to estimate the effect of dynamic range limitation around bright sources we located all sources that are well isolated, i.e. with no other detected sources within  $450''$ . Then we estimated the rms noise using the residual mosaic maps in concentric annuli with a thickness of  $15''$ . Sources were then binned based on their flux density and the median rms noise was estimated as a function of distance from the source position (see



**Figure 9.** The logarithm of the ratio of integrated to peak flux density plotted against the signal-to-noise ratio as measured by the source finder. A perfect point source would have  $\ln(S_I/S_P) = 0$  (blue dashed line). In the top panel we show all isolated sources present in the catalogue. The bottom panel is restricted to those sources whose cross-match is an isolated point source in LoTSS (see text). Blue crosses show the binned median with the extension on the x-axis showing the bin size and on the y-axis showing  $\pm 1$  median absolute deviation. Red markers show the position of the 95th percentile of the distribution in each bin. Dotted red line is a fit to the red crosses as explained in the text.

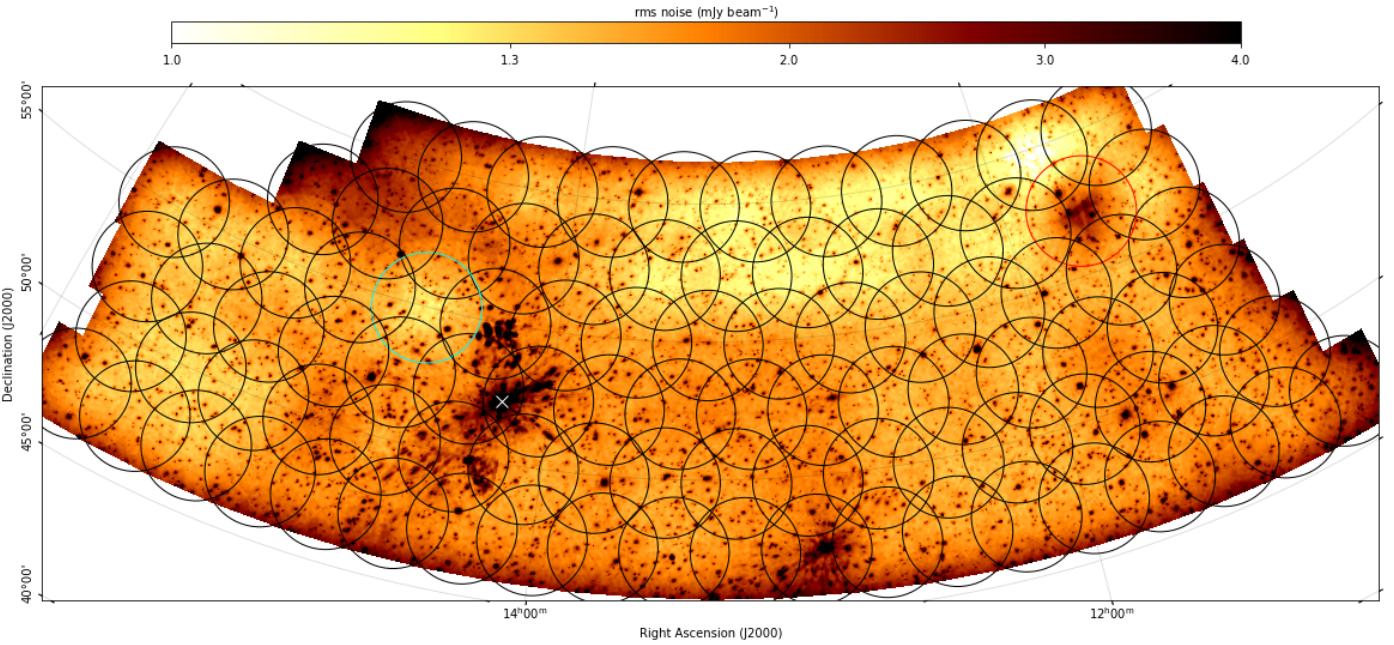
Fig. 12). The procedure shows an increase in the rms noise close to sources brighter than  $\sim 100 \text{ mJy}$ .

## 5.3. Astrometric precision and accuracy

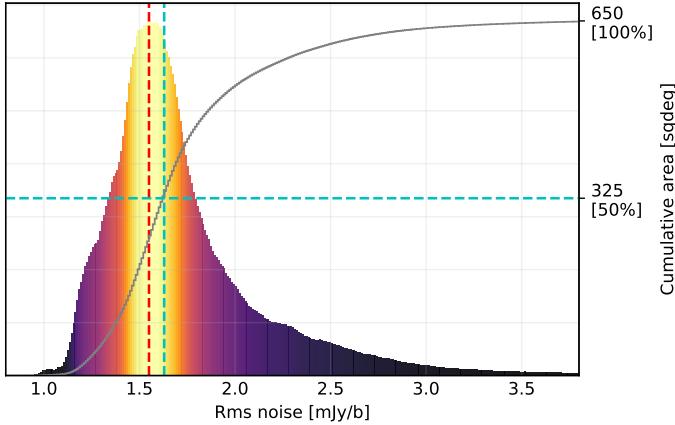
As discussed above, before performing mosaicing of individual images we applied an astrometric correction based on the distance with cross-matched FIRST sources. To estimate the global astrometric precision and accuracy, we started by keeping only point-like sources in LoLSS as defined by eq. 1. Then we selected all isolated LoLSS sources (no other source within  $45''$ ), as well as all isolated FIRST sources (no other source within  $15''$ ) and cross-matched the resulting catalogues. The cross-match was done starting with a maximum separation of  $100''$  and keeping only sources within 10 times the median absolute deviation (MAD) of all distances between matched sources. This procedure is then applied iteratively, recalculating the median absolute deviation until it converges. The final matching distance is  $6''$ , which give 12 859 LoLSS sources with an associated FIRST counterpart.

In Fig. 13 we plot the RA and Dec separation between the matched sources. We can then use as an estimation of the astrometric accuracy the mean separations  $E_{\text{RA}} = -0.08''$  and

<sup>3</sup> <http://montage.ipac.caltech.edu/>.



**Figure 10.** Rms noise map of the LOFAR LBA Sky Survey - DR1. The regions with reduced sensitivity are located around bright sources that induce dynamic range limitations. Each pointing is shown with a circle at the FWHM. The red pointing (P174+57) was not observed, while the blue pointing (P218+55) was observed for 16 hrs (twice as much as the others). The white “x” indicates the position of the bright source 3C 295, which affects several fields around it.

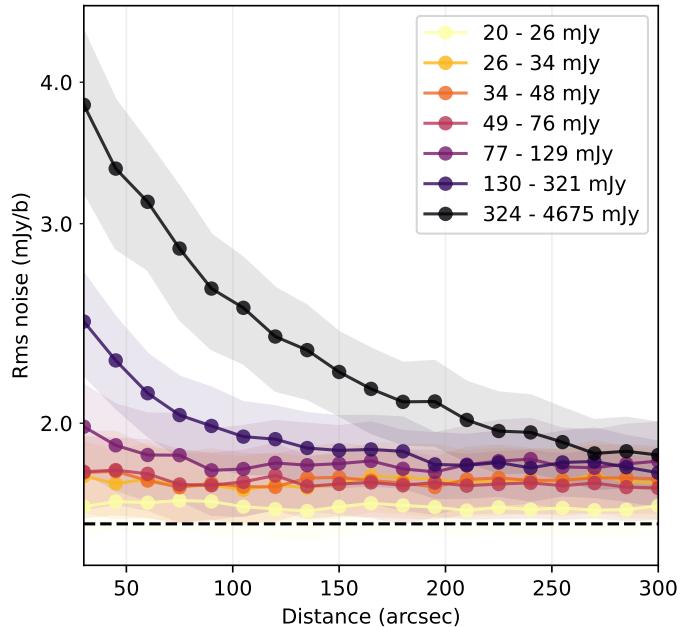


**Figure 11.** Rms noise histogram of the pixels included in the released region. The solid line shows the cumulative function. The red dashed line indicates the position of the most common value of the distribution ( $1.55 \text{ mJy beam}^{-1}$ ). The blue dashed line shows the position of the 50% percentile (median) at  $1.63 \text{ mJy beam}^{-1}$ . This means that half of the covered area ( $325 \text{ deg}^2$ ) has a lower rms noise than that. The long tail of high rms noise is due to bright sources in the field.

$E_{\text{Dec}} = 0.04''$ , and as an estimation of the astrometric precision the standard deviation  $\sigma_{\text{RA}} = 1.48''$  and  $\sigma_{\text{Dec}} = 1.17''$ .

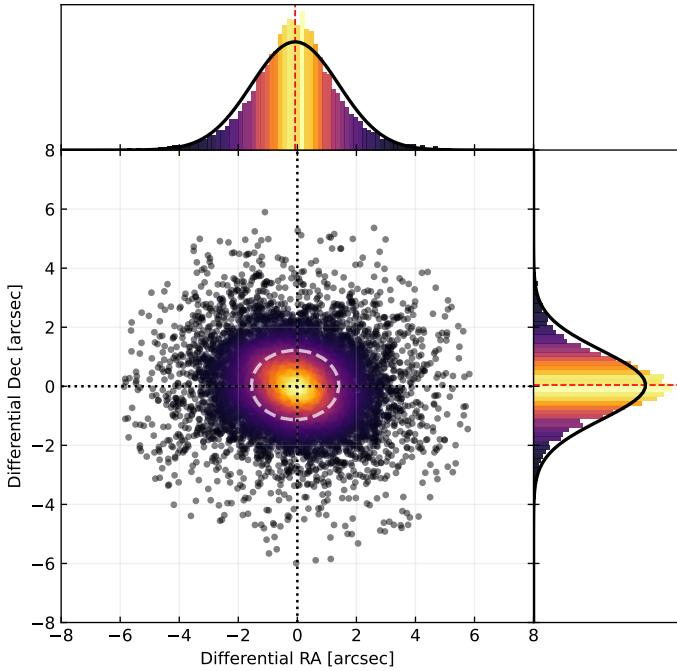
#### 5.4. Flux density uncertainties

The major source of uncertainties in the LoLSS flux density scale is the LOFAR beam analytic model that is used to compensate for the flux suppression due to both the dipole beam and the station beam. For LoLSS observations, the instrumental bandpass was estimated using the calibrator beam and a calibrator



**Figure 12.** Dynamic range limitation around bright sources. The x-axis reports the distance in arcsec from bright isolated sources. The y-axis shows the local rms noise. Each line shows the median for about 190 sources in different flux density intervals with a shaded region that illustrates the size of one median absolute deviation. The black dashed line is at  $1.63 \text{ mJy beam}^{-1}$ .

model outlined by Scaife & Heald (2012). The bandpass solutions were then transferred to the target beams so that the flux density scale matches the one used for the calibrator model. No further corrections were applied to the flux density scale, which makes LoLSS independent of other radio surveys. Therefore, the

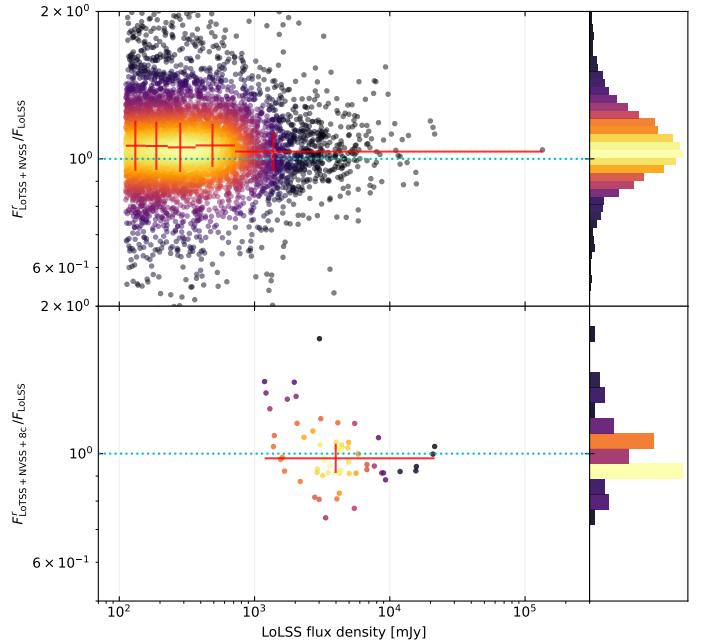


**Figure 13.** Astrometric accuracy of the sources present in the catalogue measured using the FIRST survey as a reference. The mean errors are  $E_{\text{RA}} = -0.08''$  and  $E_{\text{Dec}} = 0.04''$  with standard deviation  $\sigma_{\text{RA}} = 1.48''$  and  $\sigma_{\text{Dec}} = 1.17''$ . The standard deviation of the distribution is shown with a white dashed line. In the histograms, the black solid lines show Gaussians with the aforementioned means and standard deviations, the red dashed lines indicate the position of the means.

accuracy is limited only by instrumental stability and the primary beam model. This is in contrast to LoTSS, where the 6C and NVSS surveys were used to re-scale the LoTSS images to the expected flux density scale (Shimwell et al. 2022, Sec. 3.3.1).

Following de Gasperin et al. (2021), we can give an initial estimation of the expected flux scale uncertainty. The first source of uncertainty is due to the Scaife & Heald (2012) flux density scale, that has a nominal error ranging between two and four percent depending on the source used as calibrator. The second uncertainty is due to primary beam model errors. These were estimated by observing two calibrators simultaneously and comparing the derived bandpass solutions. This experiment showed an error of about five percent. Finally, an estimation of the flux density uncertainty due to calibration and imaging processes was derived by measuring the flux density of the calibrator source 3C 295, which is within the survey footprint. The measured flux density of that source is 134.4 Jy compared to an expected flux density of 133.3 Jy, a difference of one percent. Adding the three uncertainties in quadrature results in a combined systematic uncertainty of 6 percent.

In de Gasperin et al. (2021), we emphasised the presence of a 10% systematic discrepancy between the flux density scale of LoLSS and LoTSS. However, the discrepancy was derived using images without direction-dependent corrections and assuming a straight extrapolation of the spectral index from NVSS and LoTSS, down to LoLSS frequencies. Consequently, it was unclear how much of this effect could be attributed to the average spectral curvature of radio sources. With the improved data of LoLSS DR1, combined with the in-band spectra, this can be evaluated in more detail.



**Figure 14.** LoLSS flux density versus the expected flux density at 54 MHz from the linear extrapolation of the spectral energy density (SED) using NVSS and LoTSS (top panel) or from a second order polynomial SED estimated using NVSS, LoTSS and 8C (bottom panel). A ratio of 1, the blue dotted lines, means a perfect match between the prediction and the flux density measured in LoLSS. Red crosses are binned medians (five bins for the top panel, one for the bottom) with  $\pm 1$  median absolute deviation shown as an extension in the y-direction and the bin size as an extension in the x-direction.

For this purpose, we used only isolated detections (i.e. no other sources within twice the survey resolution with a minimum of  $30''$ ) and we cross-matched all the surveys with LoLSS using a maximum distance for a positive match of  $5''$  for LoTSS,  $10''$  for NVSS, and  $30''$  for 8C. Firstly, we compared the measured LoLSS flux density with the expected value at 54 MHz estimated via a linear extrapolation of the SED that is derived from NVSS and LoTSS flux densities. In this case we have 14 645 usable sources. A second cut reduced the number to 6 008 when we applied a flux-limit on each survey so that every source with a spectral index  $-0.2 < \alpha < -1.2$  was catalogued in all surveys. This constraint should almost completely eliminate the bias due to the different flux density cuts of the various surveys. The results are shown in the top panel of Fig. 14. In this case, the flux density of LoLSS is lower than predicted by 3 – 6 percent (median:  $1.056 \pm 0.006$ ; MAD: 0.11), depending on the flux density. In the lower panel of the same figure, we compared the LoLSS flux density with the expected value at 54 MHz derived from a second order polynomial, therefore a curved spectrum, obtained using NVSS, LoTSS, and 8C. Although 8C has a partial overlap with LoLSS footprint and a low sensitivity and resolution ( $\sim 200 \text{ mJy beam}^{-1}$  with beam:  $270'' \times 270''$ ), the survey is at 38 MHz, a lower frequency than LoLSS, which allows a better constraint compared with an extrapolation. In this case, we have only 61 matched sources but the flux density of LoLSS is well aligned with predictions with a median separation that is two percent (median:  $0.98 \pm 0.05$ ; MAD: 0.07).

These results are in line with most of LoLSS sources having straight or slightly curved spectra with a possible reduced

flux density at 54 MHz compared with linear extrapolation from higher frequencies. The flux density scales of LoLSS and LoTSS seem to agree within the given uncertainties. We conclude that LoLSS does not possess an appreciable systematic flux scale offset. The precision of the LoLSS flux density scale is estimated to be 6 percent.

### 5.5. In-band spectral index

The in-band spectral indices of LoLSS were compared with the wide-band spectral indices as derived from other surveys. This comparison is presented in the first two panels of Fig. 15, where LoLSS in-band spectral indices are compared with the wide-band spectral indices at high frequency (LoTSS – NVSS) and low-frequency (LoLSS – LoTSS). In both cases, the in-band spectral indices are much flatter than the wide-band spectral indexes with a median in-band spectral index of  $-0.479 \pm 0.008$  (MAD: 0.21) compared with a median wide-band spectral index of  $\alpha_{\text{LoTSS-NVSS}} = -0.820 \pm 0.003$  (MAD: 0.10) and  $\alpha_{\text{LoLSS-LoTSS}} = -0.765 \pm 0.005$  (MAD: 0.14), see also Boehme et al. (in prep.) where LoLSS preliminary release sources matched to other radio surveys also hint at curved spectra. Other low-frequency radio surveys, such as GLEAM, have found a relatively small fraction (< 10% percent) of curved spectra (Callingham et al. 2017). It is important to note that GLEAM reaches down to 72 MHz, making it slightly harder to identify a low-frequency deviation from a straight spectrum.

The comparisons done so far are subject to effects due to the different frequency ranges used for the in-band spectral indices and the possible curvature of the source SEDs. To further investigate to what extent the in-band spectra are reliable, we fitted a second order polynomial function to the flux density values of LoLSS, LoTSS, and NVSS. The tangent to these polynomial, evaluated at 54 MHz is then compared with the in-band spectral index (see Fig. 16). Again the in-band spectral index are flatter than expectations, with a median difference between expectation and in-band of  $-0.268 \pm 0.008$  (MAD: 0.18). The median becomes  $-0.17 \pm 0.07$  (MAD: 0.13) when considering also 8C although this reduces the sample to just 60 sources. Our conclusion is that a certain degree of flattening is expected and encoded in the in-band spectral index; however, a systematic offset of between +0.2 and +0.3 might be present. Finally, we checked that the average in-band flux density is compatible with the one reported in the main catalogue and we find a median deviation < 0.2%.

In Fig. 17, we show the median flux density value of all matched sources for two sets of radio surveys. The plot underline the overall good alignment of the various flux scales as well as a local flattening at low frequencies. The flattening is also present in the in-band spectral index but, as discussed above, stronger than predicted.

Finally, we examined the 40 brightest sources for which 8C, LoLSS, LoTSS, and NVSS flux densities are positively matched. For these sources we show in Fig. 18 the extrapolated linear spectra using LoTSS and NVSS (yellow dotted lines) and the second order polynomial fit done using 8C, LoTSS, and NVSS (green line) and including also in-band LoLSS values (red line). In all cases, LoLSS data seem to fit well the second order polynomial function although in certain cases the in-band spectra index (black dotted lines) results are flatter than expected. The shape of the functions differ only marginally when LoLSS data are included in the fit, showing again a good overall agreement.

## 6. Discussion

### 6.1. Curved spectra

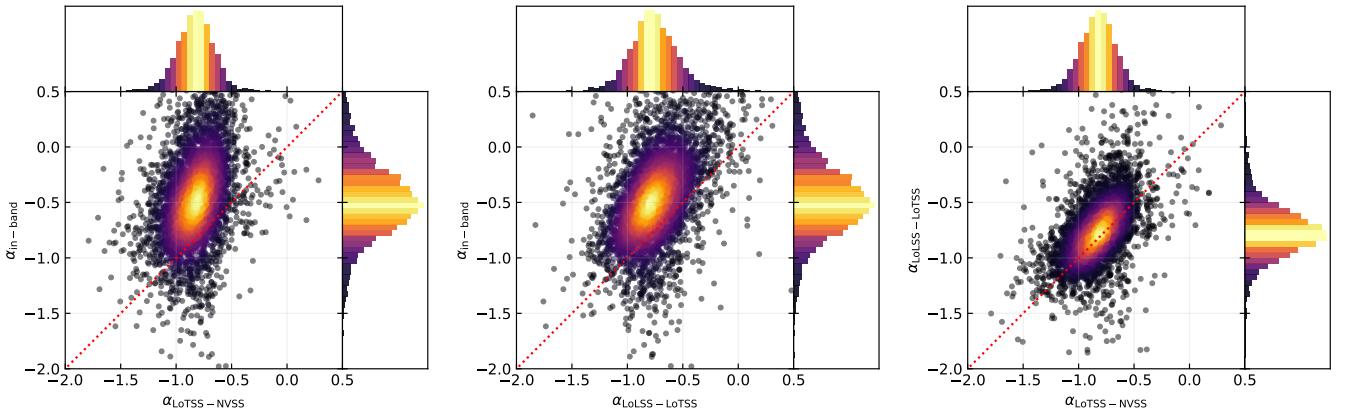
The two main LOFAR surveys, LoTSS (144 MHz) and LoLSS (54 MHz) are highly complementary, with the first having higher resolution and sensitivity and the second providing information on the critical parameter that is the low-frequency spectral index. Currently, no other existing or planned instrument besides LOFAR LBA will be able to deliver information on the radio SED down to tens of MHz, where only the brightest radio sources have been studied up till now.

The catalogue presented in this work showed that the average radio SED shows a curvature at frequencies between 50 – 1400 MHz, going from a median high-frequency spectral index  $\alpha_{\text{LoTSS-NVSS}} = -0.820 \pm 0.003$  (MAD: 0.10) to a median low-frequency spectral index  $\alpha_{\text{LoLSS-LoTSS}} = -0.765 \pm 0.005$  (MAD: 0.14), see also Boehme et al. (in prep.) where LoLSS preliminary release sources matched to other radio surveys also hint at curved spectra. Other low-frequency radio surveys, such as GLEAM, have found a relatively small fraction (< 10% percent) of curved spectra (Callingham et al. 2017). It is important to note that GLEAM reaches down to 72 MHz, making it slightly harder to identify a low-frequency deviation from a straight spectrum.

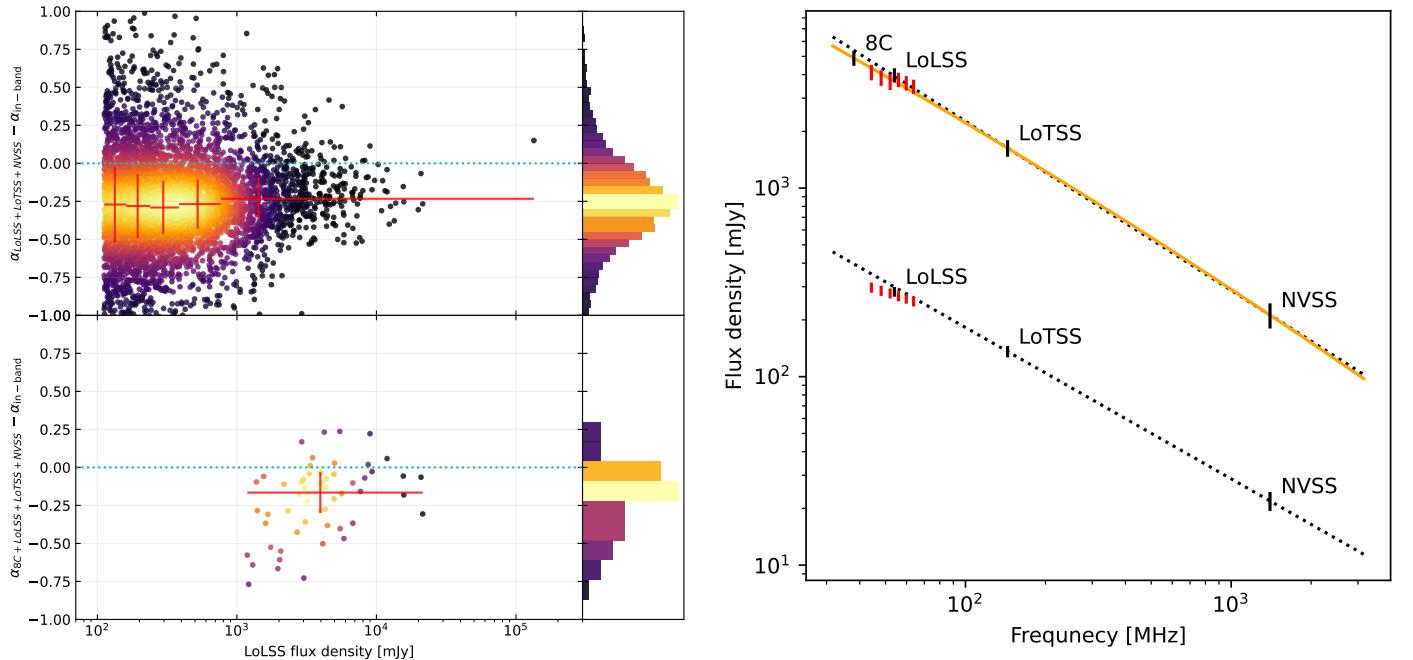
A deviation from straight spectra is not new for samples of sources observed at low-frequencies. Laing & Peacock (1980); Van Weeren et al. (2014), for instance, showed that for a sample of sources selected at 178 MHz, the deviation of the measured spectra from the flux density, predicted with a power-law fit, increased with decreasing frequency. This was an indication of curved spectra becoming increasingly dominant at frequencies < 100 MHz. However, their analysis was restricted to FR II sources whose flux density is strongly affected or dominated by the emission from their hot-spots. Consequently, their interpretation was that self-absorption in the compact hot-spots was the main driver for the downward curvature (Laing & Peacock 1980).

For the first time, our catalogue provides flux density measurements at frequencies well below 100 MHz for a large population of radio sources. At the few mJy level reached by LoLSS, the dominant population are FR I radio galaxies (Mingo et al. 2019). It is important to note that star-forming galaxies, well present in the LoTSS survey catalogue, are sparsely represented in LoLSS, and the detection of this population at 54 MHz requires a sensitivity of 0.5 – 1 mJy (Williams et al. 2021).

Unlike FR II radio galaxies, at low frequencies FR I radio galaxies are not dominated by the emission of compact regions, but rather by the diffuse emission of their lobes. This makes any absorption mechanisms less likely to cause the turnover. A possible explanation is that integrated radio SED are a superimposition of aged spectra with a rather flat ( $\alpha \sim -0.5$ ) injection index, and the ultra-low frequencies are dominated by emission from low-energy electrons that have not lost most of their energy yet. Therefore, their energy distribution generates radio emission that preserves a flatter spectral index, closer to the injection index. Alternatively, we might be witnessing the beginning of the expected low-energy cutoff of the electron distribution. Particle acceleration processes in radio AGN in fact accelerate electrons following a power-law energy distribution  $n(E)dE \propto E^{-\delta}dE$ , which translates into a straight radio SED with spectral index  $\alpha = \frac{1-\delta}{2}$ . However, the number of electrons at low energy cannot increase indefinitely, and a cutoff is required. A detailed analysis of this result is out of the scope of this paper and will be presented in a forthcoming publication.



**Figure 15.** Left: High-frequency wide-band radio spectral index (144 – 1400 MHz) from LoTSS and NVSS versus the LoLSS in-band spectral index (42 – 66 MHz). Center: the same but using a low-frequency wide-band radio spectral index (54 – 144 MHz) from LoLSS and LoTSS. Right: comparison between the high- and the low-frequency wide-band spectral index.



**Figure 16.** In-band spectral index plotted against its separation with the tangent of a second order polynomial fit to the SED derived using LoLSS, LoTSS, and NVSS (top panel) and 8C, LoLSS, LoTSS, and NVSS (bottom panel). Red crosses are binned medians (five bins for the top panel, one for the bottom) with  $\pm 1$  median absolute deviation showed as an extension in the y-direction and the bin size as an extension in the x-direction. The average in-band spectral index results flatter than expectations.

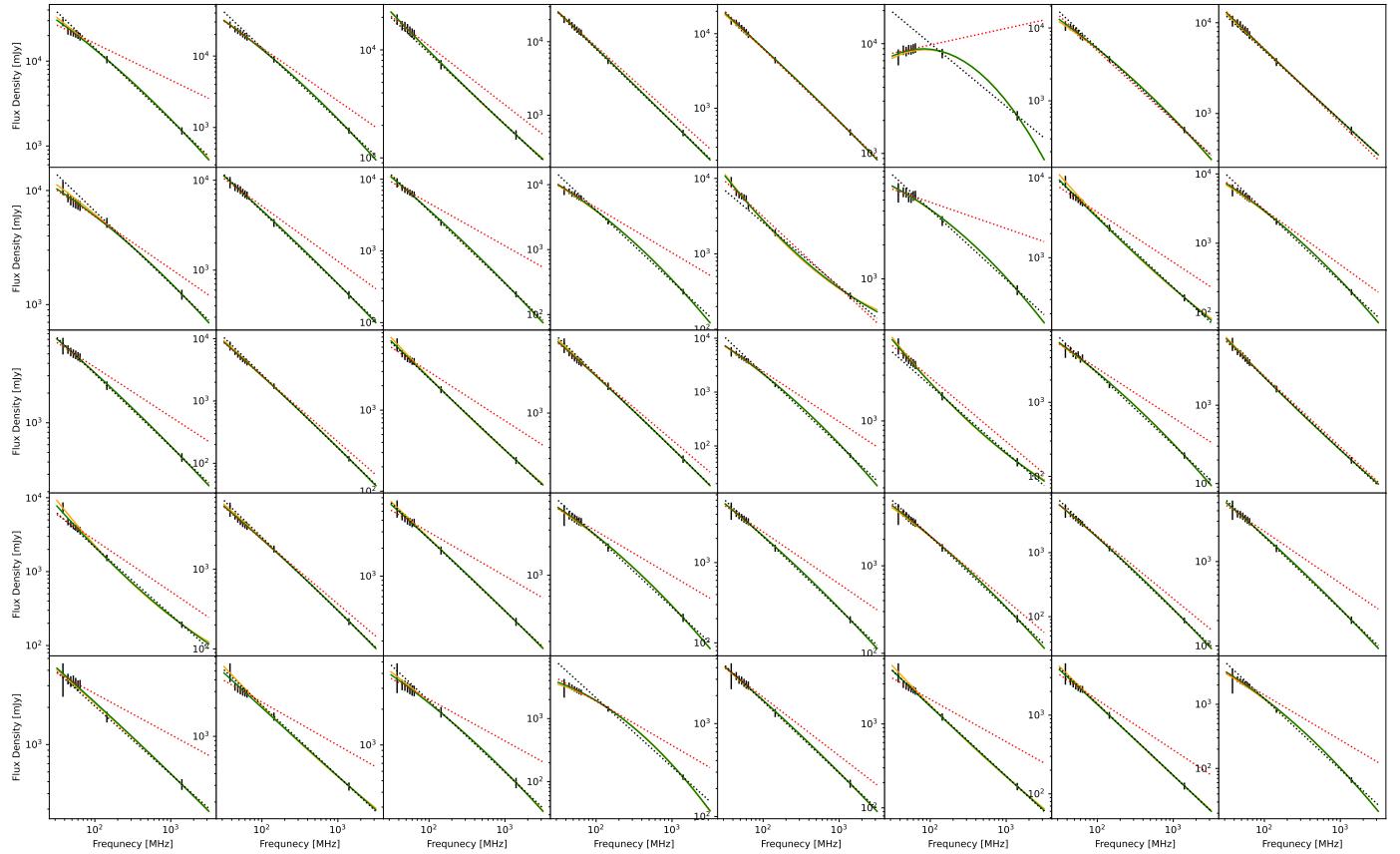
## 6.2. Source counts

The Euclidean-normalised differential source counts for LoLSS are plotted in Fig 19 where the uncertainties on the final normalized source counts were propagated from the error on the completeness correction and the Poisson errors (Gehrels 1986) on the raw counts per flux density bin. Incompleteness was accounted for by using the measured peak intensities to calculate the fractional area of the survey in which each source can be detected,  $A_i$ , with the total count in each flux density bin then

**Figure 17.** Median flux densities of all matched sources for 8C, LoLSS, LoTSS, and NVSS (61 source; top line) and LoLSS, LoTSS, and NVSS (6008 sources; bottom line). In red are the LoLSS in-band flux densities. The height of the lines shows the uncertainty on the median. Black dotted lines are first order polynomials derived from NVSS and LoTSS values. Yellow line is the second order polynomial derived from 8C, LoTSS, and NVSS.

determined as  $N = \sum 1/A_i$ . The error on this correction was estimated from the measured uncertainty on each peak intensity and subsequent error on the visibility area. A small resolution bias correction, taking into account the size distribution of sources, was made following Williams et al. (2016) and Prandoni et al. (2001). The LoLSS source counts show good agreement with these previously-determined counts for the preliminary release and reliably probe down to 10 mJy compared to  $\sim 25$  mJy for the preliminary release.

Like the preliminary release, the source counts presented here show excellent agreement with higher frequency counts,



**Figure 18.** Examples of radio SEDs of isolated bright sources with a detected 8C counterpart. In each plot, the black vertical lines show the data points of 8C (38 MHz), LoLSS (44, 48, 52, 56, 60, 64 MHz), LoTSS (144 MHz) and NVSS (1400 MHz) with uncertainties. The black dotted line connects LoTSS and NVSS data points. The red dotted line is a linear regression using only the LoLSS in-band data points. The green line is a 2nd order regression including all data points while the yellow line is a 2nd order regression without considering LoLSS data points. The green and yellow lines are almost always superimposed.

with a change in the average spectral index of the population at lower flux densities. For comparison, we considered the 1.4 GHz source counts compilation of De Zotti et al. (2010), scaled down to 54 MHz assuming two different spectral indices. Above around 100 mJy, the average spectral index is consistent with  $-0.8$ , while below this value a spectral index of  $-0.6$  gives better agreement between the high and low frequency source counts.

### 6.3. Preliminary analysis of LoLSS Stokes V data

As detailed in Section 3.4, Stokes V mosaics are produced by the LoLSS data reduction pipeline. Pulsars, brown dwarfs, stars, and exoplanets are known to emit highly circularly-polarised radiation (e.g. Vedantham et al. 2020; Callingham et al. 2021).

We searched the Stokes V mosaics of LoLSS for any significant circularly-polarised sources. We find the leakage of total intensity into Stokes V to be  $\approx 0.5\%$ , a factor of five worse than that found for V-LoTSS (Callingham et al. in prep). Such a difference between LoLSS and LoTSS is likely due to larger Faraday rotation corrections performed at lower-frequencies and possible mechanical issues with the LBA dipoles.

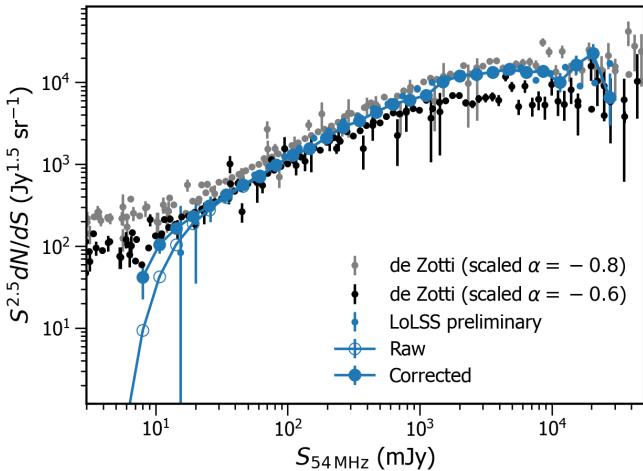
For this release of LoLSS we find only one circularly-polarised source that is not consistent with leakage. The source is pulsar PSR B1508+55, as shown in Figure 20, and has a circularly-polarised fraction of  $1.6 \pm 0.3\%$  and is an  $18\sigma$  detection in Stokes V. A complete analysis of the LoLSS Stokes V properties will be the focus of a follow-up manuscript but the

detection of PSR B1508+55 demonstrates that LoLSS Stokes V products are science-ready.

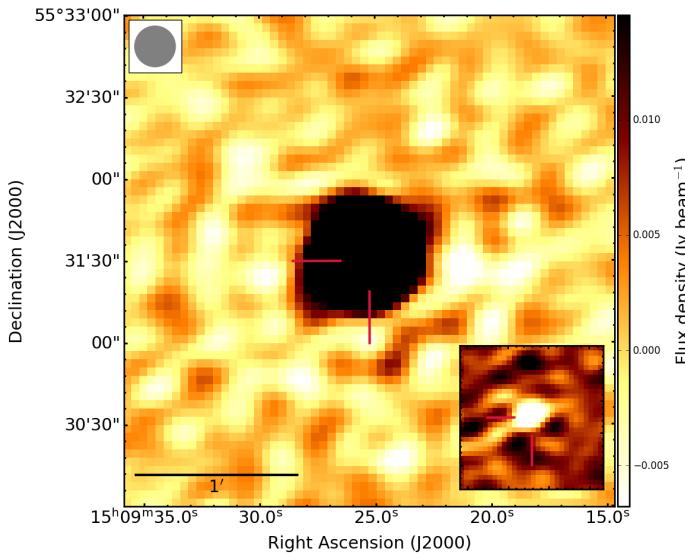
## 7. Summary

In this work, we have presented the first data release of the LOFAR LBA Sky Survey (LoLSS). The interferometer visibilities were automatically processed using the Pipeline for LOFAR LBA (PiLL); compared with the preliminary release, we have now included the derivation and correction of direction-dependent systematic errors. The typical rms noise of the survey is  $1.55 \text{ mJy beam}^{-1}$  at a resolution of  $15''$ , although several regions of the survey are limited in sensitivity by the dynamic range issues that affect the surroundings of sources with flux density  $S_v > 100 \text{ mJy}$ .

The footprint of the survey is about  $650 \text{ deg}^2$  and the final catalogue contains 42 463 sources. The catalogue is 50% complete at  $5.2 \text{ mJy}$  and 95% complete at  $11 \text{ mJy}$ . The fraction of false positives is estimated at 1.4%. Compared to the LOFAR Two-metre Sky Survey (LoTSS), where most (92%) of the detected sources are point-like, in LoLSS this fraction decreases to 78% despite the lower resolution. This is likely due to the lower sensitivity, which favours the detections of nearby sources. The survey astrometric precision is estimated to be  $\sigma_{\text{RA}} = 1.48''$  and  $\sigma_{\text{Dec}} = 1.17''$ . The flux scale accuracy is estimated to be at 6%. This has been validated by cross-checking LoLSS flux densities with the values derived from existing surveys.



**Figure 19.** Euclidean-normalised differential source counts for LoLSS between 10 mJy and 30 Jy. The open circles show the raw, uncorrected source counts, while the filled circles show the counts corrected for completeness and resolution bias. The 1.4 GHz source counts from various surveys compiled De Zotti et al. (2010), and scaled to 54 MHz assuming a spectral index of -0.8 (in gray) and -0.6 (in black) are shown for comparison.



**Figure 20.** Total intensity image of PSR B1508+55 with a Stokes V image inset. The red cross-hairs point to PSR B1508+55, and the synthesised beam is displayed at the top-left corner. The total time and frequency average flux density of PSR B1508+55 in Stokes I and V is 1.42 Jy and -25 mJy, respectively.

Six in-band flux density measurements are also derived. While their median values are in line with the full-band catalogue, the in-band spectral index appears systematically flatter (about +0.2 – +0.3) compared to expectations. However, this systematic offset cannot account for the full flattening, which is interpreted to be partially physical.

We plan to complete the observations of the northern sky at Dec > 24° by the end of 2022. These observations and those taken in the past years will be included in a forthcoming release of LoLSS. Starting from 2024, LOFAR will enter into a

phase of renovation collectively called “LOFAR 2.0” (Hessels et al. in prep.). This process will enable simultaneous LBA+HBA observations for better ionospheric calibration, the use of all LBA dipoles increasing the collecting area of each station, and the possibility to exploit a larger number of beams increasing the survey speed. Finally, a large superstation which has been completed in France (NenuFAR; Zarka et al. 2012) will boost LOFAR-VLBI sensitivity in the LBA band and potentially permit the use of international baselines in future surveys.

**Acknowledgements.** LOFAR is the LOw Frequency ARray designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and are collectively operated by the ILT foundation under a joint scientific policy. The ILT resources have benefited from the following recent major funding sources: CNRS-INSU, Observatoire de Paris and Université d’Orléans, France; BMBF, MIWF-NRW, MPG, Germany; Science Foundation Ireland (SFI), Department of Business, Enterprise and Innovation (DBEI), Ireland; NWO, The Netherlands; The Science and Technology Facilities Council, UK; Ministry of Science and Higher Education, Poland; Istituto Nazionale di Astrofisica (INAF). This research has made use of the University of Hertfordshire high-performance computing facility (<https://uhhpc.herts.ac.uk/>) and the LOFAR-UK compute facility, located at the University of Hertfordshire and supported by STFC [ST/P000096/1].

FdG and MB acknowledge support from the Deutsche Forschungsgemeinschaft under Germany’s Excellence Strategy - EXC 2121 “Quantum Universe” - 390833306. MJH acknowledges support from the UK Science and Technology Facilities Council (ST/R000905/1). RJW acknowledges support from the ERC Starting Grant ClusterWeb 804208. AB acknowledges support from the ERC-StG DRANOEL n. 714245. DJB acknowledges funding from the German Science Foundation DFG, via the Collaborative Research Center SFB1491 “Cosmic Interacting Matters - From Source to Signal.”

## References

- Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559
- Bock, D. C.-J., Large, M. I., & Sadler, E. M. 1999, Astron. J., 117, 1578
- Callingham, J. R., Ekers, R. D., Gaensler, B. M., et al. 2017, ApJ, 836
- Callingham, J. R., Vedantham, H. K., Shimwell, T. W., et al. 2021, Nat. Astron., 5, 1233
- Cheng, Y. 1995, IEEE Trans. Pattern Anal. Mach. Intell., 17, 790
- Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, Astron. J., 8065, 1693
- de Gasperin, F., Brunetti, G., Brüggen, M., et al. 2020a, A&A, 642, A85
- de Gasperin, F., Dijkema, T. J., Drabent, A., et al. 2019, A&A, 5, A5
- de Gasperin, F., Lazio, T. J. W., & Knapp, M. 2020b, A&A, 644, A157
- de Gasperin, F., Mevius, M., Rafferty, D. A., Intema, H. T., & Fallows, R. A. 2018, A&A, 615, A179
- de Gasperin, F., Williams, W. L., Best, P., et al. 2021, Astron. Astrophys., 648, A104
- de Zotti, G., Massardi, M., Negrello, M., & Wall, J. 2010, Radio and millimeter continuum surveys and their astrophysical implications
- Edler, H. W., de Gasperin, F., Brunetti, G., et al. 2022, A&A, arXiv:2207.11040
- Gaensler, B., Landecker, T., Taylor, A., & POSSUM Collaboration. 2010, in Am. Astron. Soc. Meet. Abstr., Vol. 42, 470.13
- Gehrels, N. 1986, ApJ, 303, 336
- Groeneveld, C., Van Weeren, R. J., Miley, G. K., et al. 2022, Astron. Astrophys., 658, A9
- Heald, G. H., Pizzo, R. F., Orrú, E., et al. 2015, A&A, 582, A123
- Hill, G. J., Gebhardt, K., Komatsu, E., et al. 2008, Panor. Views Galaxy Form. Evol., 399, 115
- Hurley-Walker, N., Callingham, J. R., Hancock, P. J., et al. 2017, MNRAS, 464, 1146
- Hurley-Walker, N., Galvin, T. J., Duchesne, S. W., et al. 2022, arXiv:2204.12762
- Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, A&A, 598, A78
- Intema, H. T., van der Tol, S., Cotton, W. D., et al. 2009, A&A, 501, 1185
- Lacy, M., Baum, S. A., Chandler, C. J., et al. 2020, Publ. Astron. Soc. Pacific, 132, 035001
- Laing, R. A. & Peacock, J. A. 1980, MNRAS, 190, 903
- Lane, W. M., Cotton, W. D., van Velzen, S., et al. 2014, MNRAS, 440, 327
- Mahony, E. K., Morganti, R., Prandoni, I., et al. 2016, MNRAS, 463, 2997
- Mevius, M., van der Tol, S., Pandey, V. N., et al. 2016, Radio Sci., 51, 927
- Mingo, B., Croston, J. H., Hardcastle, M. J., et al. 2019, MNRAS, 18, 2701
- Mohan, N. & Rafferty, D. 2015, Astrophys. Source Code Libr., 1502.007

- Morabito, L. K., Jackson, N. J., Mooney, S., et al. 2022, A&A, 658, A1
- Norris, R. P. 2017, Nat. Astron., 1, 671
- Norris, R. P., Hopkins, A. M., Afonso, J., et al. 2011, Publ. Astron. Soc. Aust., 28, 215
- Norris, R. P., Marvil, J., Collier, J. D., et al. 2021, Publ. Astron. Soc. Aust., 38, e046
- Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, MNRAS, 405, 155
- Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606
- Offringa, A. R. & Smirnov, O. 2017, MNRAS, 471, 301
- Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012, A&A, 539, A95
- Prandoni, I., Gregorini, L., Parma, P., et al. 2001, Astron. Astrophys., 365, 392
- Rees, N. 1990, MNRAS, 244, 233
- Rengelink, R. B., Tang, Y., de Bruyn, a. G., et al. 1997, Astron. Astrophys. Suppl., 124, 259
- Renting, G. A. & Holties, H. A. 2012, in Astron. Data Anal. Softw. Syst. XXI
- Sabater, J., Best, P. N., Tasse, C., et al. 2021, Astron. Astrophys., 648
- Saff, E. B. & Kuijlaars, A. B. 1997, Math. Intell., 19, 5
- Scaife, A. M. M. & Heald, G. H. 2012, MNRAS, 423, 30
- Shimwell, T. W., Hardcastle, M. J., Tasse, C., et al. 2022, A&A, 659, A1
- Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017, A&A, 598, A104
- Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, Astron. Astrophys., 622, A1
- Sweijen, F., van Weeren, R. J., Röttgering, H. J., et al. 2022, Nat. Astron., 6, 350
- Tasse, C. 2014, Astron. Astrophys., 566, A127
- Tasse, C., Hugo, B., Mirmont, M., et al. 2018, A&A, 611, A87
- Tasse, C., Shimwell, T., Hardcastle, M. J., et al. 2021, Astron. Astrophys., 648
- van der Tol, S., Jeffs, B., & van der Veen, A.-J. 2007, IEEE Trans. Signal Process., 55, 4497
- van Diepen, G., Dijkema, T. J., & Offringa, A. 2018, DPPP: Default Pre-Processing Pipeline
- van Haarlem, M. P., Wise, M. W., Gunst, a. W., et al. 2013, A&A, 556, A2
- van Weeren, R. J., Williams, W. L., Hardcastle, M. J., et al. 2016, Astrophys. J. Suppl. Ser., 223, 2
- Van Weeren, R. J., Williams, W. L., Tasse, C., et al. 2014, ApJ, 793, 82
- Vedantham, H. K., Callingham, J. R., Shimwell, T. W., et al. 2020, Nat. Astron., 1
- Williams, W. L., De Gasperin, F., Hardcastle, M. J., et al. 2021, Astron. Astrophys., 655, A40
- Williams, W. L., Van Weeren, R. J., Röttgering, H. J., et al. 2016, MNRAS, 460, 2385
- Wilman, R. J., Miller, L., Jarvis, M. J., et al. 2008, MNRAS, 388, 1335
- Zarka, P., Girard, J., Tagger, M., & Denis, L. 2012, in SF2A-2012 Proc. Annu. Meet. French Soc. Astron. Astrophys., 687–694