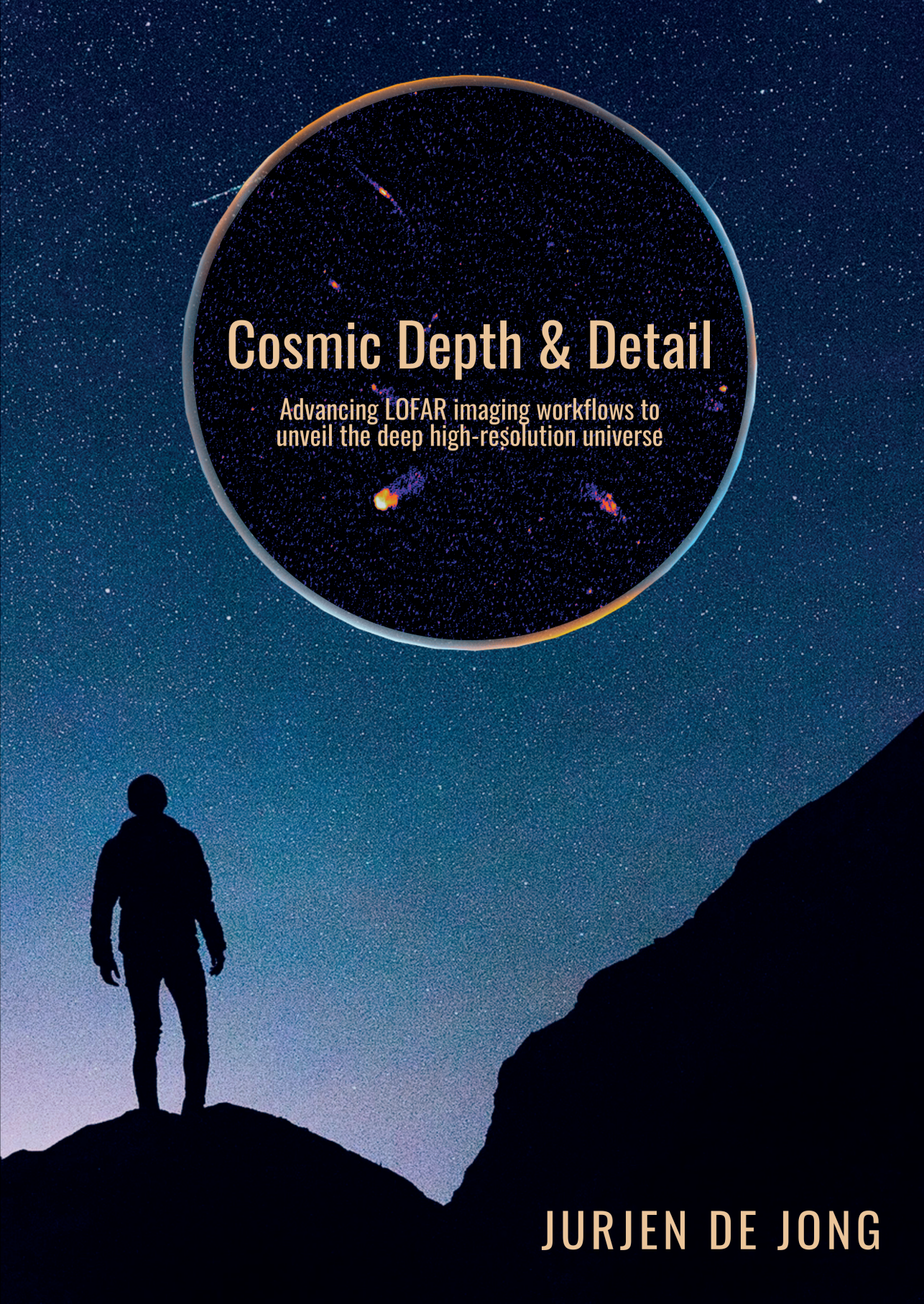

Cosmic Depth & Detail
Advancing LOFAR imaging workflows to
unveil the deep high-resolution universe
JURJEN DE JONG

Cosmic Depth & Detail

Advancing LOFAR imaging workflows to unveil the deep high-resolution universe

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op vrijdag 9 mei 2025
klokke 13:00 uur

door

Jurjen Michiel Gerrit Hendrik Jan de Jong

geboren te Terneuzen, Nederland
in 1993

Promotor:
Prof. dr. H.J.A. Röttgering

Co-promotores:
Dr. R.J. van Weeren
Dr. J.B.R. Oonk (SURF; Universiteit Leiden)

Promotiecommissie:
Prof. dr. K.J. Batenburg (CWI; Universiteit Leiden)
Dr. A.G.J. van Leeuwen (ASTRON)
Prof. dr. L.K. Morabito (Durham University)
Dr. A.R. Offringa (ASTRON; Rijksuniversiteit Groningen)
Prof. dr. I.A.G. Snellen
Prof. dr. P.P. van der Werf

ISBN: 978-94-6496-361-8

Cover: Background picture taken by photographer Joshua Earle, modified by author with help from Satu Nygren. The front cover features a person looking into a part of the deep high-resolution radio sky from the ELAIS-N1 deep field. This is an important sky area imaged and studied in this thesis. The back cover showcases the LOFAR LBA and HBA antennas, positioned beneath a sky featuring FRI and FRII type radio galaxies at 0.3″ resolution, overlaid on their lower resolution 6″ counterparts.

This work is part of the project CORTEX (NWA.1160.18.316) of the research programme NWA-ORC which is (partly) financed by the Dutch Research Council (NWO).

Printed by Gildeprint.
A copy of this thesis is available at `https://openaccess.leidenuniv.nl`.

© Jurjen de Jong, 2025

"Astronomy, as nothing else can do, teaches men humility."

– Arthur C. Clarke

Contents

1 Introduction **1**

1.1 Birth of radio astronomy . . . 1
1.2 Radio emission across the Universe . . . 2
1.2.1 Radio-loud active galactic nuclei . . . 2
1.2.2 (Pre-)merging galaxy clusters . . . 4
1.3 Catching radio waves . . . 6
1.3.1 Interferometry . . . 8
1.3.2 The Low Frequency Array . . . 9
1.3.3 Sensitivity vs. resolution with LOFAR . . . 10
1.4 Data processing with LOFAR . . . 12
1.4.1 Calibration at low frequencies . . . 12
1.4.2 Imaging large FoV . . . 13
1.5 Pipelines and computational cost . . . 13
1.5.1 Computational Bottlenecks . . . 14
1.5.2 CORTEX . . . 14
1.5.3 Wide and deep surveys . . . 15
1.6 This thesis . . . 15
1.7 Future prospects . . . 17
1.7.1 Towards a better understanding of AGN and galaxy clusters 17
1.7.2 LOFAR 2.0 . . . 17
1.7.3 FAIR community-wide collaboration . . . 18
1.7.4 Technological advancements . . . 18

2 Deep study of A399-401: Application of a wide-field facet calibration **19**

2.1 Introduction . . . 21
2.2 Observations and data reduction . . . 23
2.2.1 Data . . . 23
2.2.2 Calibration . . . 24
2.3 Results . . . 31
2.3.1 Diffuse emission . . . 31
2.3.2 AGNs . . . 37
2.4 Thermal and non-thermal scaling relations . . . 38
2.4.1 Mass-power relation . . . 39
2.4.2 Point-to-point analysis . . . 39
2.5 Discussion . . . 42

2.5.1 Origin of the radio bridge . . . 43
2.5.2 Radio haloes . . . 44
2.6 Conclusions . . . 46
2.A Computing recalibration . . . 47
2.B `DDF-Pipeline` versus recalibration for A399-401 . . . 48
2.C `Halo-FDCA` results . . . 49

3 Cosmic evolution of FRI and FRII sources out to $z = 2.5$ 53
3.1 Introduction . . . 55
3.2 Data . . . 57
3.2.1 Catalogues . . . 57
3.2.2 Source selection . . . 58
3.3 Redshift simulations . . . 61
3.3.1 Surface brightness and redshift relation . . . 61
3.3.2 Redshifting algorithm . . . 62
3.3.3 Source components . . . 63
3.3.4 FR classification . . . 64
3.4 Constructing radio luminosity functions . . . 66
3.4.1 $V_{\rm max}$ method . . . 66
3.4.2 Completeness corrections . . . 68
3.5 Results . . . 72
3.5.1 Local radio luminosity function . . . 72
3.5.2 Radio luminosity function up to $z = 2.5$ . . . 73
3.6 Discussion . . . 76
3.6.1 Energy losses and the RLF . . . 77
3.6.2 Comparing FRI/FRII RLFs . . . 78
3.6.3 Space density enhancements . . . 80
3.6.4 Future prospects . . . 81
3.7 Conclusions . . . 81
3.A Testing the RLF reliability . . . 82

4 Into the depths: Unveiling ELAIS-N1 with LOFAR's deepest sub-arcsecond wide-field images 87
4.1 Introduction . . . 89
4.2 Data description . . . 91
4.3 Calibration . . . 93
4.3.1 Initial Dutch calibration . . . 94
4.3.2 Direction independent calibration of full array . . . 97
4.3.3 Direction-dependent calibration of full array . . . 106
4.4 Wide-field imaging . . . 120

4.4.1 Method . . . 120
4.4.2 Facets and mosaics . . . 123
4.4.3 Computing costs . . . 125
4.5 Cataloguing . . . 126
4.6 Discussion . . . 129
4.6.1 RMS noise . . . 129
4.6.2 Smearing . . . 131
4.6.3 Astrometry . . . 132
4.6.4 Flux scale . . . 135
4.6.5 Sensitivity versus resolution . . . 135
4.7 Summary and conclusion . . . 137
4.A Towards an automated VLBI pipeline . . . 139
4.B Potential self-calibration issues . . . 141

5 Unlocking ultra-deep wide-field imaging with sidereal visibility averaging 145
5.1 Introduction . . . 147
5.2 Sidereal visibility averaging . . . 150
5.2.1 Frequency and time axis . . . 150
5.2.2 Algorithm . . . 151
5.3 Data . . . 154
5.4 Results . . . 156
5.4.1 Image quality . . . 156
5.4.2 Computing resources . . . 159
5.5 Discussion . . . 160
5.5.1 Image quality vs. compute resources . . . 160
5.5.2 Precession, nutation, and aberration . . . 162
5.5.3 Doppler shifts . . . 163
5.6 Summary and conclusion . . . 164

6 Advanced strategy for deep sub-arcsecond wide-field facet calibration with LOFAR 167
6.1 Introduction . . . 169
6.2 Data processing framework . . . 171
6.3 Automated long-baseline self-calibration . . . 174
6.3.1 Calibrator selection . . . 174
6.3.2 Solution interval metric . . . 176
6.3.3 Automatic self-calibration assessment . . . 178
6.4 Improved DD calibration and imaging strategy . . . 182
6.4.1 Dutch station calibration . . . 182

6.4.2 International station calibration . . . 186
6.4.3 1.2″ facet subtraction . . . 190
6.4.4 Final calibration refinement . . . 191
6.4.5 Final imaging . . . 195
6.5 Data . . . 195
6.6 Results . . . 196
6.6.1 Image quality improvements . . . 197
6.6.2 Computing costs . . . 198
6.7 Discussion . . . 204
6.7.1 Modular data processing . . . 204
6.7.2 Advanced decision making . . . 205
6.7.3 Data compression and I/O . . . 207
6.8 Summary and conclusion . . . 208

Bibliography **211**

Popular scientific summary **235**

Populairwetenschappelijke samenvatting **245**

List of publications **255**

Curriculum Vitae **259**

Acknowledgements **263**

1 Introduction

1.1. Birth of radio astronomy

For most of our history, visible light was the only form of radiation that we were aware of, not realizing it was just a small part of a vast electromagnetic spectrum. This changed when infrared radiation was discovered in the early 1800s (Herschel, 1800; Rowan-Robinson, 2013). Over time, new regions of the electromagnetic spectrum were uncovered, providing valuable insights not only into terrestrial phenomena but also into extraterrestrial processes. Among these newly explored regions was radio, which turned out to open a new window into understanding the Universe.

As with many discoveries, the first detections of radio waves of extraterrestrial origin were made by 'accident'. Early reports of mysterious 'hissing' noises at radio frequencies were documented in the 1920s and 1930s (e.g. Oswald, 1930; Arakawa, 1936), which were later identified as being caused by solar activity (e.g. Heightman, 1936; Ham, 1975; Kellermann et al., 2020). The first confirmed detection of radio emission from an extraterrestrial source came in 1933, when physicist and engineer Karl Jansky, while investigating interference in telecommunications, identified signals originating from the centre of the Milky Way at a frequency of 20.5 MHz (Jansky, 1933). Inspired by this discovery, Grote Reber – an engineer, 'amateur' radio astronomer, and radio operator – confirmed Jansky's source of radio emission at 162 MHz. He also detected radio signals from the Andromeda Galaxy, as well as from other areas in the sky, including the Cygnus and Cassiopeia regions (Reber, 1940, 1944). The advancements in radar technology during the Second World War inspired the development of radio astronomy in the postwar era (e.g. Sullivan, 2009; Elbers, 2015; Kellermann et al., 2020). It was therefore after the war that interest in radio astronomy grew, with more engineers and astronomers dedicating their efforts

to capturing and studying signals from the radio sky.

1.2. Radio emission across the Universe

The next most notable discoveries in the radio were the brightest radio sources in the sky, which are grouped into the so-called 'A-team'. This group includes radio galaxies such as Cygnus A, Virgo A, Hercules A, and Centaurus A, as well as supernova remnants Taurus A, Cassiopeia A, and Puppis A (e.g. Hey et al., 1946; Bolton & Stanley, 1948b,a; Mills, 1952b,a; Baade & Minkowski, 1954). These detections sparked the interest in conducting large-scale surveys, beginning with the Cambridge catalogues (Shakeshaft et al., 1955; Edge et al., 1959; Pilkington & Scott, 1965), and eventually leading to large surveys utilizing more advanced instruments, such as the Faint Images of the Radio Sky at Twenty centimetres (FIRST; Becker et al., 1995), the Westerbork Northern Sky Survey (WENSS; Rengelink et al., 1997), the NRAO VLA Sky Survey (NVSS; Condon et al., 1998), the Galactic and Extra-Galactic All-Sky MWA Survey (GLEAM; Wayth et al., 2015) the TIFR GMRT Sky Survey (TGSS; Intema et al., 2017), and the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al., 2017).

Over time, various discoveries were made at radio frequencies as well, including discoveries of the 21-cm line (Ewen & Purcell, 1951), quasars (Hoyle & Fowler, 1963; Schmidt, 1963), the cosmic microwave background (CMB; Penzias & Wilson, 1965), rapidly rotating neutron stars (pulsars; Hewish et al., 1968), radio haloes (Willson, 1970), fast radio bursts (FRBs; Lorimer et al., 2007), brown dwarfs (Berger et al., 2001), and more. While each of these objects deserves its own detailed description, we highlight in the following subsections only the background of objects most relevant to the work presented in this thesis.

1.2.1. Radio-loud active galactic nuclei

Among the brightest compact sources that were discovered in the early 1960s were a class of sources called 'quasi-stellar' (e.g. Hoyle & Fowler, 1963), which became later known as 'quasars' (e.g. Schmidt, 1970). After linking these objects to optical counterparts with known redshifts (e.g. Matthews & Sandage, 1963), it turned out that these sources corresponded to distant galaxies, rather than being from stellar origin (Schmidt, 1963, 1968). Eventually, the widely accepted explanation for the origin of these objects converged to one in which large amounts of material get pulled towards the black hole at the core of galaxies, forming a rotating hot accretion disk around it, from where vast amounts of energy are emitted across the electromagnetic spectrum (e.g. Lynden-Bell, 1969; Netzer, 2013; Padovani et al., 2017). These regions are now known as active galactic nuclei (AGN). The primary mechanism

behind the radio emission from galaxies with AGN is charged particles spiralling at relativistic speeds through magnetic fields, generating non-thermal synchrotron radiation.

When AGN emit a significant portion of their energy through radio waves they are called radio-loud AGN (RLAGN; e.g. Peacock et al., 1986). These types of objects are typically associated with jet-like structures originating from their cores (e.g. Padovani et al., 2017). These jets are energetic outflows extending from kpc up to Mpc scales, far from their host galaxies (e.g. Northover, 1973; Longair et al., 1973; Willis et al., 1974; Bridle & Perley, 1984; Oei et al., 2022, 2024a), and inject vast amounts of energy into the intergalactic medium (IGM) and intra-cluster medium (ICM). This affects galaxy formation within the jet's environment, since it can, for instance, prevent gas from cooling and collapsing to form new stars (e.g. Croton et al., 2006; Nesvadba et al., 2008; Fabian, 2012; McNamara & Nulsen, 2012; Dubois et al., 2013; Tremblay et al., 2016; Morganti, 2017).



Figure 1.1: Example of an FRI (*left panel*) and FRII (*right panel*), indicating the host galaxy and their jets. Images are taken from Hardcastle & Croston (2020).

Based on their morphologies, RLAGN with jets are separated into two different main morphological classes: FRI and FRII radio galaxies (Fanaroff & Riley, 1974). The FRI type has the brightest region near the centre and gradually fades out towards the edge (core-bright), while the FRII type has its brightest hotspots at the edges of the jets (edge-bright). An example of an FRI and FRII are shown in Figure 1.1. It is currently thought that the distinction between these two types of sources stems from the power of their radio jets in combination with the environments they inhabit (e.g. Bicknell, 1995; Kaiser & Best, 2007; Mingo et al., 2019). The FRIs correspond to jets that fail to push through the dense environments in which they

live, causing the jets to diffuse as they extend outward. In contrast, FRIIs have jets powerful enough to sustain their collimated jets while pushing through their environment. As a result, radio galaxies with FRI-type jets are typically found in denser environments like galaxy clusters, while galaxies with FRII-type jets tend to reside often in more isolated regions with abundant gas, providing the resources necessary for their jets to remain undisrupted (e.g. Bicknell, 1994; Croston et al., 2019).

The physical origin of the binary division of RLAGN into FRIs and FRIIs has frequently been challenged due to selection biases introduced by the sensitivity and resolution limits of radio telescopes (e.g. Best, 2009; Singal & Rajpurohit, 2014; Mingo et al., 2019; Magliocchetti, 2022; Mostert et al., 2024), as well as the identification of new populations beyond the binary FR-classification. These include FR0s, which are sources that are about 30 times more core-dominated than 'standard' FRIs due to a lack of extended emission (Baldi et al., 2015). There are also hybrid morphologies, which are a mix of both FRI and FRII morphologies (Gopal-Krishna & Wiita, 2000). These ongoing revisions and new classes challenge our understanding of radio galaxy jets, their physical origins, and evolution, leaving us with questions that remain partially or entirely unanswered. Most recent studies suggest that there is no link between different types of accretion and a radio galaxy having FRI or FRII type of jets (Mingo et al., 2022). However, could there be other mechanisms within the host galaxy that may still contribute to jet disruption? Or is jet disruption only driven by changes in the larger-scale environment of the galaxy? Do the FRI and FRII populations evolve differently as a function of redshift and radio power? How does this connect to our understanding of the evolution of galaxies or, on an even larger scale, the cosmic web? These questions highlight the need for multi-wavelength data, including large samples of FRI and FRII radio galaxies from deep radio observations, using different resolutions to strike a balance between surface brightness sensitivity and the ability to resolve their jets. We touch upon this in Chapter 3.

1.2.2. (Pre-)merging galaxy clusters

FRIs and FRIIs with sizes $\geq$ 0.7 Mpc (e.g. Dabhade et al., 2020)[1] are typically classified as giant radio(-loud) galaxies (GRGs). These objects can serve as tracers of the large-scale structure of the cosmic web in our Universe (e.g. Malarecki et al., 2015; Oei et al., 2024b). At even larger scales, galaxy clusters are another important tracer of the large-scale structure. These clusters of galaxies can span several Mpc and contain masses up to $10^{15}M_{\odot}$, making them the largest and most massive

[1]Or in other studies with size thresholds $\geq$ 1 Mpc (e.g. Lara et al., 2001).

gravitationally bound structures in the Universe. It is therefore that the formation, growth, and merging of galaxy clusters, play an important role in the evolution of the large-scale structure of our Universe (e.g. Springel et al., 2005; Kravtsov & Borgani, 2012).

The mergers of galaxy clusters are very energetic events, releasing ~10^{65} ergs (e.g. Markevitch & Vikhlinin, 2007). When the ICMs of merging clusters collide, the gas is compressed and heated through shocks and adiabatic compression, raising the temperature to levels that produce X-ray emission (e.g. Böhringer & Werner, 2010). During mergers, shocks and turbulence accelerate cosmic-ray electrons to relativistic speeds, generating non-thermal synchrotron emission through interactions with magnetic fields, which is most prominently detected at low radio frequencies (e.g. van Weeren et al., 2021).

The two main types of extended radio emission associated with galaxy clusters have been classified as radio (mini-)haloes and relics (e.g. Willson, 1970; Jaffe & Rudnick, 1979; Burns et al., 1992; Brunetti et al., 2009; Cassano et al., 2010; van Weeren et al., 2013, 2017; Bonafede et al., 2022). Radio haloes are often located at the centres of galaxy clusters and span sizes on the order of Mpcs. Their origin is currently explained by two main models: turbulent reacceleration of cosmic-ray electrons within the ICM (De Young, 1992; Brunetti et al., 2001; Gitti et al., 2002), or secondary electrons produced by hadronic collisions (e.g. Dennison, 1980; Blasi & Colafrancesco, 1999). Relics, on the other hand, are arc-shaped structures extending up to 2 Mpc and are found more often at the outskirts of clusters. They are believed to originate from collisionless shocks that accelerate cosmic-ray electrons (Blandford & Eichler, 1987; Ryu et al., 2003), or from the reacceleration of pre-existing fossil cosmic-ray electrons (e.g. Markevitch et al., 2005; van Weeren et al., 2016a), which are relic particles from past energetic events.

More recently, the first diffuse radio bridges connecting two pre-merging galaxy clusters have been discovered as well (Govoni et al., 2019; Botteon et al., 2020b). This is a direct detection at radio frequencies of filamentary structure between galaxy clusters, as part of the large-scale structure of the cosmic web. Images with the well-known pre-merging clusters Abell 399 and Abell 401 with corresponding radio bridge in the middle are shown in Figure 1.2. The origin of the non-thermal radio emission from radio bridges between pre-merging clusters has been the subject of debate in recent years, but the latest studies seem to favour a turbulent reacceleration model (Brunetti & Vazza, 2020; Nunhokee et al., 2021; Pignataro et al., 2024), similar to the most favoured mechanism for generating radio haloes. Some of the currently open questions related to radio bridges are: Is turbulent reacceleration the only and best explanation for radio bridges? How common are (large-scale) radio bridges? What is their connection with the evolution of the large-scale structure

of our Universe? To answer these questions, very deep observations are required at low frequencies, preferably combined with X-ray observations to compare thermal and non-thermal emissions, since the reacceleration of electrons occurs in similar regions as where the ICM is heated (e.g. Brunetti & Jones, 2014). We explore this in Chapter 2.

1.3. Catching radio waves

Although radio astronomy began with Jansky's work at very low frequencies, using equipment originally developed for early radio communication, the field later shifted its focus to higher frequencies, as more advanced telescopes were designed and built at these frequencies. Among these are for instance the Westerbork Synthesis Radio Telescope (WSRT Baars et al., 1973), the Very Large Array (VLA; Thompson et al., 1980), or the more recently constructed Atacama Cosmology Telescope (ALMA; Swetz et al., 2011). At higher frequencies, astronomers have the opportunity to study radio emissions from molecular clouds (e.g. Yusef-Zadeh, 2012), protoplanetary disks (e.g. Villenave et al., 2020), star-forming regions (e.g. Rosero et al., 2016), among other astronomical objects which emit thermal emission.

We already hinted that detecting non-thermal radio emission, such as associated with the jets discussed in Section 1.2.1 or the radio bridges mentioned in Section 1.2.2, requires telescopes specifically designed to observe low-frequency radio waves. Some low-frequency telescopes, like for example the Cambridge Interferometer (Ryle & Hewish, 1955) or the Ukrainian T-shaped Radio Telescope (UTR; Braude et al., 1978),[2] were operational for many decades. However, these instruments were limited by their resolution. This is because the resolution improves as frequency increases for a given telescope size, according to the following relation

$$\theta \propto \frac{\lambda}{D_{\rm tel}}, \tag{1.1}$$

where θ is the resolution, λ is the wavelength, and $D_{\rm tel}$ the diameter of the telescope (or baseline as we will see in the next subsection). This implies that improving the ability of radio telescopes to resolve more objects and substructures at low frequencies requires more advanced designs and techniques.

[2]Sadly, this telescope was, during my PhD, heavily damaged during the Russian invasion of Ukraine in 2022.

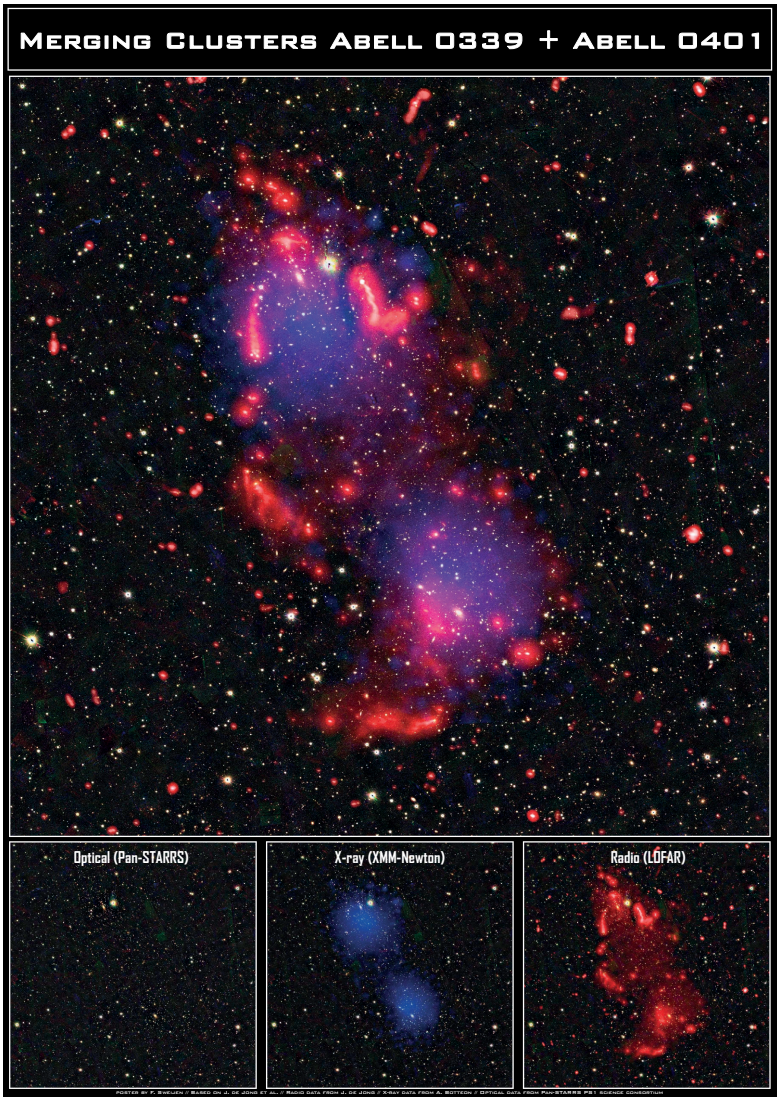


Figure 1.2: Image of pre-merging clusters Abell 399-401 at optical (Pan-STARRS; Chambers et al., 2016), X-ray (XMM-Newton; Sakelliou & Ponman, 2004), and detected at radio frequencies at 144 MHz (de Jong et al., 2022). Each galaxy cluster – one in the north and the other in the south – exhibits a radio halo, which largely overlaps with X-ray emission. This indicates that these clusters have experienced a galaxy cluster merger in the past. At the same time, there is also fainter emission between the galaxy clusters represented by a radio bridge, showing that these clusters are also in the process of merging with each other. Radio galaxies with their optical counterparts are also depicted. This image utilises the data analysed in Chapter 2 and was designed by F. Sweijen.

1.3.1. Interferometry

A method to improve resolution at lower frequencies is through interferometry,[3] which is a technique where signals received by two or more widely spaced dishes or antennas are correlated to extract information from objects on smaller angular scales (e.g. Ryle & Vonberg, 1946; Kellermann & Moran, 2001; Quirrenbach, 2009). This relies on the principle of interference, where radio waves from the same source in the sky, arriving at different antennas, are combined in such a way that their constructive or destructive interference patterns at different angular scales are used to extract information about the source. The correlated signals between antenna pairs, called 'visibilities', are represented by complex numbers (with a phase and amplitude) and correspond to baselines, which represent the distances between each pair of antennas or dishes. These visibilities with their corresponding baseline coordinates (commonly expressed in the uv-plane or uvw-plane), are samples of the Fourier transforms of the sky brightness. This indicates that they can be mapped into image space using (fast) Fourier transforms (FFT; Cooley & Tukey, 1965). In this way, interferometry effectively creates a virtual telescope with a maximum diameter equal to the longest baseline, enabling significantly higher resolution than any single-dish telescope could achieve on its own. In this context, $D_{\rm tel}$ in Equation 1.1 refers to the length of the longest baseline.

The challenge for interferometers, compared to single-dish telescopes, is their limited coverage of the uv-plane, which affects the image reconstruction process. A straightforward way to improve this coverage is through making use of Earth's rotation, a technique known as aperture synthesis (Ryle & Hewish, 1960). This allows telescopes to observe the same region of the sky from different angles over time, effectively sampling different parts of the uv-plane at various frequencies. The incomplete sampling of the uv-plane introduces also artefacts, typically from side-lobes in the point spread function (PSF), which can distort or obscure (fainter) features in the image. Mitigating this effect is typically achieved during the conversion of visibilities into more accurate images using (variations of) the CLEAN algorithm (Högbom, 1974; Clark, 1980). This algorithm iteratively deconvolves the radio image by identifying and subtracting point sources starting with the initial 'dirty' image, progressively reconstructing the true sky brightness distribution by adding the subtracted components in each iteration to a model image. The final components in the model image are then convolved with a Gaussian beam and added back to the residual noise background, resulting in a cleaner image with reduced side-lobe interference. This algorithm can be further extended with various enhancements, such as multi-scale CLEAN, which accounts for emission structures at multiple spatial scales (e.g. Cornwell, 2008).

[3]Note that this technique can be used at non-radio wavelengths as well.

1.3.2. The Low Frequency Array

Since the early 2000s, advancements in technology, financial resources, and renewed interest have enabled the development of state-of-the-art low-frequency interferometers with longer baselines and higher sensitivities. These include the Giant Metrewave Radio Telescope (GMRT; Swarup et al., 1991), the Murchison Widefield Array (MWA; Tingay et al., 2013), the Long Wavelength Array (LWA; Ellingson et al., 2009), the upcoming Square Kilometer Array (SKA; Carilli & Rawlings, 2004), and the International Low Frequency Array (LOFAR) Telescope (ILT; van Haarlem et al., 2013). LOFAR is the primary instrument that serves us in the next two chapters of this thesis to address some of the outstanding scientific questions discussed in Sections 1.2.1 and 1.2.2, motivating us later in the thesis to explore ways to enhance its capabilities.

LOFAR is a unique instrument that detects low frequency radio waves between 10 and 80 MHz with its low-band antennas (LBAs) and between 110 and 240 MHz with its high-band antennas (HBAs). It makes use of phased-array technology (e.g. Herd & Conway, 2016), where signals from many individual dipole antennas are combined electronically. This allows LOFAR to 'steer' its focus across the sky by introducing delays in the data, without physically moving its antennas. The absence of moving parts, combined with LOFAR's focus on detecting long wavelengths, reduces the need for strict mechanical engineering tolerances compared to telescopes designed for higher frequencies or with moving parts. This enables a cost-effective design and construction of a large number of antennas throughout Europe. LOFAR has a dense concentration of core and remote stations in the Netherlands and international stations across many European countries. The current longest baseline stretches 1980 km between Birr (Ireland) and Łazy (Poland). An overview of the station layout across Europe is presented in Figure 1.3.

The size of the field of view (FoV) of a LOFAR station is given by

$$\mathrm{FoV} \propto \left(\frac{\lambda}{D_{\mathrm{stat}}}\right)^2,$$

where D_{stat} is the station diameter. This relation implies that LOFAR is able to obtain a large FoV in the order of multiple square degrees. In fact, the FoV achievable using only the Dutch stations at 150 MHz is more than 10 deg^2 (e.g. Shimwell et al., 2017), while an international station is larger and therefore reaches at the same frequencies, combined with the Dutch stations, a FoV of about 6 deg^2 (e.g. Sweijen et al., 2022c). LOFAR's capability to image large FoVs makes it a great instrument to perform surveys over large sky areas, catching many sources at once.

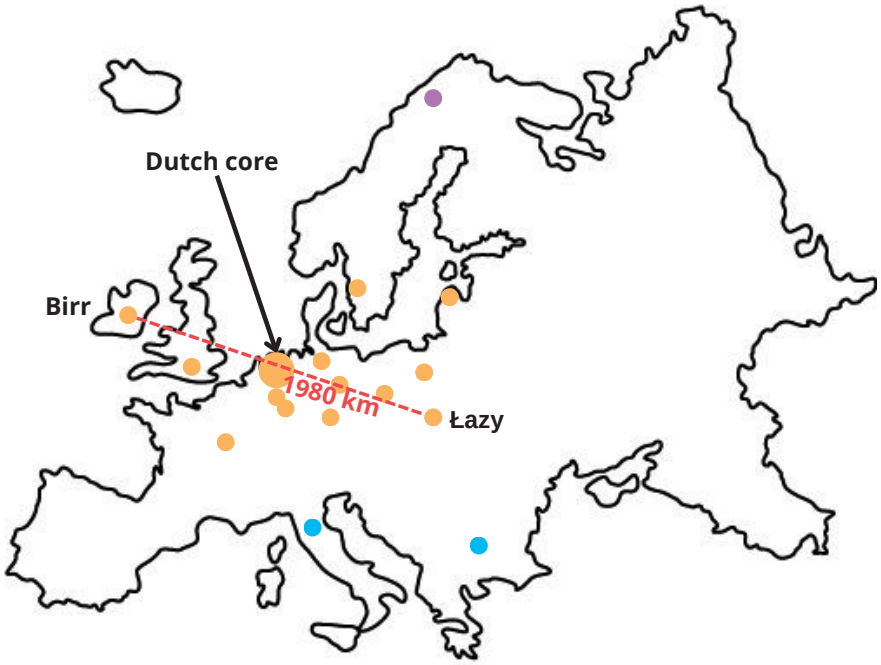


Figure 1.3: LOFAR stations across Europe. The orange circles indicate stations currently active and from which data are utilised in this thesis, with the Dutch LOFAR core stations represented by a larger circle. The blue circles mark LOFAR stations currently being constructed and soon to be added to the international LOFAR network. The purple circle is the Kilpisjärvi Atmospheric Imaging Receiver Array (KAIRA; McKay-Bukowski et al., 2015), which is a LOFAR station in Finnish Lapland that operates independently of the main LOFAR network. We have marked the currently active longest baseline of 1980 km between Birr and Łazy with a red dashed line.

1.3.3. Sensitivity vs. resolution with LOFAR

The Dutch LOFAR stations provide a $6''$ resolution and a sensitivity of approximately 70 μJy beam^{-1} for one 8 hours observation (Shimwell et al., 2019). By incorporating the international stations, the resolution can improve to $0.3''$, along with an improved sensitivity by about a factor of 2, due to the larger collecting area of the international stations (van Haarlem et al., 2013; Varenius et al., 2015). This sensitivity improvement refers to the point source sensitivity, as imaging at lower resolutions still performs better in tracing extended low surface brightness emission from large-scale objects (e.g. radio haloes, relics, or bridges), which may otherwise be hidden in the background noise in high-resolution images. Intermediate arcsecond-scale resolution offers therefore a valuable alternative as well, striking a balance between resolution, sensitivity, and computational cost (e.g. Ye et al.,

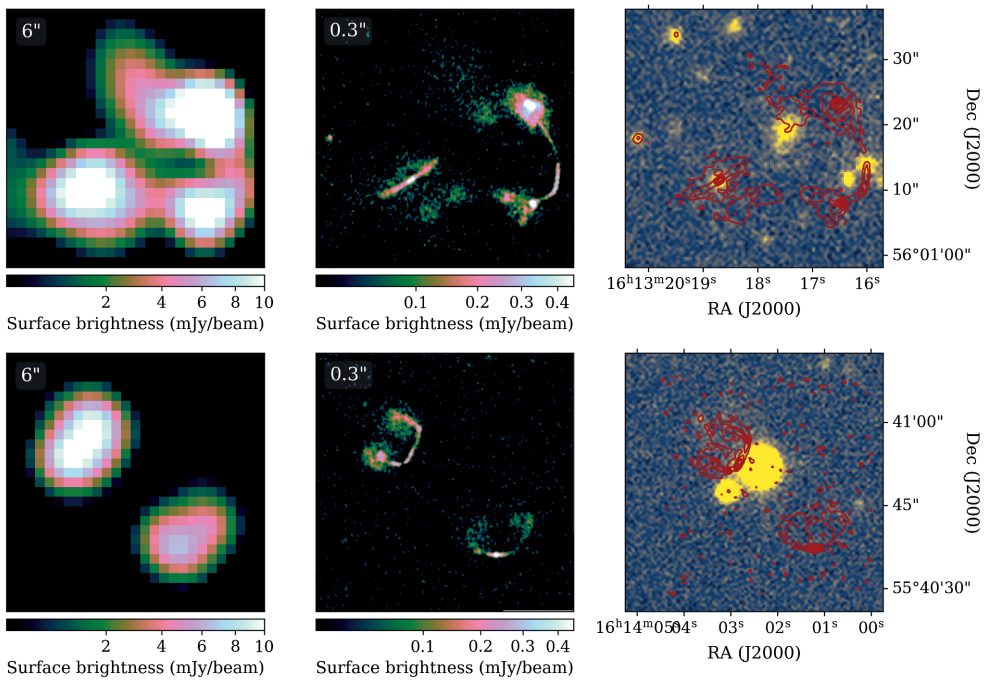


Figure 1.4: Demonstrating the power of resolving sources at higher resolution. Both panels are from left to right postage stamp images from the following maps: 6″ resolution radio map from (Shimwell et al., 2025), 0.3″ resolution radio map from de Jong et al. (2024), and optical map from DESI (Dey et al., 2019) with 0.6″ radio contours from de Jong et al. (2024) on top. *Upper panel:* Although the 6″ resolution is sensitive to capture low surface brightness emission, it does in this case not allow for accurate morphological classifications of RLAGN in FRI or FRII types, while on 0.3″ resolution the galaxies are highly resolved. *Lower panel:* While at 6″ the sources are difficult to classify and there is no clear optical counterpart, do the high-resolution counterparts show that these are individual bent RLAGN jets.

2024).

For the studies of RLAGN, low surface brightness sensitivity is advantageous for detecting diffuse jets of FRIs when the object is resolved and classifiable at a given resolution (e.g. Mingo et al., 2019). Yet, in cases such as presented in Figure 1.4, where we compare detections of neighbouring RLAGN at 6″ and 0.3″, it is clear that while some of the RLAGN are detected at 6″ resolution, they are not enough resolved to successfully classify these sources. Jurlin et al. (2024) recently illustrated the benefit of using the multi-resolution ability of LOFAR, by revealing recent activity of RLAGNs close to their core with the international long baselines, while uncovering older emissions from past activity using the Dutch shorter baselines. In the context

of galaxy clusters, van Weeren et al. (2024) demonstrated that combining multiple resolutions from LOFAR offers a more comprehensive understanding of the various processes occurring in and around galaxy clusters as well. While these examples illustrate the need to balance resolution and sensitivity for specific science cases, it is also crucial to emphasise that processing the data to produce these images does not come for free, as will be outlined in the next section.

1.4. Data processing with LOFAR

The combination of high resolution, high sensitivity, and large FoV positions LOFAR as a powerful instrument for producing deep, high-resolution wide-field images of the low-frequency radio sky. However, collecting so much information in a single observation also comes at a cost. Hence, efficient and robust processing pipelines that calibrate and image the data are important to efficiently map the raw visibility data to a final science-ready image.

1.4.1. Calibration at low frequencies

For standard data processing, several effects need first to be addressed. These include calibration for different systematic effects, such as clock offsets, polarisation offsets, ionospheric effects, beam errors, and Faraday rotation (e.g. van Weeren et al., 2016b; de Gasperin et al., 2019a; Mevius, 2018). At low frequencies, the main calibration challenges are coming from direction-dependent effects (DDEs). These are variations in observed signals caused by changes in direction across the FoV of a telescope. At the low frequencies observed by LOFAR, these variations are particularly introduced by the ionosphere. This is a layer in Earth's atmosphere filled with charged particles that interact with incoming radio waves, especially at low frequencies. These interactions cause phase delays, which vary across the sky and change over time, leading to image distortions and positional shifts. Multiple software packages have been developed to deal with this issue, such as `DP3`[4] (van Diepen et al., 2018a; Dijkema et al., 2023), `SPAM`[5] (Intema et al., 2009), `Sagecal`[6] (Kazemi et al., 2011), `KillMS`[7] (Tasse, 2014a,b; Smirnov & Tasse, 2015), and `facetselfcal`[8] (van Weeren et al., 2021). To account for the DDE variation across the FoV, the corrections from these packages are typically applied within a facet-based approach. This requires the selection and calibration of a number of calibrator sources across

[4] `https://dp3.readthedocs.io`
[5] `http://www.intema.nl/doku.php?id=huibintema:spam:start`
[6] `https://github.com/nlesc-dirac/sagecal`
[7] `https://github.com/saopicc/killMS`
[8] `https://github.com/rvweeren/lofar_facet_selfcal`

the FoV with enough signal at the largest baselines. The field is then divided into facets using a Voronoi tessellation (Schwab, 1984; van Weeren et al., 2016b; Tasse et al., 2018), wherein the calibration corrections are applied to smaller sky areas (the facets) corresponding to the nearest calibrator source.

1.4.2. Imaging large FoV

Having done all the calibration, the corrected visibilities need to be converted to images using Fourier transforms (typically in the form of FFTs). Although LOFAR's large FoV is an advantage to capturing many sources at once, it also implies that it is not possible to estimate the sky to be a flat plane, as for imaging smaller FoVs. So, to incorporate curvature of the Earth and the sky, it is essential to account for the so-called 'w-term', which describes the non-coplanarity of an interferometer (e.g. Perley, 1999; Cornwell, 2008). This additional term makes applying Fourier transforms computationally more expensive. Fortunately, several efficient imaging software packages are around to effectively take into account the curvature to generate wide-field images. One of these is `WSClean`[9] (Offringa et al., 2014), which uses w-stacking to group visibilities with similar w-values into different layers in the Fourier domain. Another software package is `DDFacet`[10] (Tasse et al., 2018), which uses w-kernels to correct for the sky curvature within facets.

1.5. Pipelines and computational cost

For LOFAR data with only Dutch stations, the `DDF`-pipeline[11] was developed (Shimwell et al., 2019; Tasse et al., 2021). This pipeline calibrates and processes visibilities automatically up to a fully DDE-corrected wide-field image at $6''$ resolution. This has resulted in the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al., 2017, 2019, 2022; Williams et al., 2019), which has nearly completed imaging the entire northern hemisphere. Yet, a fully automated pipeline for direction-dependent calibration and wide-field imaging data that includes both Dutch and international LOFAR stations does currently not exist, despite having a general outline of the necessary steps (Morabito et al., 2022a, 2025a; Sweijen et al., 2022c; Ye et al., 2024; de Jong et al., 2024). This is because reducing this data has been experimental until now, and the high computational cost associated with full data processing must first be managed before scaling up.

[9] `https://wsclean.readthedocs.io/en/latest`
[10] `https://github.com/saopicc/DDFacet`
[11] `https://github.com/mhardcastle/ddf-pipeline`

1.5.1. Computational Bottlenecks

Prior to the work presented in this thesis, the computational cost for reducing a single 8 hours LOFAR observation for sub-arcsecond wide-field imaging summed up to about ~250,000 CPU core hours (Sweijen et al., 2022c).[12] This large cost is due to the substantial visibility data volumes, which for one LOFAR observation is about 16 terabytes (TB). Using data compression techniques this can be brought down to about 4 TB (Offringa, 2016), or alternatively being averaged with baseline-dependent averaging (BDA; Cotton, 1986, 2009; Wijnholds et al., 2018).[13] This data volume cannot be further reduced, since it is essential to contain the full time and frequency resolution for the final imaging to avoid time and bandwidth smearing. In particular, the final imaging, despite having efficient software (as outlined in Section 1.4.2), has proven to be the main computational bottleneck, since this is where the large number of visibilities with calibration solutions applied have to be iteratively converted to about 8 billion image pixels (Sweijen et al., 2022c; de Jong et al., 2024). Given these large computational costs associated with the processing of LOFAR data for sub-arcsecond wide-field imaging, it is essential to optimise our algorithms and data reduction strategies for greater efficiency.

1.5.2. CORTEX

As part of the Centre for Optimal Real-Time Machine Studies of the Explosive Universe (CORTEX) project,[14] we aim to address the computing and automation challenges for processing LOFAR data for deep sub-arcsecond wide-field imaging. First of all, our goal is to further automate the processing of LOFAR high-resolution data, reducing the need for human intervention. This includes automatically selecting the best calibrator sources and parameters for correcting DDEs during LOFAR data reduction but also validating the selection and calibration. This can for instance be done by finding ways to assess the signal-to-noise (S/N) for different calibrators at the longest baselines directly from visibility data but also by incorporating state-of-the-art neural networks to assess image quality post-calibration and fine-tune parameters accordingly, as will be discussed in Chapter 4. Second of all, to reduce the computational costs of processing large datasets, it is important to revisit the current data reduction strategy and develop techniques to reduce the data volume before imaging, as will be discussed in Chapters 5 and 6.

[12] This means that it would take about 7 years to process this data on a typical 4-core laptop.

[13] Dysco compression does currently not work in combination with BDA. However, alternative compression techniques are currently being explored and may be able to combine lossy or loss-less data compression with BDA compression (e.g. Dodson et al., 2024).

[14] `https://www.nwo.nl/en/projects/nwa116018316`

1.5.3. Wide and deep surveys

By developing an automatic robust pipeline and by bringing down the computational cost, not only surveys across large sky areas (similar to LoTSS) are enabled but also ultra-deep imaging at high resolutions, deeper than what can be achieved at 6″ resolution. At 6″ resolution confusion noise – additional background noise caused by blending faint, unresolved sources (e.g. Condon et al., 2012) – becomes a limiting factor for reaching sensitivities below 10 μJy beam^{-1} (e.g. Sabater et al., 2021). Having beams 400 times smaller in area with LOFAR's 0.3″ resolution will therefore mitigate this issue. For instance, having hundreds of hours of integration time already available of the ELAIS-N1 deep field in the LOFAR long-term archive (LTA), gives us the opportunity to produce an ultra-deep 2.5×2.5 degrees image with sensitivities in the order of μJy beam^{-1}.

1.6. This thesis

In this thesis, we aim to leverage LOFAR's exceptional sensitivity (depth) and resolution (detail) to study our Universe at low frequencies, while at the same time pushing its capabilities to open new windows for scientific exploration. We begin by demonstrating LOFAR's capability for creating deep sensitive wide-field images of large-scale structures through the production of the deepest image of a known radio bridge, aiming to explain its origin. We then utilise at smaller scales the resolution and sensitivity of the Dutch LOFAR stations with the goal of answering open questions about the cosmic evolution of extended RLAGN morphologies. Although this leads to new insights, it highlights the resolution limitations of Dutch-only observations as well, motivating us to enhance data processing techniques for creating deep sub-arcsecond wide-field images in the second half of this thesis. This effort results in the deepest sub-arcsecond wide-field image to date, along with new techniques to push our work further with reduced need for human intervention while saving computational cost. The chapters in this thesis are outlined as follows:

- In **Chapter 2**, we apply an advanced calibration technique to create the deepest wide-field image of a known radio bridge using LOFAR data. This enables us to investigate its particle (re)acceleration mechanisms and explore its origin. Through a point-to-point comparison of radio and X-ray data, our findings indicate that turbulent (second-order Fermi) reacceleration is the most likely cause of this phenomenon We suggest a similar origin for the radio haloes of the individual clusters as well.

- In **Chapter 3**, we study the cosmic evolution of FRI and FRII radio galaxies up to redshift $z = 2.5$ by selecting classified FRI and FRII radio sources

detected by LOFAR. This provides us with 100 times more FRIs above $z = 0.3$ compared to previous work. With radio luminosity functions we find space density enhancements towards higher reshifts for the most powerful radio galaxies, while we observe a space density decrease over redshift for less powerful sources. The former can be physically explained by a higher gas density in an earlier universe, while the latter is likely related to selection biases that can only be overcome by improving the spatial resolution and depth of our observations.

- In **Chapter 4**, we create the deepest sub-arcsecond 2.5×2.5 degrees image with LOFAR of the ELAIS-N1 deep field with a central RMS noise of about 14 μJy beam^{-1}. This is accomplished by combining four different observations of the same pointing centre, enhancing the direction-independent calibration, and further automating the direction-dependent calibrator selection. While computing costs are reduced, they remain significant, underscoring the need to minimise these costs for deeper imaging or wide surveys.

- In **Chapter 5**, we address a large fraction of the cost associated with deep imaging by revisiting and improving the technique we call 'sidereal visibility averaging'. This method makes use of the fact that baseline coordinates repeat every sidereal day, which allows us to average visibilities at similar baseline coordinates when combining multiple observations. This reduces for instance the computing costs for imaging by a factor of 10 when producing an ultra-deep image with 500 hours of LOFAR data. Although the method has proven to be reliable, we also stress the fact that it is important to take into account precession and aberration, which introduce offsets in the baseline coordinates between observations.

- In **Chapter 6**, we address remaining bottlenecks that prevent the processing of a large number of observations of the same sky area for sub-arcsecond wide-field imaging with LOFAR. This includes image quality improvements, the introduction of a streamlined framework, a refined and further automated DDE calibration using neural networks, and the lowering of the total computing costs up to a factor of 4 compared to the work from Chapter 4. This work lays the foundation for processing and imaging hundreds of hours of existing LOFAR data of the ELAIS-N1 field, ultimately enabling the creation of a wide-field sub-arcsecond image with unprecedented depth in the order of a few μJy beam^{-1}.

1.7. Future prospects

LOFAR will remain for at least another decade a leading instrument to study radio frequency emission below 200 MHz. This is because its capability to achieve sub-arcsecond resolution will remain unmatched at these frequencies, even with the advent of the SKA (Carilli & Rawlings, 2004). As already mentioned in Section 1.5.3, having both high-resolution and high sensitivity will be the main way to beat confusion noise and probing the depths of our Universe. The work presented in this thesis, contributes to this promising future for LOFAR, as we continue to process and analyse increasing amounts of LOFAR data.

1.7.1. Towards a better understanding of AGN and galaxy clusters

Scientifically, LOFAR's current and upcoming deep (sub-)arcsecond data allows us to better separate radio emission due to AGN and star formation, even within the same galaxy, using brightness temperature measurements (e.g. Morabito et al., 2022b, 2025b). Furthermore, high-resolution views of more nearby radio galaxies give us the ability to unravel the detailed physics and dynamics of their jets as well (e.g. Mahatma et al., 2023; Mahatma, 2023; Jurlin et al., 2024). Moreover, sub-arcsecond spatial detail enables the possibility to more accurately measure radio galaxy sizes over large sky areas (e.g. Sweijen et al., accepted), helping us for example to enhance our understanding of the evolution of their jets on a more statistical basis. Similarly, we can with enough deep sub-arcsecond resolution data also correct for the selection biases identified in Chapter 3, and obtain a deeper understanding of how radio jet morphologies relate to the cosmic evolution at redshifts beyond $z = 0.8$. At larger scales, deep high-resolution images of the sky at 144 MHz may also provide unprecedented insights into emissions from various scales in galaxy clusters, from thin filaments constituting the tails of radio galaxies to diffuse threads embedded in the ICM (De Rubeis et al. in prep.).

1.7.2. LOFAR 2.0

The upcoming upgrade to LOFAR 2.0 will further improve LOFAR's capabilities. One of the main highlights of this upgrade is the synchronisation of the clocks of the Dutch stations. This will mitigate clock drifts for these stations and therefore improve calibration (de Gasperin et al., 2019b). The LOFAR 2.0 upgrade will also enhance the sensitivity in the LBA and enable simultaneous LBA and HBA observations (e.g. Edler et al., 2021). This will benefit studying objects at wider frequency bandwidth at the same time (spectral studies), but may also help to improve LBA high-resolution imaging (e.g. Morabito et al., 2016; Groeneveld et al., 2022), by uti-

lizing combined calibration at HBA and LBA to solve for the ionospheric-induced delays.

1.7.3. FAIR community-wide collaboration

As the complexity of our data reduction strategies increases, it will also be essential to unite efforts across the LOFAR community and across multiple disciplines. Tackling new challenges efficiently will require a collaborative approach, where skilled software developers play a key role in advancing the data reduction pipelines and making those more robust, while astronomers contribute to the development and help steer focus on achieving the scientific goals. These synergies will require thorough communication as data reduction strategies need to be optimised and computational bottlenecks need to be addressed, but the primary goal to push scientific boundaries should never be lost out of sight. Adhering also to FAIR principles – being Findable, Accessible, Interoperable, and Reusable – will make it easier for other researchers to find, access, and build upon future data and tools (Wilkinson et al., 2016; O'Toole & Tocknell, 2022), accelerating advancements in LOFAR sub-arcsecond imaging and studying the Universe across multiple wavelengths and domains.

1.7.4. Technological advancements

Looking further ahead, we can expect more implementations of code and algorithms for the use of GPUs, pushing the efficiency of our data processing. And artificial intelligence (AI) will not only play a bigger role in our daily lives but also continue to enhance our ability to process data and deepen our understanding of the Universe. Likewise, emerging technologies such as quantum computing may hold the potential to revolutionise how we live and work as well. However, despite the promises of these advancements, it is crucial to approach their implementation critically and carefully. Ethical considerations (e.g. Kop et al., 2023; Resnik & Hosseini, 2024) and energy consumption (e.g. Kruithof et al., 2023) must remain important as we improve our software and integrate new technologies into scientific research and beyond. Ultimately, the success of innovative solutions depends not only on their capabilities but also on the ability to use them wisely.

2

Deep study of A399-401: Application of a wide-field facet calibration

J.M.G.H.J. de Jong, R.J. van Weeren, A. Botteon, J.B.R. Oonk, G. Brunetti, T.W. Shimwell, R. Cassano, H.J.A. Röttgering, C. Tasse

Astronomy & Astrophysics, Volume 668, id.A107, 19 pp.

Abstract

Diffuse synchrotron emission pervades numerous galaxy clusters, indicating the existence of cosmic rays and magnetic fields throughout the intra-cluster medium. This emission is generated by shocks and turbulence that are activated during cluster merger events and cause a (re-)acceleration of particles to highly relativistic energies. Similar emission has recently been detected in megaparsec-scale filaments connecting pairs of pre-merging clusters. These instances are the first in which diffuse emission has been found outside of the main cluster regions. We aim in this work to examine the particle acceleration mechanism in the megaparsec-scale bridge between Abell 399 and Abell 401 and assess in particular whether the synchrotron emission originates from first- or second-order Fermi re-acceleration. We also consider the possible influence of active galactic nuclei (AGNs). To accomplish this, we use ~40 hrs LOw-Frequency ARray (LOFAR) observations with an improved direction-dependent calibration to produce deep radio images at 144 MHz with a sensitivity of $\sigma = 79$ μJy beam^{-1} at a $5.9'' \times 10.5''$ resolution. Using a point-to-point analysis, we investigate the correlation between the radio and X-ray brightness from which we can constrain the particle re-acceleration mechanism. Our radio images show the radio bridge between the radio haloes at high significance. We find a trend between the radio and X-ray emission in the bridge and the radio haloes, along with evidence for AGN particle injection and re-energised fossil plasma in the radio bridge and haloes. We argue that second-order Fermi re-acceleration is currently the most favoured process to explain the origin of the radio bridge and radio haloes. We also find indications for a scenario in which past AGN particle injection may have introduced significant scatter in the relation between the radio and X-ray emission in the bridge, but may also supply the fossil plasma needed for in situ re-acceleration.

2.1. Introduction

Structures in our Universe are growing hierarchically, with smaller systems merging to form larger structures. The largest gravitationally bound structures are galaxy clusters, and when these merge with each other, $\sim 10^{64}$ erg is released into the intracluster medium (ICM) on timescales of billions of years (Markevitch & Vikhlinin, 2007; Hoeft et al., 2008). The ICM is a diluted plasma that permeates the cluster volume and primarily emits thermal bremsstrahlung at X-ray wavelengths. Synchrotron radio emission has been observed in numerous clusters (see van Weeren et al., 2019, for a recent review). The presence of this emission indicates the existence of cosmic rays and magnetic fields in the ICM. The general consensus is that shocks and turbulence, generated during cluster merger events, cause the (re)acceleration of particles to highly relativistic energies (Brunetti & Jones, 2014). Recently, diffuse radio emission has also been detected between pairs of clusters at megaparsec (Mpc)-scales (Govoni et al., 2019; Botteon et al., 2020b). These so-called radio bridges might trace regions in which the gas is compressed during the initial phase (i.e., the pre-merger phase) of the collision between massive galaxy clusters. Radio observations of cluster bridges open new windows for studying the acceleration of cosmic rays in environments with a density that is lower than typical in clusters (Brunetti & Vazza, 2020). The detection of radio bridges also brings us closer to the detection and study of plasma conditions in the densest phase of the so-called warm-hot intergalactic medium (WHIM) (Vazza et al., 2019). However, because only a few bridges are known and only very few theory papers have been published about their possible origin, the investigation of the origin of the magnetic fields and cosmic rays in the radio bridges is still in an initial stage in these low-density environments.

Radio bridges associated with the pre-merging clusters Abell 1758N and Abell 1758S (A1758) at $z = 0.279$ (Botteon et al., 2018b, 2020b) and the pre-merging clusters Abell 399 and Abell 401 (A399-401) at $z = 0.072$ (Govoni et al., 2019) have been recently discovered. These radio bridges are between two comparable systems and were discovered with LOw Frequency ARray (LOFAR) observations at 144 MHz. Follow-up studies at different frequencies have been performed recently (Botteon et al., 2020b; Nunhokee et al., 2021). Moreover, Abell 1430 might have a radio bridge between two merging clusters (a main cluster and subcluster), but this has not been fully confirmed (Hoeft et al., 2021). Other types of radio bridges have been discovered between the Coma cluster and the NGC4839 group ($z \approx 0.0231$) at 346 MHz (Kim et al., 1989) and 144 MHz (Bonafede et al., 2021) and between the cluster Abell 3562 and the radio source J 1332-3146a in the group SC 1329-313 in the Shapley supercluster ($z \approx 0.048$) at GHz frequencies (Venturi et al., 2022). Of all the bridges between pre-merging clusters, A399-401 has been most

frequently and deeply studied with X-ray observations and Sunyaev-Zeldovich (SZ) effect measurements (Fujita et al., 1996; Fabian et al., 1997; Markevitch et al., 1998; Sakelliou & Ponman, 2004; Fujita et al., 2008; Murgia et al., 2010; Planck Collaboration et al., 2013; Akamatsu et al., 2017; Bonjean et al., 2018; Hincks et al., 2022). It has already been known for a while that A401 has a radio halo (Harris et al., 1980; Roland et al., 1981; Bacchi et al., 2003), but Murgia et al. (2010) identified a radio halo in A399 as well, which made A399-401 the first detected double radio-halo system. The detection of these radio haloes suggests that the clusters themselves are also undergoing their own mergers.

Because of energy losses, relativistic electrons can only travel up to sub-Mpc distances at 140 MHz (Jaffe, 1977) in their lifetime. These age constraints mean that particles must be generated in situ to explain how diffuse radio emission can originate on Mpc scales in the A399-401 bridge. Govoni et al. (2019) proposed that radio bridges may result from first-order Fermi (Fermi-I) re-acceleration of a volume-filling population of fossil relativistic electrons by weak, $\mathcal{M} \leq 2 - 3$, shocks under favourable projection effects. Alternatively, Brunetti & Vazza (2020) suggested that the synchrotron emission from the radio bridge could be a result of second-order Fermi (Fermi-II) re-acceleration, where turbulence plays a major role by amplifying magnetic fields and reaccelerating particles. In this case, preexisting relativistic particles and magnetic fields interact with the turbulence, which re-energises them in the region between the two clusters. Nunhokee et al. (2021) recently constrained a steep spectrum ($\alpha > 1.5$) supporting a turbulent Fermi-II re-acceleration origin. Botteon et al. (2020b) found a trend between the radio and X-ray emission in the bridge A1758 by studying the spatial correlation between the two emission components. This suggests that the radio and X-ray emissions are generated in comparable volumes, which supports turbulent re-acceleration. Strong spatial correlations have been observed for radio haloes as well (Govoni et al., 2001; Feretti et al., 2001; Giacintucci et al., 2005; Rajpurohit et al., 2018, 2021; Botteon et al., 2020a; Ignesti et al., 2020; Biava et al., 2021; Duchesne et al., 2021; Bonafede et al., 2021), where Fermi-II re-acceleration in most cases been understood to be the most relevant particle acceleration process for giant radio haloes (van Weeren et al., 2019).

The goal of this paper is to study the morphology and origin of the synchrotron emission in A399-401 in more detail. We use new, deep radio data for this aim and an improved direction-dependent (DD) calibration method. With the new radio map, we study the diffuse emission from the radio haloes and radio bridge in greater depth and investigate new features related to the origin of the reaccelerated particles. Additionally, we correlate our new radio surface brightness map with an X-ray surface brightness map from A399-401 as a tool for inferring the mechanism

IDs	Date
L626678	7/12/2017
L626692	30/11/2017
L626706	16/11/2017
L632229	13/12/2017
L632511	27/12/2017
L632525	20/12/2017

Table 2.1: Observation IDs with observation dates.

behind the re-acceleration of electrons in the diffuse emission from the bridge and radio haloes.

We start by describing the data and the data reduction method in Section 2.2. The radio images are discussed in Section 2.3. In Section 2.4 we consider the relation between the radio and X-ray emission. All the results are discussed in Section 2.5. Finally, we conclude our work in Section 2.6.

We use a ΛCDM cosmology model with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. The images in this paper are made in the J2000 coordinate system.

2.2. Observations and data reduction

In this section, we describe the radio data that we used and how we calibrated them to arrive at a final image that we used for our science. Because we wish to relate the radio and X-ray emission, we also reduced X-ray observations.

2.2.1. Data

For the study of A399-A401, we used 6×8 hrs of observations from LOFAR (van Haarlem et al., 2013) from project `LC9_015` (PI: van Weeren). The observation IDs and observation dates are given in Table 2.1. Every observation has a pointing centre at a right ascension of 02h 58m 21s and a declination of +13°17' 10″ (J2000 equinox). The data cover the frequency range from 120–168 MHz and were observed only with high-band antennas. We only used the stations located in the Netherlands. `L626692` used 60 stations, while the other observations all used 62 stations.

During the testing of our calibration method (further discussed in Section 2.2.2), we decided to flag the last 1h 20m from all the observations, which leaves a total of 40 hrs of observation time. This was necessary because the calibration solutions started to diverge in this part, which was most likely caused by a low elevation that decreased the sensitivity and imposed a thicker ionosphere to look through. Moreover, we also manually flagged a few sub-bands between 126 and 128 MHz

because they were contaminated by radio frequency interference (RFI). The central frequency of our data is 144 MHz.

2.2.2. Calibration

The two main parts in the calibration of LOFAR data are direction-independent (DI) and DD calibration. The DI calibration follows the LOFAR Two-metre Sky Survey (LoTSS), where `Prefactor` version 3 with the Default Preprocessing Pipeline (`DP3`) is used (van Weeren et al., 2016c; Williams et al., 2016; de Gasperin et al., 2019b).[1] This includes RFI flagging (Offringa et al., 2012), bandpass corrections, removing data that are affected by bright off-axis sources, clock-TEC separation, polarisation alignment, ionospheric rotation measure (RM) corrections, and calibrating against a sky model from external radio surveys. The implementation of the automated DI calibration pipeline itself is discussed in Mechev et al. (2017). The DI calibration has been left untouched because the main remaining calibration issues were coming from DD solutions near bright sources around the clusters.

The DD effects are caused by ionospheric effects and imperfect beam models, which can be corrected with Jones matrices (Hamaker et al., 1996; Shimwell et al., 2019), derived from the visibilities. Over time, several DD correction algorithms have been developed, for instance, `SPAM` (Intema et al., 2009), `Sagecal` (Kazemi et al., 2011), or facet calibration (van Weeren et al., 2016c). For LoTSS, the `DDF-Pipeline` has been developed by the LOFAR Surveys Key Science Project, which is based on `KillMS` to derive the Jones matrices and to apply the solutions during the imaging of the entire field of view (FoV) with `DDFacet` (Tasse, 2014a; Tasse et al., 2018, 2021; Smirnov & Tasse, 2015; Shimwell et al., 2019).[2] With these DD calibrations, LoTSS 8-hrs observations reach $6''$ angular resolution and a typical median sensitivity of $\sigma \approx 80$ μJy beam^{-1} over the entire LoTSS-DR2 area (Shimwell et al., 2022).

Although the output from the automated `DDF-Pipeline` is sufficient to do most science with, there is still room for improvements, especially for targets with large angular extent. We therefore decided to further enhance the data reduction for this specific field. Our goal was to correct for artefacts around bright sources near A399 and A401, which determine how deep we can look into the substructure from the diffuse radio bridge and how much we can detect from the radio haloes. One of the main parts to improve is the selection of specific directions to derive and apply DD calibrations. The `DDF-Pipeline` makes corrections in a user-specified number of directions (45 is used for LoTSS standard processing). The number of directions constrains the facet layout and the final calibrated image because it is assumed that

[1] `https://github.com/lofar-astron/prefactor`
[2] `https://github.com/mhardcastle/ddf-pipeline`

the DD calibration solutions are constant throughout a facet. Because the directions are determined in an automated way, these layouts are not always optimal. This motivated us to use the recalibration method described in van Weeren et al. (2021). This calibration method has already been successfully used in numerous other works (Botteon et al., 2019, 2020a,b, 2021, 2022; Cassano et al., 2019; Hardcastle et al., 2019a; Osinga et al., 2021; Hoang et al., 2021). Because A399-401 covers a large area, however, we need to apply this method for many directions, which required including additional steps. In the following sections, we describe the recipe for a single direction ($N = 1$) from van Weeren et al. (2021), followed by an explanation of how we upgraded this to several directions ($N > 1$), and applied it to A399-401.

2.2.2.1. N=1

First, we made a square box region with the `DS9` software around a bright compact source (Joye & Mandel, 2003). In this area, self-calibration was applied and DD effects were corrected for. All sources outside of this box were subtracted from the visibilities with the DD solutions and sky model from the `DDF-Pipeline` (extraction step, see van Weeren et al., 2021). Ideally, the box had sides with a size between 0.25° and 0.4°. Box sizes need to be large enough for the flux density to be high enough so that diverging solutions are avoided in the self-calibration, whereas to improve upon the `DDF-pipeline`, the boxes need to be smaller than the facets used there because both assume constant calibration solutions across the facet. After the extraction, we phase-shifted the *uv*-data to the centre of the box and averaged the time and frequency to 16s and 0.39 MHz to reduce the data size by a factor of 8, which is sufficient for smearing purposes and does not lead to ionospheric calibration problems. With `Dysco`, we further compressed the data volume (Offringa, 2016). Then followed several rounds of self-calibration on the extracted box (self-calibration step). The starting point were the DI calibrations from the `DDF-Pipeline`. In all self-calibration rounds, we performed several so-called `tecandphase` calibrations with `DP3` (van Diepen et al., 2018b) to solve for the total electron content (TEC). We achieved this with solution intervals between 16s and 48s, followed by Stokes I gain calibrations with preapplied `tecandphase` solutions and solution intervals between 16 and 48 minutes along the time axis and solution intervals between 2 MHz and 6 MHz along the frequency axis. These solution intervals were automatically determined for each box region based on the amount of apparent compact source flux, as this differs per box. After all rounds of self-calibration, we created the final image of the facet. This was done with `WSClean` (Offringa et al., 2014) or the `DDFacet` imager (Tasse et al., 2018). See Figure 2.2 for an example of the result after eight self-calibration cycles, imaged with `WSClean`.

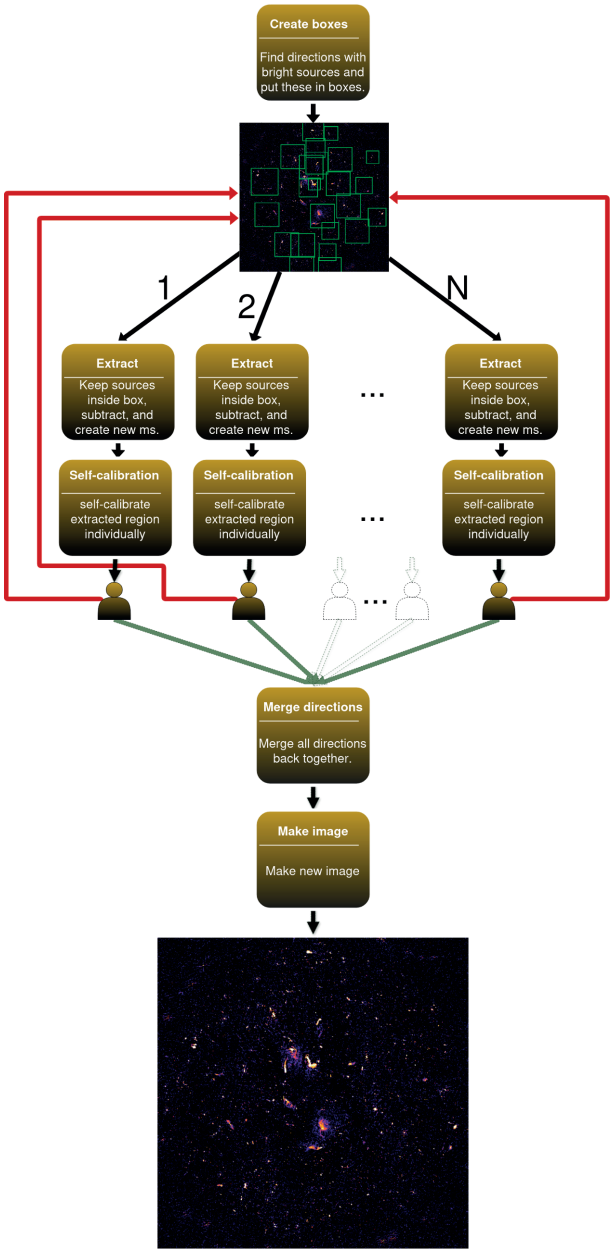


Figure 2.1: Flowchart showing every major step from the recalibration recipe described in Sections 2.2.2.1 and 2.2.2.2.

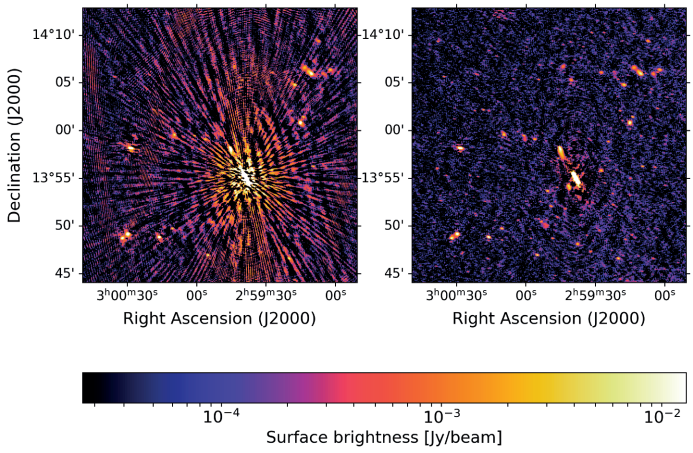


Figure 2.2: Self-calibration of an individual box. *Left panel*: Image of an extracted box before self-calibration and with the DI solutions alone. *Right panel*: Image of the same region after eight self-calibration cycles. Visible artefacts around the source disappear while correcting for phase effects. These images are made with `WSClean` version 3 (Offringa et al., 2014).

2.2.2.2. N>1

We scaled the method in Section 2.2.2.1 in order to allow the use of an arbitrary number of box regions ($N > 1$). In every box, we included at least one bright source. To limit the manual steps and to save time, we implemented an automated box-region generator. We found bright sources by scanning for all pixels where the surface brightness is higher than 70 mJy beam^{-1} in the image with a resolution of $6''$ from LoTSS that we wished to recalibrate. This pixel value was chosen because we found it to correspond to the approximate flux limit for a stable calibration for each box. We started with the brightest source and placed an initial box with a size of $0.4°\times0.4°$. Because a smaller box size speeds up the self-calibration but enough flux is necessary, the algorithm optimises the size for each box with final sizes between $0.25°$ and $0.4°$ while at the same time finding an optimal box centre not farther than $0.2°$ from the initial position. After the full box layout was optimised, we manually further fine-tuned the result to obtain the optimal composition. This was deemed to be necessary for difficult cases in which many compact bright sources are near each other, which makes it difficult for the algorithm to decide whether to include them in different or the same boxes. When the box layout was approved (see, e.g., the left panel in Figure 2.3), we followed all the same steps as in the $N = 1$ case for every individual box. Boxes may overlap because the solutions that are applied to

a part of the sky are taken from the closest box centre. In some cases, this overlap is necessary to obtain enough flux for the self-calibration. When all the individual self-calibrations were completed, we validated the quality of every set of solutions, such that no corrupt or diverging solutions were later applied in the imaging step. After the calibration, the solutions were merged into a single `HDF5` solution file per observation.[3] The box layout can be mapped to a facet layout, as we show in Figure 2.3. These facets represent the final solution area with solutions from the closest box to each pixel in the image. The solutions from these facets were applied in the final imaging step. To do this, it is only possible to use an imager that supports facets (in our case, the facet mode of `WSClean` version 3). All the main steps from box selection until imaging are summarised in Figure 2.1.

2.2.2.3. Facet calibration for A399-401

Using the method from Section 2.2.2.2 for $N > 1$, we recalibrated an area with a radius of 1.2° from the pointing centre of A399-401 with N=24 boxes. This small region size was chosen to reduce the computational cost by a factor ~4 compared to recalibrating and imaging the full field of view of our pointing.[4] This choice does not affect the result of our main target of interest, which is in the centre of the field and extends for ~0.5°. Everything outside this area was subtracted from the visibilities. In Figure 2.3 the final box and facet layout for our field are shown. Every box corresponds to a different facet.

For A399-401 we added another reduction in the computational costs by using eight self-calibration cycles instead of the standard ten cycles for every box in van Weeren et al. (2021), as we realised that the noise level in the last two rounds of self-calibration did not reduce or there were no improvements at all. Another speedup was added by applying an additional factor 4 of time and frequency averaging in the first five rounds and by returning to the unaveraged data for the last three rounds. This did not affect the final result in a noticeable way, as we obtained similar results with or without this additional averaging. As all extractions and self-calibrations can run independently of each other on different computing nodes (i.e., it is an embarrassingly parallel problem), parallelisation accelerated the total processing time with a factor ~20. In Appendix 2.A, we provide details about the computational cost of our recalibration.

After obtaining all self-calibration solutions, we had one merged `HDF5` solution file and dataset for each observation. These were then used for the final imaging with the facet mode from `WSClean` version 3, using multiscale multifrequency de-

[3] `https://github.com/jurjen93/lofar_helpers/blob/master/h5_merger.py`

[4] We used 50336 CPU core hrs for the recalibration (see Appendix 2.A), which would have been ~189000 CPU core hrs for the full field of view.

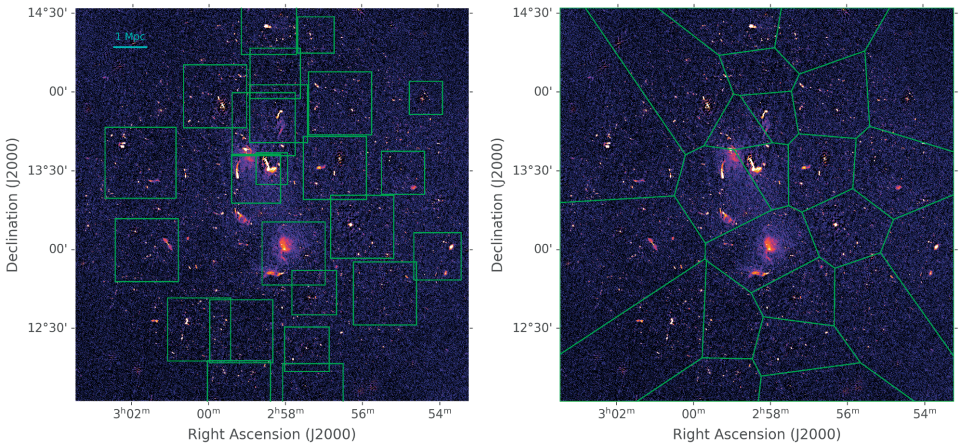


Figure 2.3: Mapping from box layout to facet layout. *Left panel*: Box layout for A399-401 within a radius of 0.6° from the pointing centre, where all individual self-calibration boxes are contoured in green. *Right panel*: Facet layout corresponding to the box layout from the left panel. The size of 1 Mpc is given in the top left corner for $z = 0.072$.

convolution and Briggs weighting in facet mode (Briggs, 1995; Offringa et al., 2014; Offringa & Smirnov, 2017). `DDFacet` also has a facet mode, but we chose `WSClean` because in our tests on this field, the deconvolution performs better for extended emission with this imager. Our final image had a resolution of $5.9'' \times 10.5''$, and we reached a sensitivity of $\sigma = 79$ μJy beam^{-1} at 144 MHz. We also further tapered the visibilities to obtain an image at $24.6'' \times 27.1''$ with $\sigma = 230$ μJy beam^{-1}, and at $72.8'' \times 75.9''$ with $\sigma = 809$ μJy beam^{-1}. These lower-resolution images have a better surface brightness sensitivity and allow us to better recover the diffuse extended emission from the radio bridge.

2.2.2.4. Advantages and disadvantages of recalibration

We can compare our highest-resolution recalibrated radio map with the radio map produced by the standard `DDF-pipeline`, which is based on the same observations. This pipeline is also used for LoTSS. By visual inspection, we see fewer artefacts around bright compact sources, and the diffuse emission is better reconstructed in our radio map than in the `DDF` image. We quantified this by studying the dynamic range around these compact sources. For most cases, this also improved (by a factor ~1.6). In Appendix 2.B we elaborate on this comparison. Overall, we can conclude that the recalibration method we used is a useful tool for calibrating a large area (larger than ~0.8°) in which calibration artefacts remain around bright sources after using the `DDF-Pipeline`. However, the high additional computational costs make

it a very expensive method at present (see Appendix 2.A). The flowchart from Figure 2.1 is not a full working pipeline either, which makes the implementation not straightforward. These advantages and disadvantages need to be considered or optimised in future usage of this method.

2.2.2.5. Removing compact sources

Because we are interested in the diffuse radio emission from the A399-401 radio bridge and radio haloes and aim in Section 2.4.2 to compare this emission with an X-ray map, we also created additional images from which the contribution from discrete compact sources was removed. As there is no perfect way to do this, we applied two different methods. Both have their advantages and disadvantages.

In the first method, we start by obtaining a compact source model by making an image from which we remove the shortest baselines corresponding to a certain physical scale. This prevents extended emission from entering the model. Then, we subtract the clean components from this high-resolution model from the starting *uv* data. With these new *uv*-subtracted data, we can make an image that is tapered to a lower resolution of $72.8'' \times 75.9''$, where contribution from compact sources is subtracted and extended emission is enhanced. We tried several baseline cuts corresponding to 200 kpc, 300 kpc, 400 kpc, and 500 kpc at the redshift of A399-401. Based on visually inspecting and comparing the final results with the original unsubtracted image, we decided to use the 300 kpc scale, as this gave the best balance between removing compact sources and having no noticeable impact on the diffuse emission. This corresponds to $216''$ and 943λ. Although the *uv*-subtract method succeeds in keeping the diffuse emission and removing most of the compact sources, there are often leftover sources mainly from extended AGNs, which can affect flux density measurements.

For the second method, we use the open map filter from Rudnick (2002) (R02 filter) to remove compact sources directly in image space. This method applies a sliding minimum filter, followed by a sliding maximum filter with the same kernel size on the image data. The R02 filter is sensitive to the noise and the kernel size. This becomes more prominent when we filter in more diffuse areas with a low signal-to-noise ratio. On the other hand, this filter is very efficient in removing all compact sources smaller than the used kernel size. However, it does not remove compact sources larger than the kernel size and can leave residual emission from partially subtracted extended AGNs. By experimenting with different settings, we decided to apply this filter on our $24.6'' \times 27.1''$ map with a kernel size of $42''$ (corresponding to 60 kpc at the redshift of A399-401) and further smooth this to $72.8'' \times 75.9''$ to have the same resolution as the other radio map.

2.2.2.6. X-ray data

We retrieved archival *XMM-Newton* observations of A399-401 from the Science Archive[5]. In particular, we made use of three pointings: 0112260101 (covering A399), 0112260301 (covering A401), and 0112260201 (covering the region between the two clusters). The European Photon Imaging Camera (EPIC) observations were processed with the *XMM-Newton* Scientific Analysis System (SAS v16.1.0) and the Extended Source Analysis Software (ESAS). After filtering bad time intervals due to soft proton flares, we produced an EPIC mosaic image in the $0.5 - 2.0$ keV band combining the three ObsIDs. This was used to compare the X-ray and radio (from LOFAR) surface brightnesses of the observed emission. For a detailed analysis of the *XMM-Newton* observations, we refer to Sakelliou & Ponman (2004).

2.3. Results

Figures 2.4 and 2.5 present the LOFAR observation of A399-401 at three different resolutions. In Figure 2.6 we show the *uv*-subtract and R02 filtered images, with grids and slices that are explained below. The radio map from our highest-resolution map is more than four times deeper than the $10''$ radio map from Govoni et al. (2019) ($\sigma = 79$ μJy beam^{-1} versus $\sigma = 300$ μJy beam^{-1}). We also show zoomed images in Figure 2.7, similar to Figures S3 and S4 in Govoni et al. (2019), and use labels in Figure 2.4 and 2.7 to mark a number of particularly interesting regions that are referred to throughout this paper.

2.3.1. Diffuse emission

In Figure 2.4 we clearly detect two radio haloes belonging to A401 in the north and A399 in the south. At the lower resolutions in Figure 2.5, we also directly observe the diffuse emission from the radio bridge. Below we discuss the morphology of the radio bridge and haloes and measure the flux densities and radio powers.

2.3.1.1. Morphology

In the two panels from Figure 2.5, we observe a prominent brightness depression west of the bridge close to A399. We can detect this depression also from slices 1 to 5 in the radio surface brightness profile in Figure 2.8, where we slice through the radio bridge images in Figure 2.6 in the northeast direction for both the R02 filter and *uv*-subtract radio maps. We identify various compact radio sources in the bridge area, but with the exception of the sources in region D from Figure 2.7,

[5] `http://nxsa.esac.esa.int/nxsa-web`

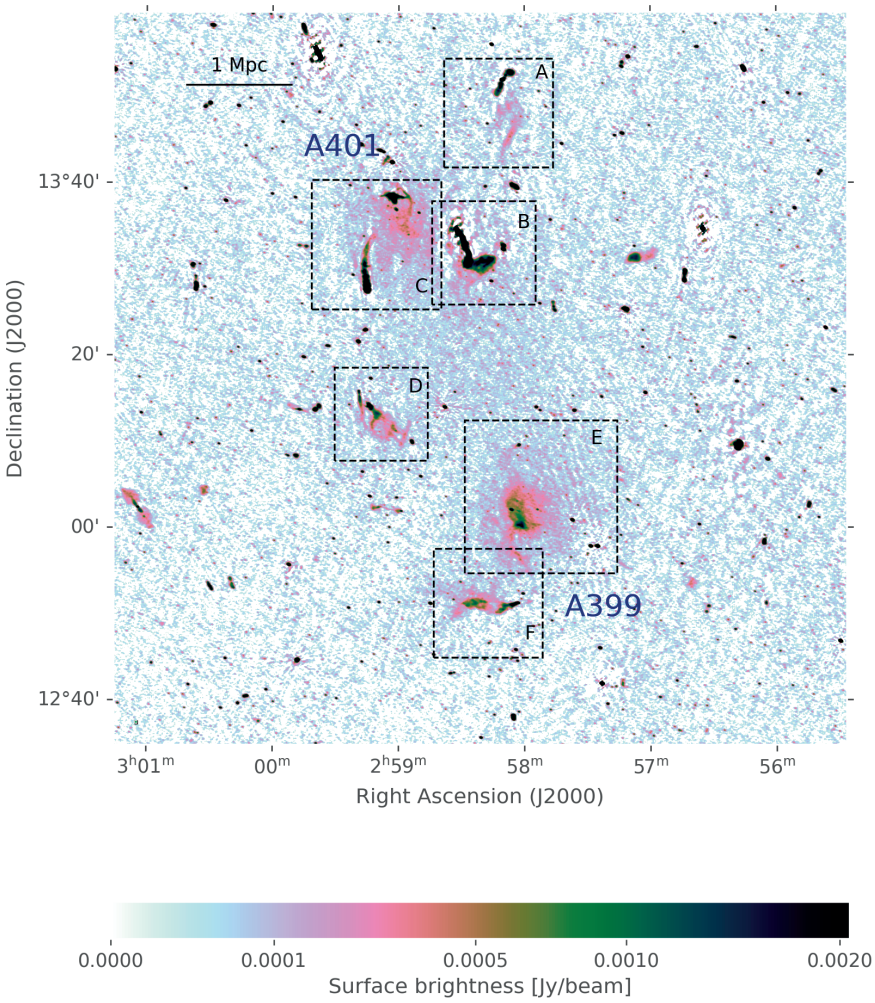


Figure 2.4: Radio map of A399-401 at a resolution of $5.9'' \times 10.5''$ with $\sigma = 79\ \mu$Jy beam^{-1}. The black highlighted regions correspond to the zoomed images in Figure 2.7. The beam size is given in the lower left corner, and the scale of 1 Mpc at $z = 0.072$ is given in the upper left corner. The square-root colour scale of the images extends from 0 to 25σ.

we do not detect any indication of a morphological relation between them and the radio bridge.

In Figure 2.7 we highlight the radio halo from A401 in region C. The largest linear size (LLS) for this radio halo, measured within the 3σ contour, is 1.6 Mpc. In the region covered by the radio halo, we detect a morphologically complex bright source (C1) that has no direct optical counterpart. This structure has a straight

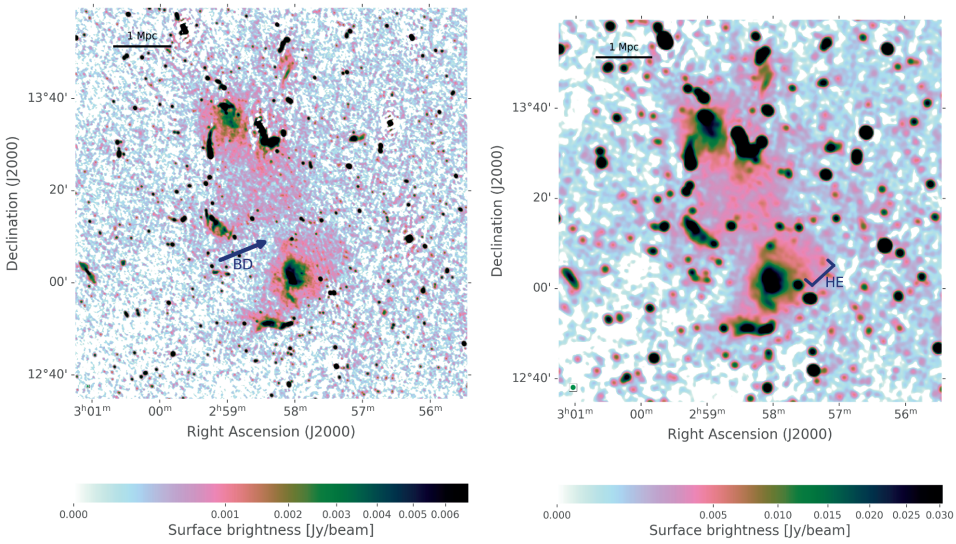


Figure 2.5: Radio maps of the A399-401 system at lower resolutions. *Left panel*: Resolution of $24.6'' \times 27.1''$ with $\sigma = 230$ μJy beam^{-1}. *Right panel*: Resolution $72.8'' \times 75.9''$ with $\sigma = 809$ μJy beam^{-1}. The dark blue brightness depression (BD) and the halo extension (HE) are described in the text. The beam size in all images is given in the lower left corner, and the scale of 1 Mpc at $z = 0.072$ is given in the upper left corner. The square-root colour scale of the images extends from 0 to 25σ.

feature with a full LLS of 300 kpc on its western and eastern side, but on its western side it, seems to connect to a bent jet-like structure (C2) that might be associated with the elliptical galaxy 2MASX J02585500+1338243 at $z = 0.079$ (Hill & Oegerle, 1993). In region E from Figure 2.7, we see the radio halo from A399. From a 3σ isophote, we find an LLS of 1.3 Mpc. In the southern area, attached to the halo, we detect a diffuse bent structure with two components with an LLS of $\sim$230 kpc and $\sim$150 kpc (E1). In the middle of these components, we can notice a small dip in the emission where we find the elliptical galaxy 2MASX J02580602+1257559 at $z = 0.075$ (Hill & Oegerle, 1993). This dip, at the location of the optically detected galaxy, suggests that this is a switched-off radio galaxy. The radio halo also has a brightness enhancement without an optical counterpart (E2) with an LLS of $\sim$100 kpc. The lower resolutions in Figure 2.5 show that the radio halo extends to the northwest, which we treated as a separate component for further analysis and in the discussion in Sections 2.4.2 and 2.5.

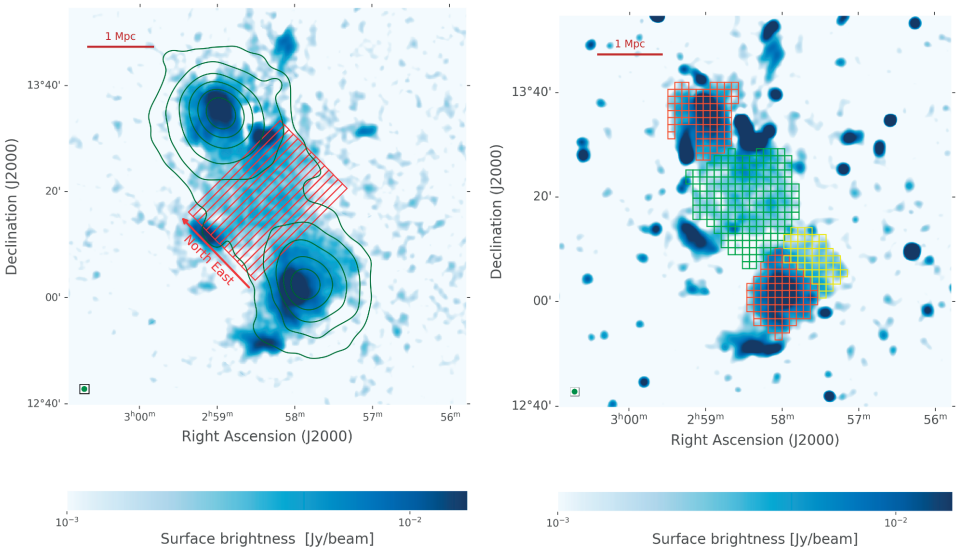


Figure 2.6: Radio maps of A399-401 where most compact sources are removed. *Left panel*: Radio image in blue: *uv*-subtract image described in the text. The square-root colour scale extends from 1σ to 25σ. Green contour lines are from X-ray data from XMM Newton, which are smoothed with a Gaussian kernel. The red slices ($1100'' \times 75''$) point northeast. *Right panel*: Radio image in blue: R02 filtered image described in the text. The square-root colour scale of the images extends from 1σ to 25σ. It has a grid overlay on the two clusters and on the radio bridge on top of the radio image with the R02 filter. The orange grids cover the haloes, the green grid covers the radio bridge, and the yellow grid covers the northwest radio halo extension from A399. The cell size in the grid is $80'' \times 80''$. The scale for 1 Mpc at $z = 0.072$ is given in both images in the upper left corner.

2.3.1.2. Flux densities and radio powers

To calculate the integrated flux densities for the haloes, we fit the radio haloes with the halo-flux density calculator (`Halo-FDCA`) (Boxelaar et al., 2021). This software applies a Markov chain Monte Carlo (MCMC) method to fit an exponential surface brightness profile,

$$I(r) = I_0 e^{G(\mathbf{r})},$$

where I_0 is the central radio surface brightness, and $G(\mathbf{r})$ is a quantity related to the morphology of the halo as a function of the two-dimensional distance from the centre ($\mathbf{r}$). The radio power is calculated by

$$P_{144\ \mathrm{MHz}} = \frac{4\pi D_L^2}{(1+z)^{1+\alpha}} S_{144\ \mathrm{MHz}},$$

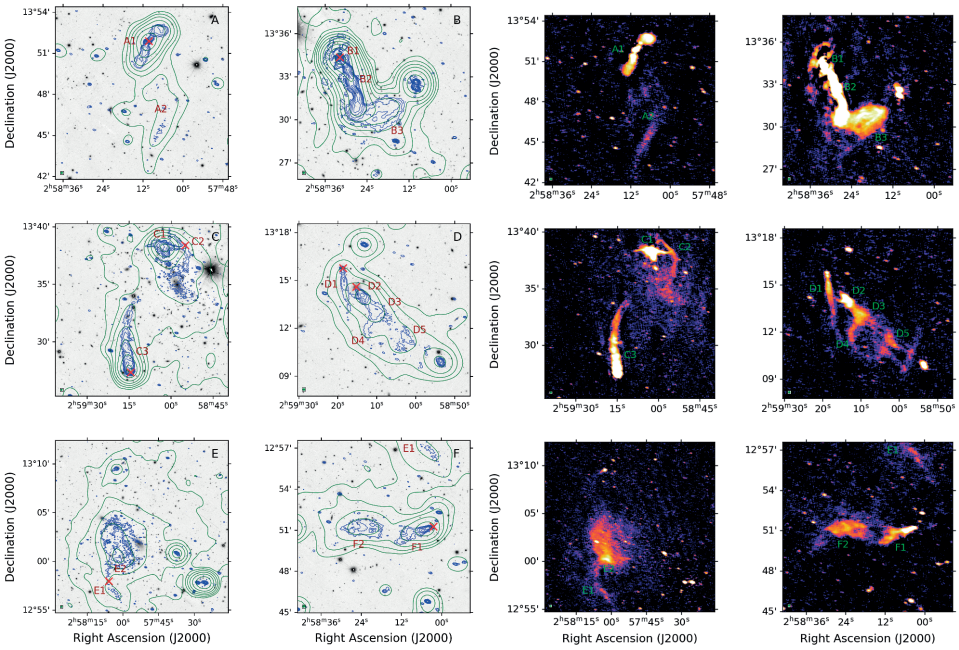


Figure 2.7: Zoomed images of the regions in Figure 2.4. These are the same sources as in Fig. S4 from Govoni et al. (2019) and ordered in the same way to facilitate comparison. *Left two panels*: Blue contours at a resolution of $5.9'' \times 10.5''$ and with $\sigma = 79$ μJy beam^{-1} , and green contours at a resolution of $72.8'' \times 75.9''$ and $\sigma = 809$ μJy beam^{-1} , both at 144 MHz. Red crosses are elliptical host galaxies from the radio sources, and we label the radio components discussed in the text in brown. The background gray-scale images are optical sources from Pan-STARRS DR1 (Chambers et al., 2016). *Right two panels*: Same regions as in the left panel at $5.9'' \times 10.5''$ with a surface brightness colour plot. The square-root colour scale of the images extends 1σ to 25σ to better highlight the plasma, and the labels are the same as in the left panel in green.

where D_L is the luminosity distance and $S_{144\ \mathrm{MHz}}$ is the integrated total flux for a spectral index α. For A399 and A401, we find that a circular model with $G(\mathbf{r}) = -\frac{\mathbf{r}}{r_e}$ works well, where r_e is the e-folding distance to the halo centre. With this fit, we can estimate the flux density and radio power with corresponding uncertainties. Following Murgia et al. (2009), we decided to integrate up to $3r_e$. For the radio halo from A399, we obtain $r_e = 208 \pm 6$ kpc and find $S_{144\ \mathrm{MHz}} = 0.98 \pm 0.10$ Jy and $P_{144\ \mathrm{MHz}} = 0.99 \pm 0.11 \times 10^{25}$ W Hz^{-1} by using the best estimated spectral index ($\alpha = 1.75 \pm 0.14$; Nunhokee et al., 2021). For the radio halo in A401, we obtain $r_e = 208 \pm 7$ kpc and find $S_{144\ \mathrm{MHz}} = 0.77 \pm 0.08$ Jy and $P_{144\ \mathrm{MHz}} = 0.99 \pm 0.03 \times 10^{25}$ W Hz^{-1} with the best estimated spectral index ($\alpha = 1.63 \pm 0.07$; Govoni et al., 2019). The output from `Halo-FDCA` is shown in Appendix 2.C. `Halo-FDCA` is created for

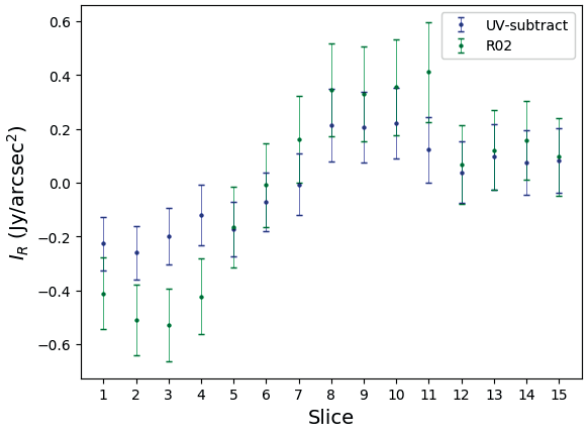


Figure 2.8: Mean radio surface brightness from the slices from the left panel in Figure 2.6, manually masked for bright AGNs, for the R02 filter and *uv*-subtracted radio data. The slice numbers correspond to the order of the slices in the direction of the arrow. The R02 filter removes more from the diffuse emission than the *uv*-subtract. Error bars include the statistic and systematic uncertainties.

radio haloes. For the radio bridge, we therefore created a manual region that we associated with the bridge (covering a similar area as the green boxes in the right panel in Figure 2.6) and integrated over this area in the *uv*-subtract image. We obtain $S_{144\ \mathrm{Mhz}} = 0.55 \pm 0.06$ Jy, and by adopting the current lowest estimated spectral index for the bridge ($\alpha = 1.5$; Nunhokee et al., 2021), we find $P_{144\ \mathrm{MHz}} = 0.75 \pm 0.08 \times 10^{25}$ W Hz^{-1} as the upper limit. All values are listed together in Table 2.2. The uncertainties include systematic, subtraction, and statistical errors. The systematic error takes into account the uncertainty of the flux scale calibration, the subtraction error takes into account the uncertainty from remaining residual emission from discrete sources after subtraction, and the statistical error takes into account the sensitivity of the image (see also Section 5 from Botteon et al., 2022).

Murgia et al. (2010) also used a circular exponential fit to calculate the flux density for the radio halo from A399. They found $r_e = 186 \pm 16$ kpc at 1.4 GHz with VLA data, which is lower than but still consistent within the error bars with our value. Govoni et al. (2019) used a different radio map, with a lower resolution and sensitivity than our map ($10''$ and $\sigma = 300$ μJy beam^{-1} respectively), to measure the flux densities from the radio haloes over a 5σ isophote. Despite these different methods, our values for the radio haloes are consistent within the error bars with those from Govoni et al. (2019) ($S_{140\ \mathrm{MHz}} = 826 \pm 126.5$ mJy for A401; $S_{140\ \mathrm{MHz}} = 807 \pm 124.7$ mJy for A399). Govoni et al. (2019) measured $S_{140\ \mathrm{MHz}} = 822 \pm 147$ mJy over 3.9 Mpc^2 for the radio bridge. This area is significantly larger than the

	r_e [kpc]	$S_{144\ \mathrm{Mhz}}$ [Jy]	$P_{144\ \mathrm{MHz}}$ [W/Hz]	α
A399	208 ± 6	0.98 ± 0.10	$1.26 \pm 0.13 \times 10^{25}$	1.75 ± 0.14
A401	241 ± 7	0.77 ± 0.08	$0.99 \pm 0.11 \times 10^{25}$	1.63 ± 0.07
Bridge	N.A.	0.55 ± 0.06	$0.75 \pm 0.08 \times 10^{25}$	> 1.5

Table 2.2: Measured physical properties at 144 Mhz with spectral indices from Govoni et al. (2019) and Nunhokee et al. (2021).

2.7 Mpc^2 that we find for the bridge. Our bridge area is more conservatively chosen than Govoni et al. (2019) because they extrapolated the average surface brightness from cluster core to cluster core for the masked regions (radio haloes and sources), while we excluded the radio haloes from the bridge area entirely. For the average surface brightness, we both find ~0.38 μJy arcsec^{-2}, which means that our results (independent of the area we constrain for the bridge) are consistent with each other.

2.3.2. AGNs

In our radio images, we detected several interesting radio components that are associated with galaxies in or near A399-401. These bright radio sources make up a large part of all the radio emission in A399-401. We therefore discuss the radio components in our radio maps in detail. Most of the radio emission is likely associated with AGNs, as already discussed by Govoni et al. (2019). Remnant plasma from AGN is also a possible ingredient to explain the origin of the emission from the radio bridge and radio haloes (this is further discussed in Section 2.5).

In region A from Figure 2.7, we detect the radio galaxy (A1) corresponding to 2MASX J02581042+1351526, which was associated by Harris et al. (1980) as a probable member of A401 based on its distance to the cluster centre (~1.5 Mpc) and brightness. This object has two lobes with a full LLS of ~300 kpc, and it is connected to diffuse emission pointed toward the A399-401 system (A2). Region B in Figure 2.7 has a bright active radio core (B1) corresponding to the elliptical galaxy 87GB 025547.6+132220 at $z = 0.084$ (Huchra et al., 2012). Whereas there was a gap between the long tail (B2) and the core (B1) in the images from Govoni et al. (2019), we detect it as one connected structure. The radio emission from B2 has an LLS of ~320 kpc. Connected at the end of this emission, we detect remnant plasma stretched to the west over a similar LLS of ~270 kpc (B3). We observe signs of plasma extending southward into the radio bridge filament at the edges of B2 and B3. To the southeast of the radio halo from A401, we detect a radio galaxy (C3) in region C from Figure 2.7 that we associate with 2MASX J02591487+1327117 at $z = 0.078$ (Hill & Oegerle, 1993). Component C3 has a ~550 kpc long prominent tail extending toward the radio halo.

In region D in Figure 2.7, we detect two sources next to each other: 2MASX J02591878+1315467, located northeast at $z = 0.073$, and southwest from this source, we find 2MASX J02591535+1314347 at $z = 0.078$ (Hill & Oegerle, 1993). The first is an elliptical galaxy associated with a tail with an LLS of ~220 kpc (D1), and the second is an elliptical galaxy associated with the bright radio core to the southwest (D2) with an LLS of ~90 kpc. From the core, a fainter component extends southwest (D3) with an LLS of ~80 kpc and a long bent structure pointed to the south (D4). Following the same direction to the southwest, the diffuse emission again becomes brighter (D5), and in the lower-resolution images in Figure 2.5, this area seems to be connected to the bridge. No optical galaxy is associated with this emission. In region F in Figure 2.7, immediately below the radio halo from A399, we detect two regions with brighter plasma that lie next to each other. While the ~250 kpc long plasma on the west side (F1) can be associated with the optical elliptical galaxy 2MASX J02580300+1251138, which is located at $z = 0.075$ (Hill & Oegerle, 1993), we do not find an obvious optical counterpart for F2, which has about the same length. Component F1 corresponds to a tailed source, and we suspect that F2 might be the extension of F1, given its morphology (see the discussion in Section 2.5.2).

2.4. Thermal and non-thermal scaling relations

It has been shown observationally (e.g., Cassano et al., 2013; van Weeren et al., 2021) that the thermal emission from the ICM and non-thermal radio emission are related by the following scaling relation:

$$P_\nu \propto M^\gamma,$$

which is the relation of the radio power integrated over the entire cluster P_ν at a wavelength ν and the cluster mass M derived from X-ray or SZ measurements with slope γ. It was suggested that non-thermal emission is powered by the dissipation of gravitational energy (e.g. Cassano et al., 2006).

Instead of using a statistical population of clusters to determine the thermal and non-thermal relation in the ICM by means of integrated quantities (P and M), we can also use spatially resolved quantities on single objects by inspecting the following scaling relation:

$$I_R \propto I_X^a,$$

where I_R is the radio surface brightness, and I_X is the X-ray surface brightness with slope a. This relation has been derived for many radio haloes with a point-to-point analysis (Govoni et al., 2001; Feretti et al., 2001; Giacintucci et al., 2005;

Rajpurohit et al., 2018; Botteon et al., 2020a; Rajpurohit et al., 2021; Ignesti et al., 2020; Biava et al., 2021; Duchesne et al., 2021; Bonafede et al., 2021). A point-to-point analysis has also been performed for the radio bridge of A1758 (Botteon et al., 2020b). A strong correlation reflects the physics of the interplay between the thermal and non-thermal components (e.g., Brunetti, 2004; Brunetti & Jones, 2014), where the particle density and magnetic field strength (traced by the radio surface brightness) decline faster than the thermal gas density (traced by the X-ray surface brightness) if $a > 1$, or vice versa, if $a < 1$. We study these scaling relations for A399-401 below.

2.4.1. Mass-power relation

To determine where the radio haloes from A399 and A401 are located in the cluster mass radio power diagram, we used the most recent relation at a close frequency from van Weeren et al. (2021). Following Cassano et al. (2013), they derived the following scaling relation:

$$P_{150\ \mathrm{MHz}} \approx M_{500}^{6.13\pm1.11}$$

for a sample of clusters in a rest-frame of $P_{150\ \mathrm{MHz}}$, and where M_{500} is the cluster mass within a radius whithin which the average density is equal to 500 times the critical density of the Universe, taken from the PSZ2 *Planck* catalogue (Planck Collaboration et al., 2016). We included A399 and A401 in the figure from van Weeren et al. (2021) (see Figure 2.9). The radio powers from Table 2.2 were scaled to 150 MHz using the spectral indices from Govoni et al. (2019). To be consistent with van Weeren et al. (2021), we used the masses from Planck Collaboration et al. (2016) ($5.24^{+0.29}_{-0.23} \times 10^{14}$ $\mathrm{M}_\odot$ for A399; $6.75^{+0.22}_{-0.17} \times 10^{14}$ $\mathrm{M}_\odot$ for A401). The radio haloes are located close to the best-fit relation, implying that their radio powers are similar to those of other clusters with similar masses.

Table 2.2 indicates that the radio power of the bridge is comparable to that of both of the radio haloes, while we know from Hincks et al. (2022) that the mass of the bridge is only roughly 8% of the total mass of A399-401. This means that the bridge does not fit into the same mass-power relation as the radio haloes, which is no surprise if bridges and radio haloes have different origins through different physical mechanisms. More observations of radio bridges are needed to infer whether a different mass-power relation exists.

2.4.2. Point-to-point analysis

The spatial correlation between the radio and X-ray emission reflects the strength of the connection between the thermal and non-thermal components in the ICM. Therefore, a point-to-point analysis can be used as a tool for comparing the radio

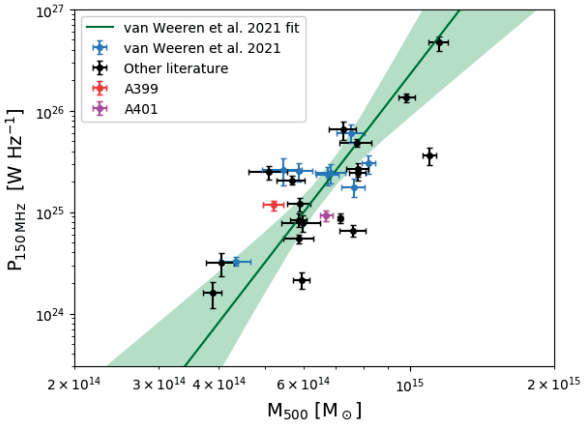


Figure 2.9: Mass and radio power relation. The fitted regression line with a 3σ confidence interval in the shaded area and the data points in cyan come from the sample in van Weeren et al. (2021). The literature data points in black are taken, similar to van Weeren et al. (2021), from previous LOFAR and *GMRT* studies, and a correlation line is fit using the bivariate correlated errors and intrinsic scatter (`BCES`) orthogonal regression algorithm (Akritas & Bershady, 1996). We left out the radio halo candidates. The radio haloes A399 and A401 are added. Error bars include statistical and systematic uncertainties.

and X-ray emission and for obtaining information about the mechanisms of acceleration and transport of relativistic particles, and the amplification of magnetic fields in radio haloes and bridges. This helps us to understand the origin of the radio bridge and haloes better.

Similarly to the radio power and mass relation, we derived the radio and X-ray relation in log-log space,

$$\log(I_R) = a \log(I_X) + b,$$

where we refer to a as the slope. To quantify the correlation, we derived the Spearman correlation coefficient (r_s). We used the *uv*-subtract and the R02 filtered map, which each have their advantages and disadvantages, as we explained in Section 2.2.2.5. We generated a grid that covers the radio bridge and radio haloes separately (see the right panel in Figure 2.6). There is no clear boundary between the radio haloes and the radio bridge, therefore we used a 5σ radio contour around the radio haloes to define the border between the radio bridge and halo areas. Furthermore, the northwestern radio halo extension from A399 and the core from A399 each have their own grids (yellow and orange, respectively) because we show below that this will help to explain the origin of the radio halo from A399. We chose a grid cell size of $80'' \times 80''$. This size is slightly larger than the beam size for the radio and X-ray map and therefore prevents a correlation between contiguous cells in the grid.

	r_s	a
A399	0.60 ± 0.20	0.33 ± 0.11
A399 core	0.47 ± 0.15	0.32 ± 0.07
A399 HE	0.71 ± 0.15	0.35 ± 0.12
A401	0.91 ± 0.04	0.63 ± 0.06
Bridge	0.41 ± 0.14	0.27 ± 0.07

Table 2.3: Correlation coefficients and slopes for $I_R - I_X^a$.

Larger cell sizes improve the signal-to-noise ratio but reduce the number of data points for a linear regression, especially for radio haloes where there is less area to cover. We calculated the average surface brightness and errors for every cell, including the statistical and systematic uncertainties for the radio and X-ray emission. From the X-ray surface brightness, we subtracted the sky background contribution, which we derived to be $1.27 \cdot 10^{-7}$ count/s/arcsec^2 in the 0.5-2.0 keV band. This contributes up to ~40% of the emission at the more diffuse edge of the bridge region from A399-401. We adopted a radio surface brightness threshold of 2σ and removed cells covering areas that are not related to the radio haloes or bridge, such as objects associated with AGNs, which are only partially removed in the source subtraction. The 2σ threshold is needed to prevent any effect from unreliable flux densities and noise at a low signal-to-noise ratio. Adopting a higher threshold can effectively flatten the slope values. To reduce this effect, we followed Botteon et al. (2020a) and used `LIRA`, which is a Bayesian regression method that allows fitting data points that have a (2σ) threshold in the y-variable (Sereno, 2016).[6] With this regression method, we obtain a mean and error that reflect the errors on the radio and X-ray measurements as well.

To reduce the choice sensitivity of the grid, we generated grids in a Monte Carlo (MC) fashion. This is similar to what is described in Ignesti (2022). In our approach, we made multiple grid realisations with small random offsets of $40''$ at most (half of the grid resolution) around the starting and ending points of the grid. In this way, we generated 100 different grid layouts for the R02 filtered and *uv*-subtract radio maps. An example of a grid layout is shown in Figure 2.6 (right panel). Finally, we calculated the final values and errors for the slopes and the Spearman correlation coefficients with the bootstrapping method similar to Ignesti (2022), such that we propagated the errors from individual fits to the final values. All results are presented in Table 2.3. The correlation plots are shown for one particular grid choice in Figure 2.10.

All correlations are positive and all slopes are sublinear ($a < 1$) in Table 2.3. The

[6] `https://cran.r-project.org/web/packages/lira/`

radio halo extension from A399 has a much steeper slope and a stronger correlation between the radio and X-ray emission than the core from A399. In Section 2.5.2 we provide more detail about this.

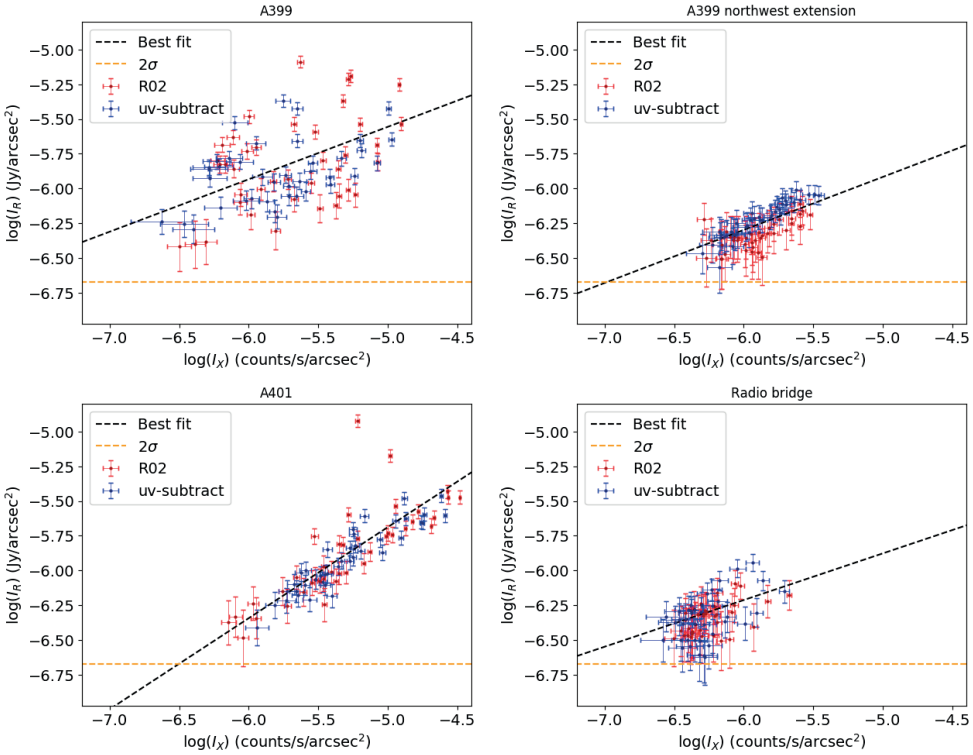


Figure 2.10: Radio and X-ray surface brightness correlation plots for every cell from the grid used for the point-to-point analysis in the right panel of Figure 2.6 for the R02 filter and *uv*-subtract maps. This is just one grid from the MC grid generation with 40 randomly drawn points to improve the readability of the plot. *Top left panel*: Radio halo of A399. *Top right panel*: Radio halo extension of A399. *Lower left panel*: Radio halo of A401. *Lower right panel*: Radio bridge. Error bars include the statistic and systematic uncertainties. The best linear fit with `LIRA` is given with a dashed black line. Only radio brightness values higher than 2σ are included.

2.5. Discussion

Because radio bridges are a very recent discovery, the origin of the radio emission from these structures remains an open question. The Mpc scale size of the A399-401 radio bridge and the maximum sub-Mpc distances that relativistic particles

can travel due to age constraints require an in situ mechanism as the driver of the emission of synchrotron radiation throughout the bridge region (Brunetti & Jones, 2014). An important ingredient for these models is the presence of fossil cosmic-ray electrons (e.g., Brunetti et al., 2001; Brunetti & Lazarian, 2011; Petrosian, 2001; Pinzke et al., 2017). These seed relativistic electrons in the energy range 100-500 MeV have a long lifetime in the ICM and can be injected by past shock activity, AGNs, galactic winds, or via the decay of charged pions from proton-proton collisions (e.g., Brunetti & Jones, 2014). During merger events, fossil cosmic rays can be reaccelerated via various Fermi-I or II mechanisms and/or be re-energised by adiabatic compression (see Brunetti & Jones, 2014, for a review). Evidence for the revival of AGN fossil radio plasma, for example, so-called gently re-energised tails (GReEt) and radio phoenices, has been observed in a number of clusters (e.g., de Gasperin, 2017; van Weeren et al., 2017; Mandal et al., 2020).

As the radio bridge is connecting the radio haloes from A399 and A401, it is important to understand the origin of these radio haloes as well. The fact that we observe radio haloes in the centre of these clusters also means that they are also undergoing their own mergers (Buote, 2001; Cassano et al., 2010).

2.5.1. Origin of the radio bridge

Although Govoni et al. (2019) initially suggested a model in which weak shocks reaccelerate particles via a Fermi-I type mechanism, recent studies favour a model with Fermi-II re-acceleration through turbulence to explain the origin of the A399-401 radio bridge (Brunetti & Vazza, 2020). With the point-to-point analysis, we find a trend between the radio and X-ray surface brightnesses, similar to Botteon et al. (2020b). We also find a source for fossil plasma, as we detect evidence for AGN injection of relativistic particles into the radio bridge region. In particular, we observe radio brightness enhancements around AGN jets that are likely places where plasma is being injected in the radio bridge. This is best visible at the southern tip of components B2 and B3 (see Figure 2.7). Moreover, the plasma from A2, likely coming from an AGN (see Section 2.3.2), is pointed toward the bridge. These examples make a compelling case for the scenario in which in the past, these and other AGNs have dumped fossil plasma, which now functions as the source of primary seed electrons ready to be reaccelerated through turbulence. At the same time, the fossil plasma can scatter the radio surface brightness distribution, which in return can reduce the spatial correlation between the radio and X-ray emission. With the fossil plasma we also have an important ingredient for an in situ turbulent re-acceleration model. In addition, we do not observe filamentary structures or shock surfaces in the bridge region, which is another indication that the emission is volume filling, and turbulence instead of shock models is currently

the best explanation. Fermi-II re-acceleration therefore remains the best candidate to explain the origin of the radio bridge. We suggest that in follow-up work, a high-resolution study of the spectral index across the bridge would be interesting to better understand the fate of the relativistic electrons that are injected into thermal gas and the role of the re-acceleration processes.

An observed X-ray temperature break by Akamatsu et al. (2017) in the region that we labelled D in Figure 2.7 is suggested to be a sign of an equatorial shock from the A399-401 merger axis (Ha et al., 2018) or a milder adiabatic compression between the clusters (Gu et al., 2019). Figure 2.7 shows that the emission from D5 is stretched in the same direction as the jet from D3, which is associated with an AGN in 2MASX J02591535+1314347. Therefore, we propose that the emission from D5 is a remnant tail from the AGN, which is re-energised by the equatorial shock or adiabatic compression. This is an alternative to the candidate radio relic classification by Govoni et al. (2019) or the switched-off radio galaxy explanation from Nunhokee et al. (2021). We further observe in Figure 2.5 that the emission from D5 is directly connected with the bridge. This shows that the morphology of the radio bridge could be directly affected by this shock or compression, which can also change the spatial correlation between the radio and X-ray emission.

To determine how the values from the point-to-point analysis for A399-401 compare with other radio bridges, we considered the point-to-point analysis that was performed for the radio bridge in A1758. We used the data behind Figure 4 from Botteon et al. (2020b) and derived $a = 0.25 \pm 0.08$ and $r_s = 0.52 \pm 0.22$. These values are similar to what we find for the radio bridge in A399-401 (see Table 2.3). Although we used a finer grid resolution (~115 kpc versus ~185 kpc) than Botteon et al. (2020b) for A399-401, this might indicate that the two radio bridges are produced by similar processes. When we compare the slopes from these two radio bridges with mini and giant radio haloes in the literature or with those from A399 and A401, we conclude that slopes from radio bridges are overall flatter (Govoni et al., 2001; Feretti et al., 2001; Giacintucci et al., 2005; Rajpurohit et al., 2018, 2021; Botteon et al., 2018b, 2020b; Ignesti et al., 2020; Biava et al., 2021; Bonafede et al., 2021, 2022). This might be an indication that the physical connections between the ICM and radio bridges and the ICM and radio haloes are different from each other. More radio bridges in pre-merging clusters need to be studied to conclude whether these correlation and slope values are typical for radio bridges between pre-merging clusters, and how this relates to the underlying physical processes.

2.5.2. Radio haloes

With our point-to-point analysis, we find a sublinear slope and a remarkable strong correlation between the radio and X-ray emission from A401. Similar to the radio

bridge, we also detect AGNs that inject the radio halo environment with plasma. First of all, Figure 2.7 shows that the AGN tail from C3 is directed toward the bright enhanced part in the radio halo. Second, the AGN labelled C2 affects the radio halo environment from the edges from its southern lobe. The combination of a strong radio and X-ray correlation with the observed AGN activity suggests a scenario in which turbulent Fermi-II re-acceleration of fossil plasma injected by AGNs in the cluster causes most of the radio emission from this radio halo (Brunetti et al., 2008; Brunetti & Jones, 2014; ZuHone et al., 2015). This is further supported by the steep spectrum ($\alpha = 1.75 \pm 0.14$) measured by Nunhokee et al. (2021), which can be best explained by a turbulent re-acceleration mechanism in moderately disturbed systems (Brunetti et al., 2008).

We find a weaker correlation between the radio and X-ray emission for the radio halo from A399 (with and without the northwest extension) than for the radio halo from A401 with the point-to-point analysis. The relation between the radio and X-ray emission components is likely affected by the cluster merger in A399 between a higher-mass system and a lower-mass system going from east to west, as proposed by Sakelliou & Ponman (2004) and simulated by Takizawa (1999). Evidence for this merger comes from an X-ray edge at the southeast side of the cluster core of A399 (Sakelliou & Ponman, 2004). The edge coincides with the region of enhanced radio brightness at the centre of the radio halo, where the enhanced emission labelled E2 in Figure 2.7 might be the result of a weak shock from the merger event (Murgia et al., 2010). The unrelaxed dynamical state of the cluster is reflected in the offset of the radio halo peak with respect to the X-ray peak. This is also in line with the cold front claimed by Botteon et al. (2018a). Other clusters that are in a complex merging stage show similar weaker correlations between the radio and X-ray emission from a point-to-point analysis (Shimwell et al., 2014; Duchesne et al., 2021). In contrast, we find a steeper slope and a strong radio and X-ray correlation in the northwest extension (labelled HE in the lower right panel in Figure 2.5) from the radio halo. Together with steep-spectrum from Nunhokee et al. (2021), this makes a case for turbulent Fermi-II re-acceleration of cosmic rays in A399, which is directed from the merger axis toward the northwest from the radio halo (Brunetti et al., 2008), and where a recent merger between a higher- and lower-mass system scatters the radio and X-ray relation around the core. The two radio tails from E1 (Figure 2.7) seem to be coming from a currently switched-off AGN, as the optical source is situated in the brightness dip between these jets. Its northern jet is directly connected with the radio halo and might be a source of seed particles that are needed for the turbulent re-acceleration in the radio halo. Instead, it is also possible that fossil plasma in the jets is re-energised by the currently ongoing merger in A399. Farther south of this cluster, we detect emission labelled F2 (Figure 2.7), which might be a re-

energised fossil plasma (originating in but disconnected from the AGN tail from 2MASX J02580300+1251138) by the merger in A399. This is again an alternative explanation to the candidate radio relic classification from Govoni et al. (2019).

2.6. Conclusions

A399-401 is one of a few giant intracluster radio bridges that have been observed so far (e.g., Govoni et al., 2019; Botteon et al., 2020b). We created new radio maps from A399-401 by using the improved recalibration method from van Weeren et al. (2021) and combining this with the wide-field facet imaging mode in `WSClean` version 3 on ∼40h LOFAR data from six different observations. Despite the high computational costs compared to the standard `DDF-Pipeline`, we find that this method works well for calibrating large diffuse structures where calibration artefacts around compact sources can be an issue in reconstructing the diffuse emission with the `DDF-Pipeline`. In the case of A399-401, we measure improvements of a factor ∼1.6 in dynamic range for bright compact sources in our recalibrated radio map compared with the radio map produced with the `DDF-Pipeline` . In comparison with the previously studied radio map of A399-401 (Govoni et al., 2019), we improved the resolution from $10'' \times 10''$ to $5.9'' \times 10.5''$ and the sensitivity from 300 μJy beam^{-1} to 79 μJy beam^{-1}.

By analyzing the resulting images and using a point-to-point analysis to compare the radio and X-ray surface brightness changes across a region, we find the following:

- ☐ We clearly detect the radio haloes and the radio bridge in A399-401. We report for the first time a prominent brightness depression close to the radio halo from A399, starting west of the bridge. This shows that the radio bridge is not one straight elongated structure stretching from A399 to A401.

- ☐ We find a trend between the radio and X-ray emission for the radio bridge with a point-to-point analysis. We also detect radio surface brightness enhancements around bright AGN jets, which are an indication that fossil plasma has been left by past AGN activity. This might also scatter the radio surface brightness distribution and therefore weaken the correlation between the radio and X-ray emission in a point-to-point analysis. At the same time, this fossil plasma is necessary for in situ re-acceleration. Together with the already constrained steep-spectrum ($\alpha > 1.5$; Nunhokee et al., 2021) from Govoni et al. (2019), these observations make a case for Fermi-II re-acceleration to explain the origin of the radio bridge.

- ☐ We obtain similar results from the point-to-point analyses in the radio bridges

in A1758 and A399-401. This suggests that these radio bridges might have similar origins.

- By applying the point-to-point analysis to the radio halo from A401, we find a strong correlation between the radio and X-ray emission. Together with signs of AGN activity in the radio halo and its steep spectrum ($\alpha = 1.63 \pm 0.07$; Govoni et al., 2019), we argue that it is likely that the emission from this halo originates in Fermi-II re-acceleration.

- We see the effects of a recent merger in A399 in a weaker radio and X-ray correlation compared to what we find for A401. However, we find a strong correlation in the northwest radio halo extension. We, therefore, argue that this observation, together with the steep spectrum from the radio halo in A399 ($\alpha = 1.75 \pm 0.14$; Nunhokee et al., 2021), is in favour of a scenario in which Fermi-II re-acceleration through turbulence is the main mechanism to explain the origin of the emission.

- We suspect that the two earlier classified radio relics by Govoni et al. (2019) might be re-energised fossil plasmas from earlier AGN activity. This supports the importance of re-acceleration and fossil plasma as drivers of the diffuse emission in A399-401.

Our work shows the power of refining the calibration and imaging of data from LOFAR to help us study the diffuse emission between pre-merging clusters. With our results, we can conclude that re-acceleration through turbulence and current and past AGN activity are likely important ingredients to explain most of the radio emission in A399-401 and possibly other radio bridges as well.

2.A. Computing recalibration

The images for A399-401 were produced with the recalibration method described in Section 2.2.2. We used 24 boxes within 0.6 degrees from the pointing centre. All the extractions and self-calibrations were done using processor nodes on Spider[7], which is a high-throughput data-processing platform from SURF[8] and allows to run parallel jobs.

The total number of CPU core hrs (processor units multiplied by job hrs) for the recalibration is 50336. In Figure 2.11 we see that the self-calibration almost used 2/3th of the total CPU core hrs to process the data, while the extraction used

[7] `https://spiderdocs.readthedocs.io/`
[8] `www.surf.nl`

around 1/3th. The imaging is the smallest component in the recalibration. The self-calibration and extraction step for every individual box can only be run in a serial manner, while the boxes can run in parallel of each other. This means that in the optimal case, we can speed-up with a factor of 24 for 24 boxes. However, because of the finite size of the Spider cluster and the job queue that enables sharing of the compute resources amongst many projects competing for the same resources the actual speedup achieved was a factor ~20 in real-time.

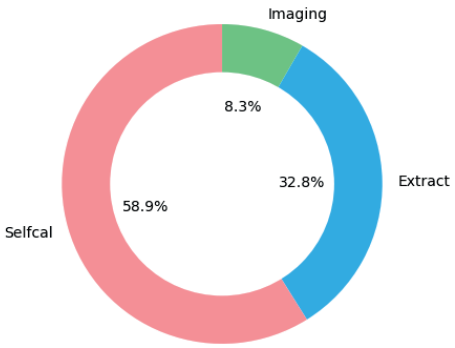


Figure 2.11: CPU core hrs in percentage from the total amount for self-calibration, extraction, and imaging for making the image from A399-401. This is based on the calibration of A399-401 with 24 boxes within 0.6 degrees from the pointing centre.

2.B. DDF-Pipeline versus recalibration for A399-401

In Section 2.2.2 we described why we decided to use a more expensive calibration method than the automated DDF-Pipeline, which is being used by the LoTSS pipeline (Shimwell et al., 2019; Tasse et al., 2021; Shimwell et al., 2022). We compare the final image from the DDF-Pipeline with our final recalibrated image that was produced with the same observations. This is not an entirely fair one-to-one comparison, as the weighting scheme in the two methods is different because of the Briggs weighting implementation in WSClean and the DDF-Pipeline image is made with its standard 100 m baseline uv-cut. With WSClean, we obtain a resolution of $5.9'' \times 10.5''$, while the DDF-Pipeline uses DDFacet and has a resolution of $6'' \times 6''$. The resulting noise levels are similar with our $\sigma = 79$ μJy beam^{-1} versus DDF-Pipeline with a lower $\sigma = 72$ μJy beam^{-1}. The individual images in Figure 2.12 show that artefacts around bright compact sources are better suppressed with our calibration method in most cases, and diffuse emission is better reconstructed (a clear exception from the improvement is the right-tailed source in the middle

panel). To quantify the artefact reduction, we also studied several bright compact sources and found an improvement in a large dynamic range $\left(\frac{\text{pixel}_{\text{max}}}{|\text{pixel}_{\text{min}}|}\right)$ in most cases, by a factor of ~1.6 on average.

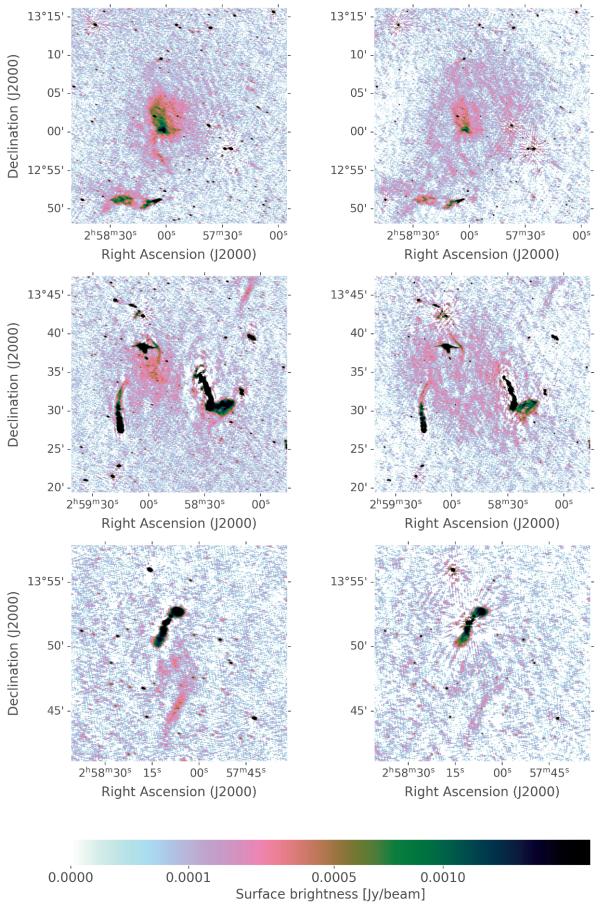


Figure 2.12: Image comparison between recalibration and the `DDF-Pipeline`. *Left column*: Images recalibrated with the method from this paper (see Section 2.2.2). *Right column*: Same images, produced with the `DDF-Pipeline` at the same colour scale. The first row shows Abell 399, the second row shows Abell 401, and the last row shows the radio galaxy 2MASX J02581043+1351519 with its extended diffuse emission tail. The square-root scaled colour bar extends from 0 to 25σ on average (average σ from both maps).

2.C. `Halo-FDCA` results

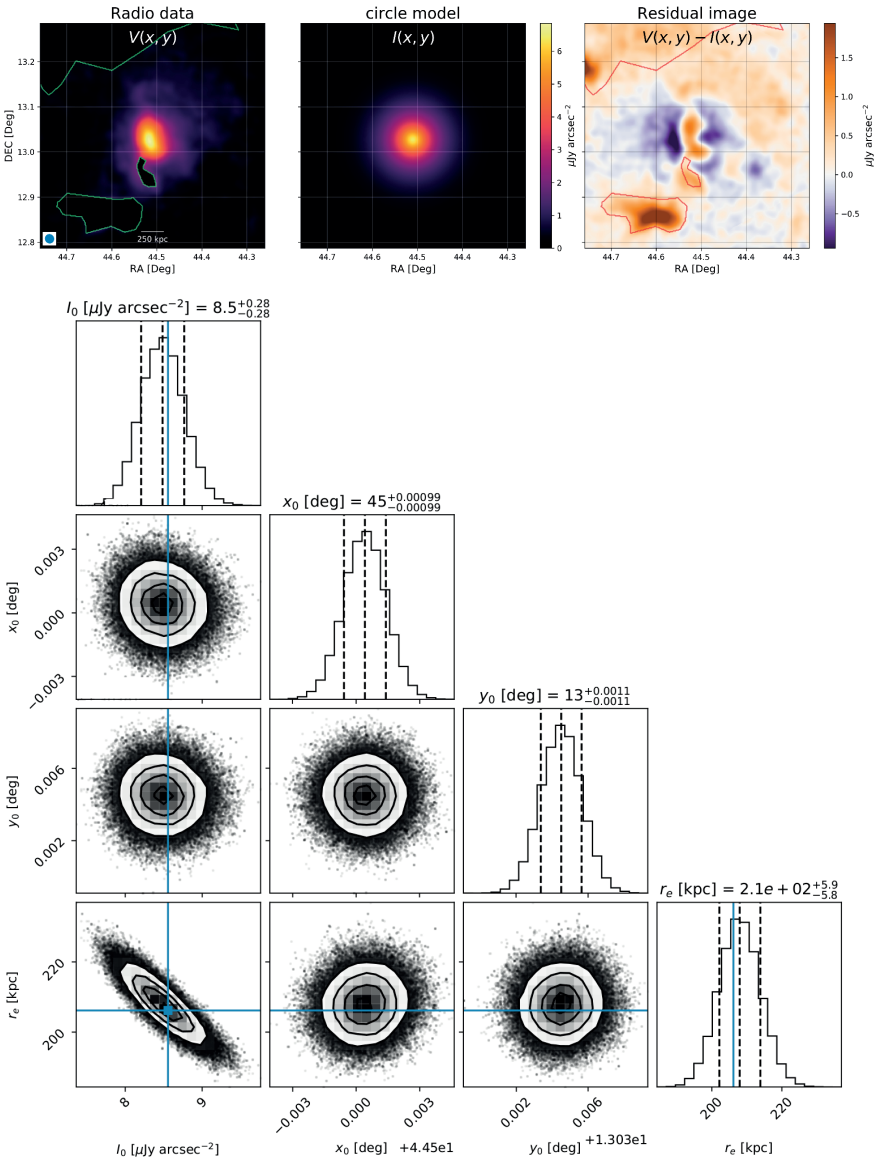


Figure 2.13: Results obtained from fitting the radio halo in A399 with `Halo-FDCA` (Boxelaar et al., 2021). *Top panel*: Image for the overlay fit with corresponding masks on bright AGNs. *Lower panels*: Markov chain Monte Carlo corner plot with the distributions of the posteriors of each fitted parameter.

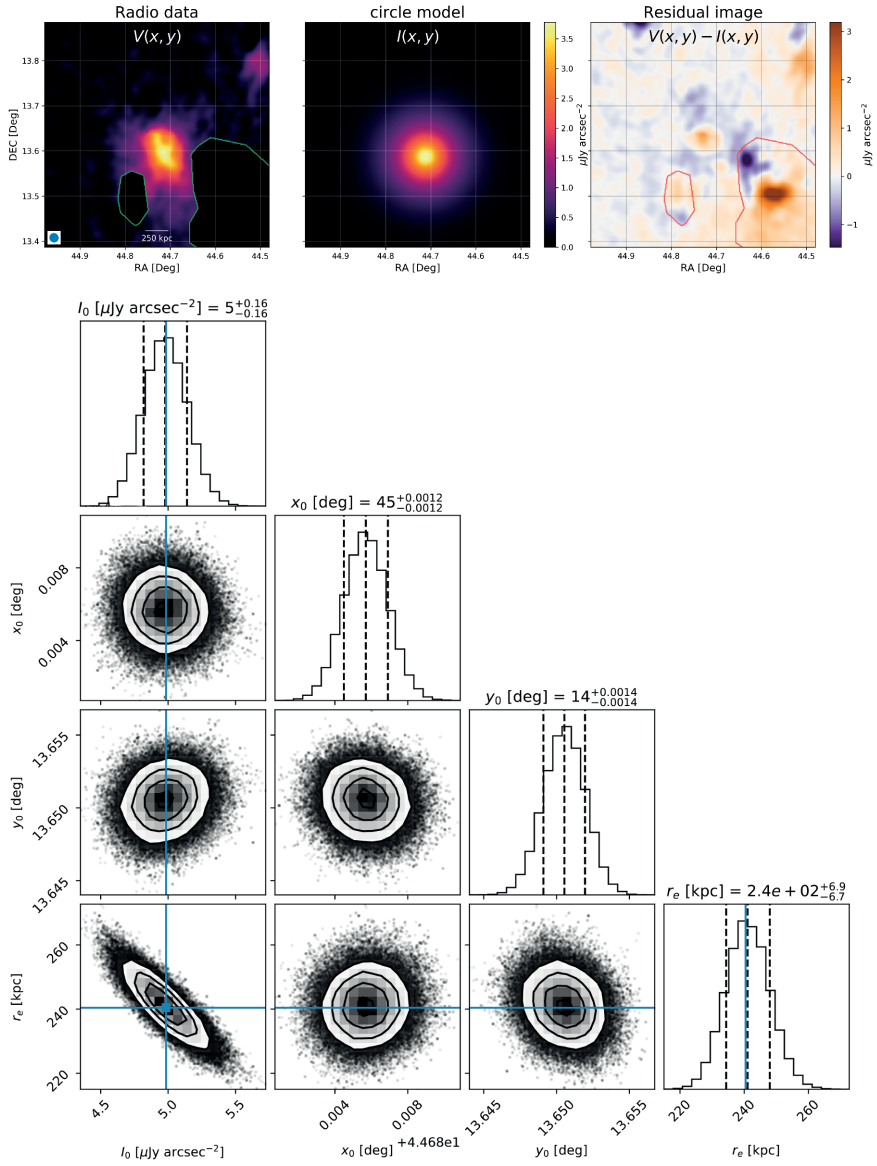


Figure 2.14: Results obtained from fitting the radio halo in A401 with `Halo-FDCA` (Boxelaar et al., 2021). *Top panel*: Image for the overlay fit with corresponding masks on bright AGNs. *Lower panels*: Markov chain Monte Carlo corner plot with the distributions of the posteriors of each fitted parameter.

Acknowledgements

This publication is part of the project CORTEX (NWA.1160.18.316) of the research programme NWA-ORC which is (partly) financed by the Dutch Research Council (NWO). This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-1287. RJvW acknowledges support from the ERC Starting Grant ClusterWeb 804208. AB acknowledges support from the VIDI research programme with project number 639.042.729, which is financed by the Netherlands Organisation for Scientific Research (NWO), and from the ERC Starting Grant DRANOEL 714245. RC acknowledges support from INAF mainstream project 'Galaxy Clusters Science with LOFAR' 1.05.01.86.05. This paper is based on data obtained with the International LOFAR Telescope (ILT). LOFAR (van Haarlem et al., 2013) 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; The IstitutoNazionale di Astrofisica (INAF), Italy. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian centre for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation. Based on observations obtained with XMM-Newton, an ESA science mission with instruments and contributions directly funded by ESA Member States and NASA.

3

Cosmic evolution of FRI and FRII sources out to $z = 2.5$

J.M.G.H.J. de Jong, H.J.A. Röttgering, R. Kondapally, B. Mingo, R.J. van Weeren, P.N. Best, L.K. Morabito, M. Magliocchetti, J.B.R. Oonk, A. Villarrubia-Aguilar, F.F. Vecchi

Astronomy & Astrophysics, Volume 683, id.A23, 17 pp.

Abstract

Radio-loud active galactic nuclei (RLAGN) play an important role in the evolution of galaxies through the effects on their environment. The two major morphological RLAGN classes are core-bright (FRI) and edge-bright (FRII) sources. With the sensitive LOw-Frequency ARray (LOFAR), we can now compare the evolution of FRI and FRII sources down to lower flux densities and with larger samples than before. Using sensitive LOFAR data, we examine in this work the cosmic space density evolution for FRIs and FRIIs by analyzing their space density evolution between $L_{150} \sim 10^{24.5}$ W Hz^{-1} and $L_{150} \sim 10^{28.5}$ W Hz^{-1} and up to $z = 2.5$. To study the space densities as a function of radio luminosity and redshift, we construct radio luminosity functions (RLFs) and simulate how sources appear across a range of redshifts, such that we can better correct for selection effects. Our RLFs do not show any sharp transitions between the space density evolution of FRI and FRII sources as a function of radio luminosity and redshift. We report a space density enhancement from low to high redshift for FRI and FRII sources brighter than $L_{150} \sim 10^{27}$ W Hz^{-1}. This can be explained by the higher gas availability in an earlier, denser universe. The FRI/FRII space density ratio does not appear to evolve strongly as a function of radio luminosity and redshift, which suggests that the jet-disruption of FRIs might be primarily caused by events occurring on scales within the host galaxy, rather than being driven by changes in the overall large-scale environment. Furthermore, while we observe a tentative decrease in the space densities of FRIs with luminosities below $L_{150} \sim 10^{26}$ W Hz^{-1} and at redshifts beyond $z = 0.8$, this may be due to residual selection biases. This highlights the need to resolve more sources at angular scales below $40''$, thereby strengthening the motivation for further developing and automating the calibration and imaging pipeline of LOFAR data to produce images at sub-arcsecond resolution.

3.1. Introduction

It is believed that most (if not all) galaxies host a supermassive black hole (SMBH) (Magorrian et al., 1998; Kormendy & Ho, 2013), whereby some of them are powered by an accretion disk around it (Lynden-Bell, 1969; Kauffmann et al., 2003; Martini et al., 2013). We classify these objects as active galactic nuclei (AGN). They are referred to as radio-loud AGN (RLAGN) when they also produce powerful collimated jets emitting at radio frequencies, due to gas falling onto the SMBH and interacting with its magnetic field, thereby generating synchrotron emission (see review by Hardcastle & Croston, 2020, and references therein). Thus, RLAGN play an important role in the evolution of the Universe, as their jets heat their environment and accelerate high energy cosmic rays. They might, in fact, be responsible for a large segment of the intergalactic magnetic field (Blandford et al., 2019). In addition, RLAGN influence the evolution of their host galaxy through feedback processes, where the energy released by the AGN can prevent cooling or expel gas. This may regulate or even quench star formation and therefore affect the growth of the host galaxies (e.g. Croton et al., 2006; Best et al., 2006; McNamara & Nulsen, 2007; Cattaneo et al., 2009; Fabian, 2012; Morganti, 2017; Hardcastle & Croston, 2020). Moreover, there also seems to be a link between the radio loudness and the host morphology, where the most powerful RLAGN are hosted by massive geometrically round-shaped galaxies (Barišić et al., 2019; Zheng et al., 2020, 2022). All of the above are reasons to argue for the importance of studying the cosmic evolution of RLAGN.

The morphologies of RLAGN can be classified based on the distance from the central galaxy to the brightest point of their jets (Fanaroff & Riley, 1974). The two main classes are the core-bright FRI morphologies and the edge-bright FRII morphologies. Fanaroff & Riley found in their sample a 'break luminosity' at a radio power of $L_{150} \sim 10^{26}$ W Hz^{-1}, where FRIs dominate below this luminosity value and FRIIs above. In the study by Ledlow & Owen (1996), it was demonstrated that the break luminosity increases for increasing host optical luminosity. This observation suggested a direct link between the morphology of radio jets and their environmental density. As a result, it was proposed that RLAGN initially exhibit an FRII morphology, but later transition to an FRI morphology when they are unable to traverse the local interstellar medium, which breaks the jet-collimation and decelerates the jet speed on kpc-scales from their host galaxy (Bicknell, 1995; Kaiser & Best, 2007). More recent studies, based on samples of radio sources less affected by selection effects compared to older studies, showed that the break luminosity does not separate FRI and FRII morphologies as strictly as was initially observed (Best, 2009; Gendre et al., 2010, 2013; Wing & Blanton, 2011; Capetti et al., 2017b; Mingo et al., 2019, 2022). The FRI/FRII morphological divide is therefore less

closely connected to radio luminosities than previously suspected. Nonetheless, the existence of FRII below the original break luminosity remains consistent with the jet-disruption model if this population is a mix of restarting FRIIs, (old) fading FRIIs, or FRIIs hosted by less massive host galaxies in less dense environments compared to FRIIs above the break luminosity (Croston et al., 2019; Mingo et al., 2019). Examining the cosmic evolution of FRI and FRII morphologies is therefore linked to studies of the evolution of the large-scale environments of radio galaxies (e.g. Croston et al., 2019).

The changing interpretation of the break luminosity demonstrates how selection biases play an important role in the ability to detect FRI and FRII sources and then affect our understanding of radio galaxy evolution as a result. Because FRIIs have hotspots at the edge of the lobes and are more powerful, they will be easier to detect than FRIs. This selection effect becomes stronger when we look at more distant RLAGN, where their jets become fainter and are closer to the flux density limit from the used instrument of the survey. It has been debated how much of the observed evolution of the FRI and FRII sources is due to selection effects and how much is due to true evolution of the radio galaxy population (Singal & Rajpurohit, 2014; Magliocchetti, 2022). The difficulty in detecting FRIs at high redshifts, with very limited sample sizes, has led to the prediction that powerful FRIs should be significantly more abundant at $z > 1$ compared to what we find locally (Snellen & Best, 2001; Jamrozy, 2004; Rigby et al., 2008). A broader study of the FR dichotomy, with the combined NVSS-FIRST (CoNFIG) catalogue (Gendre & Wall, 2008), has strengthened this prediction by finding positive space density enhancements from low to high redshift up to a factor of 10 over the local population for both FRIs and FRIIs up to $z = 2.5$ (Gendre et al., 2010, 2013). To understand the role of extended RLAGN across cosmic time, it is vital to fully characterise the cosmic evolution of the FRI and FRII sources using much larger samples and down to fainter radio luminosities than previous studies.

In this paper, we investigate the FRI/FRII space density evolution as a function of redshift and radio luminosity by constructing radio luminosity functions (RLFs) up to $z = 2.5$ with the catalogues from Mingo et al. (2019, 2022). These catalogues are based on the 150-MHz LOw-Frequency ARray (LOFAR; van Haarlem et al., 2013) surveys that have a high surface brightness sensitivity and are therefore great at detecting FRIs with low flux densities. By correcting more precisely for selection effects through redshift simulations, we derive better estimates of the maximum volume that a source can still be observed and classified over, before falling below the detection and selection limits of the sample. These simulations, in combination with the theoretical angular size distribution, help us to better correct for the incompleteness of over sample, such that we can recover an improved estimate of the

'real' FRI/FRII RLFs. This comprehensive approach, along with 100 times more FRI redshifts above $z = 0.3$ compared to previous works from Gendre et al. (2010), is also helpful in improving the FRI and FRII RLF comparisons above this redshift and to re-examine the findings from Gendre et al. regarding the space density enhancements of FRIs and FRIIs up to $z = 2.5$.

In Section 6.5 of this paper, we briefly discuss the selected data and catalogues. This is followed by an explanation of our redshift simulation in Section 3.3, which is then applied in Section 3.4 to construct RLFs. We present our results in Section 6.6. We further discuss our results in Section 3.6 and look at future prospects to improve this work in the future. Finally, we give our conclusions in Section 3.7. Throughout this work, we use a ΛCDM cosmology model with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

3.2. Data

We utilised FRI/FRII sources classified by Mingo et al. (2019, with their catalogue hereafter referred to as M19,) combined with a similarly compiled catalogue based on deeper observations from Mingo et al. (2022, with their catalogue hereafter referred to as M22). Both catalogues were constructed with data from the LOFAR Two-metre Sky Survey (LoTSS Shimwell et al., 2017, 2019) at 144 MHz with a resolution of 6″and a sensitivity ranging from ~20 to ~70 μJy beam^{-1}. The sensitive low-frequency observations used to construct these catalogues excel in identifying complex extended sources, which benefits the detection of extended diffuse jets from FRIs. In the following subsections, we discuss their content and our source selection.

3.2.1. Catalogues

Sources from LoTSS DR1 are included in M19, based on radio maps from the HETDEX Spring Field, covering 424 deg^2 with a median sensitivity of 71 μJy beam^{-1} (Shimwell et al., 2017, 2019; Williams et al., 2019), while M22 contains sources from the LoTSS-Deep Fields DR1, which are about three to four times deeper than LoTSS DR1 and covers 25 deg^2 (Tasse et al., 2021; Sabater et al., 2021; Kondapally et al., 2021; Duncan et al., 2021; Best et al., 2023). The deep fields were selected with deep wide-area multi-wavelength imaging from ultraviolet to far-infrared (see Kondapally et al. (2021) for details), consisting of the following three fields: European Large Area Infrared Space Observatory Survey-North 1 (ELAIS-N1) (Oliver et al., 2000), Lockman Hole (Lockman et al., 1986), and Boötes (Jannuzi & Dey, 1999). The corresponding LoTSS-Deep Fields DR1 radio maps respectively exhibit a median sensitivity of about 20, 22, and 32 μJy beam^{-1} for ELAIS-N1, Lockman Hole, and Boötes (Tasse et al., 2021; Sabater et al., 2021).

For every source, the M19 and M22 catalogues contain their total 150-MHz flux density, size, and host galaxy redshift and position on the sky. The sizes and radio flux densities from sources in M19 and M22 were carefully measured by adopting a flood-filling procedure and by comparing their sizes and flux densities with what was found with the `PyBDSF` Gaussian fitting tool (See for more details Section 2.5 in Mingo et al., 2019).[1] These comparisons were followed up by visual inspection. The sources from the catalogue from LoTSS DR1 to construct M19 are for 73 % identified with an optical host (Williams et al., 2019) and have 51 % spectroscopic or photometric redshifts (Duncan et al., 2019). Sources in the deep fields were over ~97% identified with radio host-galaxies associated with carefully determined photometric redshifts (or spectroscopic where available) using a hybrid approach of template fitting and machine learning methods (Duncan et al., 2019, 2021). Mingo et al. included in M22 only sources from the deep fields up to $z = 2.5$, as the spectral energy distribution (SED) fitting of all the detected radio sources in the deep fields was considered reliable for sources up to this redshift (Best et al., 2023). The FRI and FRII classifications of these sources were done using the `LoMorph` classification code and by additional visual inspection (Mingo et al., 2019).[2] They reported a classification accuracy of 89% for FRIs and 96% for FRIIs. Mingo et al. excludes sources with projected angular sizes smaller than 27″ or below 40″, with a 20″or smaller distance between their two peak emissions. This angular size cut is based on classification difficulties and the resolution limits from LoTSS DR1 and the LoTSS-Deep Fields DR1. Also, other sources that did not clearly fit in FRI or FRII sources were filtered out (see Section 2.4 from Mingo et al., 2019). Those contain double-double sources (restarting FRII) (Schoenmakers et al., 2000) and hybrid sources (Gopal-Krishna & Wiita, 2000).

3.2.2. Source selection

The sources from M19 with reliable redshifts above $z = 0.8$ are primarily quasars. This is mainly due to the scarcity of spectroscopic data for radio galaxies and the line-of-sight bias (Duncan et al., 2019; Hardcastle et al., 2019b). As a result, we have excluded sources with redshifts beyond $z = 0.8$ from M19. We do not need to make additional redshift cuts for M22, as the radio galaxies from M22 are drawn from the significantly deeper LoTSS-Deep Fields DR1, in contrast to LoTSS DR1, which offers precise redshifts up to $z = 2.5$ (Kondapally et al., 2021; Duncan et al., 2021; Best et al., 2023). Consequently, all the sources utilised in this paper with redshifts between $z = 0.8$ and $z = 2.5$ originate exclusively from M22 and make up only ~9% of our full sample. Although the sources collected from the LoTSS-Deep

[1] `https://pybdsf.readthedocs.io/`
[2] `https://github.com/bmingo/LoMorph`

Fields DR1 have higher redshift completeness than the sources collected from LoTSS DR1, we find the ratio of the number of FRI/FRIIs for the overlapping redshifts below $z < 0.8$ for M19 and M22 to be similar. Also, the redshift distributions below $z < 0.8$ are similar, that is, we find their redshifts to have the same mean and standard deviation. This demonstrates that combining M19 and M22 does not introduce additional selection effects as a function of redshift.

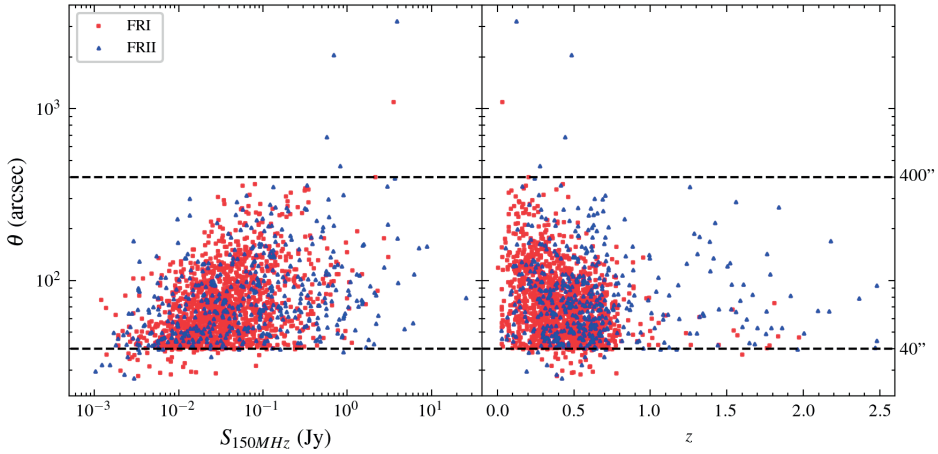


Figure 3.1: Angular size as a function of flux density (left panel) and redshift (right panel) for the 1560 sources considered in this paper. We separated our sources into 1146 FRI and 414 FRII morphologies with separated colours and drawn dashed lines for 40″ and 400″. These lines determine the angular size cuts that we applied in this paper before doing completeness corrections (see Section 3.4.2).

In Figure 3.1, we plot the angular size versus the radio flux density and redshift for all our FRIs and FRIIs. From this figure, we see that due to the selection criteria by Mingo et al., there are very few sources between 27″ and 40″in the sources from M19 and M22. Most of those sources are below $z = 0.8$. We also find a clear gap of sources above 400″, where only 0.3% of our sources are situated. This is the area where classification of bright giant radio galaxies (GRGs) becomes difficult because different selection biases for FRIs and FRIIs do play a role in relation to the surface brightness sensitivity of the used observation. First of all, the most distant parts of the FRI jets are close to or below the surface brightness limit, which (in some cases) would make only their unresolved core appear or make the measurements of their sizes smaller than they actually are. This issue becomes more prominent at the higher end of our redshift distribution where surface brightness limits play a more dominant role due to the fact that sources are on average fainter as a function of redshift. This effect can also be observed in the right panel of Figure 3.1, where FRIs are closer positioned to the 40″ boundary compared to FRIIs at

higher redshifts. Secondly, large FRII jets close to the surface brightness limit might appear to only have two bright disconnected hotspots. This poses challenges in associating them with each other, particularly when their separation is substantial. So, it is not unexpected that a significant proportion of large sources are missing in our catalogues (e.g. Oei et al., 2023). To account for the 40″ and 400″ angular size limits, we excluded sources that fall below and above these threshold values and subsequently implemented completeness corrections as a function of angular size and flux density to regain the contribution of sources beyond these boundaries to our RLFs, as we discuss in more detail in Section 3.4.2.2.

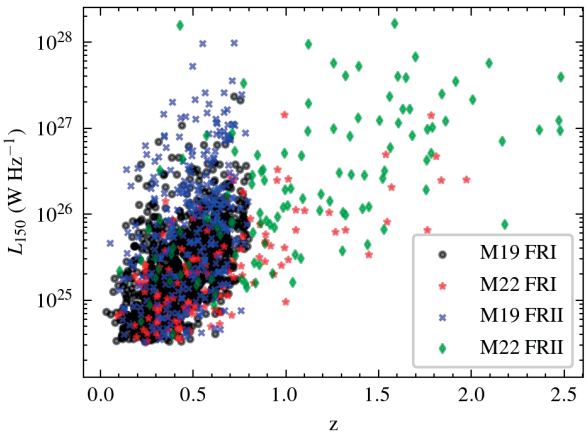


Figure 3.2: Luminosity-redshift diagram for M19 and M22 split in FRIs and FRIIs.

We further limited our analyses to sources with a radio luminosity above $L_{150} \sim 10^{24.5}$ W Hz^{-1}, as the number of sources below this radio luminosity is small. This also ensures that our completeness corrections are not strongly affected by star-forming galaxies (e.g. Sabater et al., 2019; Cochrane et al., 2023) or compact sources (e.g. Sadler, 2016; Baldi et al., 2018; O'Dea & Saikia, 2021). Moreover, due to the LoTSS flux density limits, beyond $z = 0.8$, we did not find any sources below $L_{150} \sim 10^{24.5}$ W Hz^{-1}. Thus, this decision improves the reliability of completeness corrections and, therefore, our resulting RLFs without compromising the goal of this paper, which is to investigate the evolution of radio galaxies across redshifts up to $z = 2.5$. Figure 3.2 shows the final selection of the 1560 sources with their 150 MHz radio luminosities as a function of redshift. From those, 1146 are classified as FRI sources and 414 as FRII sources.

3.3. Redshift simulations

Detection limits from radio telescopes restrict our view of the Universe. In order to incorporate these effects into the derivation of RLFs for FRIs and FRIIs, we have developed an algorithm that simulates how sources would appear after relocating them to higher redshifts, so that we can derive up to which maximum distance a source remains reliably classifiable. These redshift simulations will also help us to determine the incompleteness of our sample as a function of flux density and redshift.

3.3.1. Surface brightness and redshift relation

We can relate the total flux density (S_ν) of a source to its luminosity (L_ν) and its luminosity distance (d_L) as:

$$S_\nu = \frac{L_\nu (1+z)^{1-\alpha}}{4\pi d_L^2}, \tag{3.1}$$

where $(1+z)^{1-\alpha}$ accounts for the K-correction with spectral index α and where the radio flux density is $S_\nu \propto \nu^{-\alpha}$. The angular size θ_z of a source is related to its physical size s and the angular distance (d_A) as:

$$\begin{aligned} \theta_z &= \frac{s}{d_A} \\ &= \frac{s}{d_L} \cdot (1+z)^2, \end{aligned} \tag{3.2}$$

where we used the relation $d_A = \frac{d_L}{(1+z)^2}$. Combining the above, we derive the surface brightness evolution over redshift as:

$$\begin{aligned} \Sigma &= \frac{S_\nu}{\theta_z^2} \\ &= \frac{L_\nu}{4\pi s^2} \cdot (1+z)^{-3-\alpha}. \end{aligned} \tag{3.3}$$

As the luminosity and physical size are physical intrinsic properties of a radio galaxy, this demonstrates that the surface brightness decreases over redshift by a factor $(1+z)^{3+\alpha}$.

3.3.2. Redshifting algorithm

If we could relocate a source to a higher redshift, we would see the angular size and flux density from the source in the image change. The source would diminish in size according to the adopted cosmology until approximately $z \sim 1.6$, after which it would begin to increase once again. These angular size changes affect the ability to detect and classify FRIs and FRIIs at the same radio powers differently, due to their different morphologies. On the one hand, the FRI will be classifiable until the diffuse jets are not bright enough to be detected or the source cannot be resolved anymore due to its small angular size. On the other hand, the FRII remains classifiable as long as the hotspots appear as point sources, the components can be associated with the host, and we have the adequate resolution to separate them, while they are still bright enough above the noise. Additionally, there are situations where FR classifications can change after redshift increments from FRI to FRII and vice versa, due to the detection limits and the changing number of pixels covering the source. To simulate and better understand these effects, we developed the `redshifting` algorithm to increment their redshift by Δz.[3]

The `redshifting` algorithm takes as input an image of a source at its observed redshift. After every redshift increment, the algorithm changes the pixel sizes by multiplying the pixel scale via $\frac{\theta_z}{\theta_{z+\Delta z}}$, which is equivalent to dividing the pixel scale by the ratio between d_A at z and d_A at $z+\Delta z$. This information is used to resample the image pixels with the DeForest (2004) resampling algorithm.[4] In this way, the source appears smaller or larger after redshift increments, depending on the redshift. Following from Equation 3.3, the algorithm reduces every pixel value by a factor $\left(\frac{1+z}{1+z+\Delta z}\right)^{3+\alpha}$. Because the image noise is independent of the source's redshift, we kept it constant during our procedure of relocating a source to a higher redshift. This means for our algorithm that we only reduce pixel values that exceed our 3σ noise threshold.

We adopted a constant spectral index of $\alpha = 0.7$ for the K-correction. This spectral index value is the typical average spectral index for RLAGN and is often used to construct radio luminosity functions if the spectral indices of sources are unknown (e.g. Condon et al., 2002; Mauch & Sadler, 2007; Padovani, 2016; Prescott et al., 2016; Hardcastle et al., 2016; Calistro Rivera et al., 2017; Kondapally et al., 2022). A typical spectral index variation between $\alpha = 0.6$ or $\alpha = 0.8$ (e.g. Hardcastle et al., 2016; Calistro Rivera et al., 2017; Murphy et al., 2017) would (compared to $\alpha = 0.7$) only adjust the pixel brightness changes from our redshifting algorithm up to $\sim 5\%$ for redshift increments of $\Delta z = 0.7$. This is the mean Δz from their original z where we find that sources are still classifiable after applying the redshift-

[3] `https://github.com/jurjen93/redshifting`

[4] `https://github.com/astropy/reproject/`

ing algorithm (see the following sections) for the sources considered in this paper. Hence, it is not expected that the choice for $\alpha = 0.7$ will strongly bias our results.

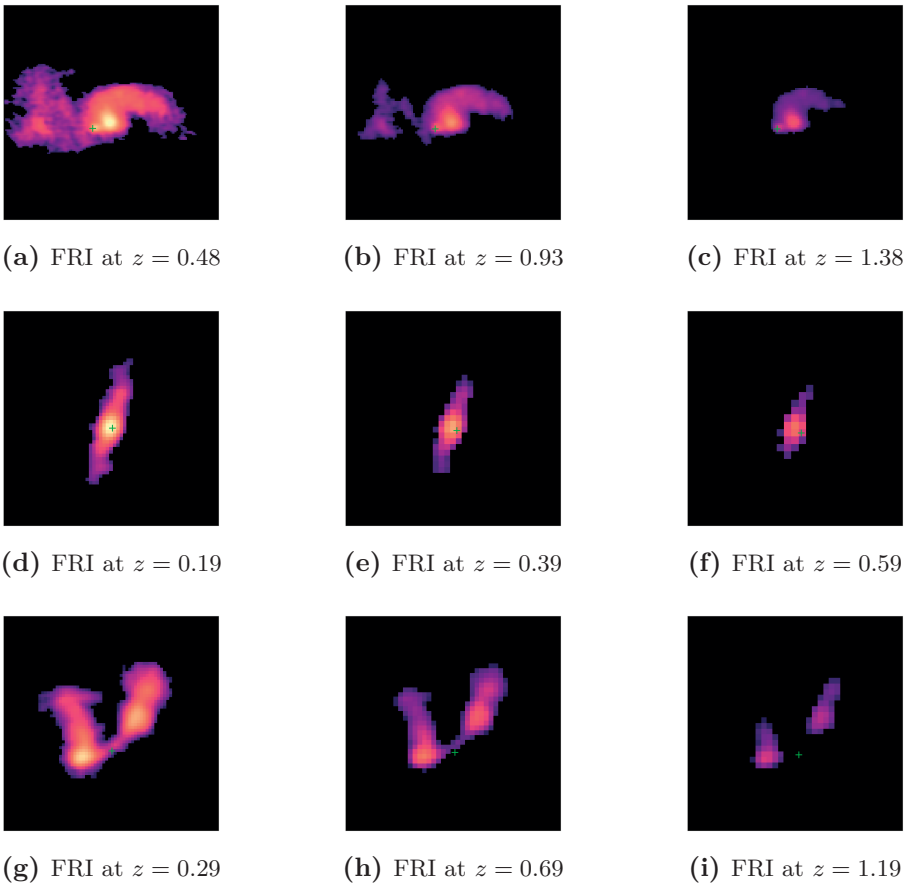
We also made sure that the total power remains constant as a function of redshift increments. This implies that we do not correct for inverse Compton (IC) scattering losses from the cosmic microwave background (CMB), which is expected to add additional selection biases (Krolik & Chen, 1991; Morabito & Harwood, 2018; Sweijen et al., 2022a). We further discuss this aspect in Section 3.6.1 regarding the implications of this choice on the interpretation of our RLFs when we do not consider this effect.

3.3.3. Source components

The association of radio source components is a difficult task and massive visual inspection through citizen science projects or neural network architectures are often used to do this for large amounts of data (Banfield et al., 2015; Williams et al., 2019; Mostert et al., 2022).[5] For all our sources, we already have initial information about the source components (based on `PyBDSF` and visual inspection, see Mingo et al., 2019). So, an additional step for associating source components is not necessary, as our focus is solely set on determining the components that remain above the noise threshold at each redshift increment. All unassociated components in the image are masked out in every source cutout image. This is necessary when multiple radio sources are close to each other, or when there are remaining islands of emission contaminating the image for classification. After relocating a source to a new redshift with `redshifting`, we draw polygons around the remaining components with emission above a particular noise level.

The sources in Figures 3.3 and 3.4 demonstrate the effect of the redshifting increments on the observed morphology for different FRIs and FRIIs. We see, in particular, how their brightness is reduced and how their appearance changes. The first two rows of Figure 3.3 display FRI sources, where their jets partly disappear below the σ-threshold when they are relocated at higher redshifts. In the last row, we have a special case where an FRI source appears as an FRII after large redshift increments. The sources in Figure 3.4 are typical cases of FRIIs with clear hotspots. These sources show how the FRII hotspots are the main components that remain visible after relocating them at higher redshifts. We can see from the first and the last row how the elongated FRII component separates into two components at higher redshifts.

[5] See also the LOFAR Galaxy Zoo project: `https://www.zooniverse.org/projects/chrismrp/radio-galaxy-zoo-lofar`



(a) FRI at $z = 0.48$ **(b)** FRI at $z = 0.93$ **(c)** FRI at $z = 1.38$

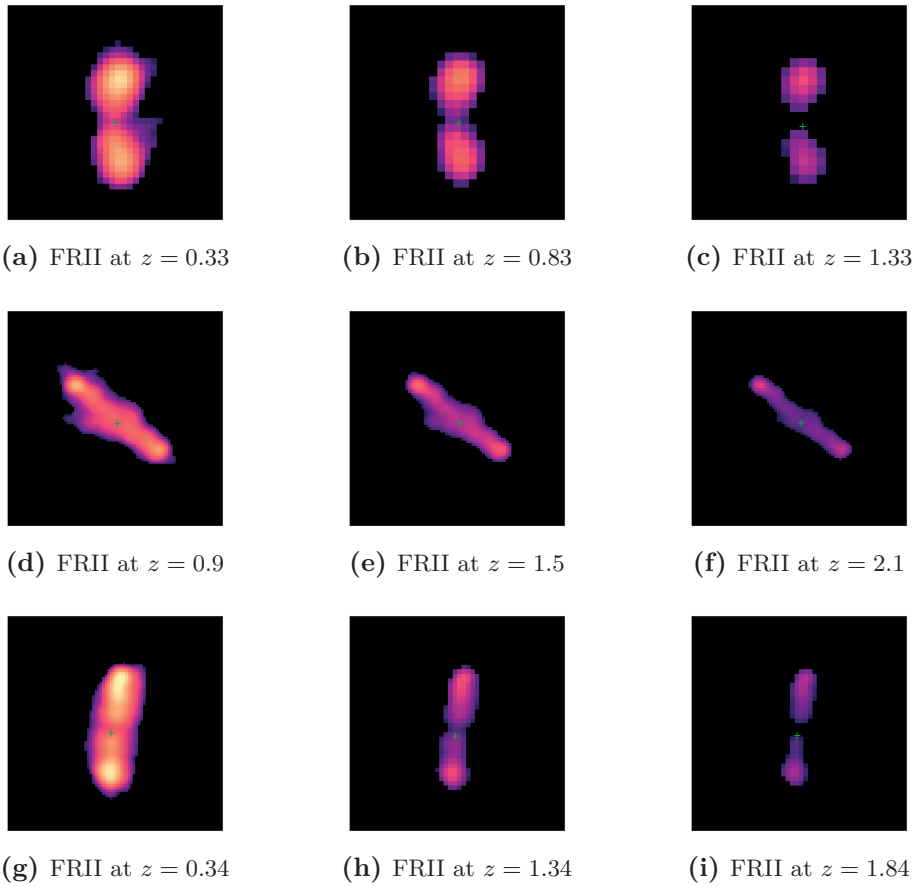
(d) FRI at $z = 0.19$ **(e)** FRI at $z = 0.39$ **(f)** FRI at $z = 0.59$

(g) FRI at $z = 0.29$ **(h)** FRI at $z = 0.69$ **(i)** FRI at $z = 1.19$

Figure 3.3: Examples of FRI sources after redshift increments by applying the `redshifting` algorithm. All unassociated emission is masked out (in this case below 4σ). Each column displays the source from the same row at different redshifts. The last row is a wide-angle tail FRI source (Owen & Rudnick, 1976), where after redshift increments, the source appears to resemble an FRII source. The optical host location is in every image given by the green cross.

3.3.4. FR classification

Because we use only sources from M19 and M22, we based our FRI/FRII classification code on the `LoMorph` code, which were used to classify our sources. This classification algorithm is based on measuring the distance from the optical host to the brightest region and the edge of the source by using flood-filling and verifying if a source is resolved or unresolved by applying `PyBDSF`. For a more detailed discussion, we refer to Section 2 from Mingo et al. (2019).



(a) FRII at $z = 0.33$ **(b)** FRII at $z = 0.83$ **(c)** FRII at $z = 1.33$

(d) FRII at $z = 0.9$ **(e)** FRII at $z = 1.5$ **(f)** FRII at $z = 2.1$

(g) FRII at $z = 0.34$ **(h)** FRII at $z = 1.34$ **(i)** FRII at $z = 1.84$

Figure 3.4: Examples of FRII sources after redshift increments by applying the `redshifting` algorithm. All unassociated emission is masked out (in this case below 4σ). Each column displays the source from the same row at different redshifts. The optical host location is in every image given by the green cross.

The classification of the source on the third row in Figure 3.3 changes both visually and in our algorithm from an FRI to an FRII morphology after being incremented to a higher redshift. This classification inconsistency occurs for about ~2% of all our sources when relocating them to higher redshifts. FRIIs could also become classified as FRIs due to the resampling of pixels, when the hotspots of an FRII would merge into a bright FRI-type core. We did not find such cases in our simulations, indicating that before this occurs, the source would no longer be classifiable. This could also be because double-double and hybrid sources have already been all filtered out from our sources (see Section 3.2.2 and Mingo et al.

(2019, 2022)). These classification inconsistencies are rare and thus demonstrate that our morphological classification is generally effective.

3.4. Constructing radio luminosity functions

RLFs are an important tool in studying the evolution of radio sources as a function of power and redshift. To construct the RLFs in this paper, we used the standard method from Schmidt (1968). This means that we calculated the density evolution as a function of redshift z and radio luminosity L via:

$$\rho(z, L) = \frac{1}{\Delta \log L} \sum_{n=1}^{N} \frac{1}{V_{\mathrm{max,n}}}, \tag{3.4}$$

where $V_{\mathrm{max,n}}$ is the maximum volume that source n can be classified and where $\Delta \log L$ is the log value of the radio luminosity bin. In the next subsections, we further discuss the V_{max} and the completeness corrections we applied to constructing the RLFs.

3.4.1. V_{max} method

To determine V_{max}, we first need to find an accurate value for the maximum redshift at which a source can still be classified (z_{max}). This value can be found by applying the `redshifting` algorithm discussed in Section 3.3.2.

3.4.1.1. Determining z_{max}

We find z_{max} by spatially relocating each source across the radio image 200 times and re-evaluating up to which maximum redshift the source is still classifiable at each spatial location with the `redshifting` algorithm ($\tilde{z}_{\mathrm{max}}$). In this way, we get a reliable measure for z_{max} by taking the mean of all the obtained $\tilde{z}_{\mathrm{max}}$ values and can find the error on this value by taking the standard deviation of the $\tilde{z}_{\mathrm{max}}$ values ($\sigma_{z_{\mathrm{max}}}$). The value for $\sigma_{z_{\mathrm{max}}}$ embeds information about the effects of the varying noise in the radio maps.

With the obtained z_{max} values, we can compare how FRIs and FRIIs are differently affected by the `redshifting` algorithm. For this, we can for example take a local sample up to $z = 0.5$ and move those up to $z = 2.5$. In Figure 3.5 we show from this example the fraction of the remaining sources that are still detectable and classifiable in steps of $z = 0.1$. Only 50% of the FRI sources we start with remain classifiable up to $z \sim 0.6$, while for FRII sources we still classify 50% of the sources up to $z \sim 1$. This is partly explained by the different effects detection limits have on

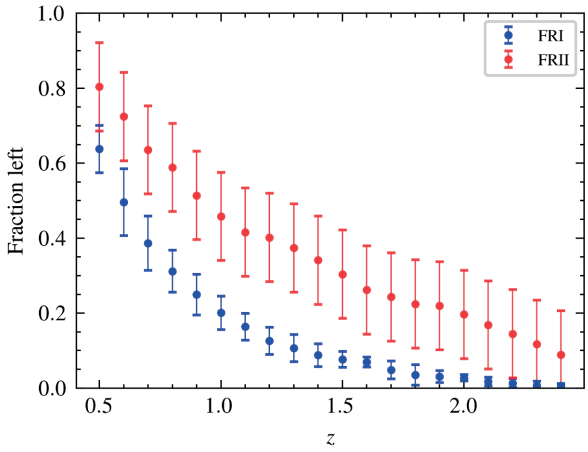


Figure 3.5: Fraction of local ($z < 0.5$) FRI and FRII sources that are still classifiable after applying redshift increments with the `redshifting` algorithm as a function of redshift. The error bars include the errors on $z_{\rm max}$ and Poisson errors.

the observed morphologies of FRI and FRII sources (see Section 3.3.2). In addition, FRII sources are on average also brighter as they are locally more abundant above the break luminosity (Fanaroff & Riley, 1974; Mingo et al., 2019). We demonstrate this in Figure 3.6 where we show how 150 randomly picked sources move in $P - z$ space from their original observed z-position to their $z_{\rm max}$ position. On the higher luminosity end, we observe, as expected, how FRII sources reach higher $z_{\rm max}$ values.

3.4.1.2. Final integral for $V_{\rm max}$

The $z_{\rm max}$ determines the maximum redshift a source can be observed in for each redshift bin, $z_{\rm bin}$, over which we evaluate the following integral

$$V_{\rm max} = \int_{z_{\rm bin}} C(S_\nu)\Theta(z)\chi(z)V(z)\mathrm{d}z, \tag{3.5}$$

where $C(S_\nu)$ is the completeness correction as a function of flux density at a particular redshift (see Section 3.4.2), then $V(z)\mathrm{d}z$ is the infinitesimal comoving volume slice between z and $z + dz$, $\Theta(z)$ is the fractional sky coverage, while $\chi(z)$ is the binary function:

$$\chi(z) = \begin{cases} 1 & \text{if } z \leq z_{\rm max}, \\ 0 & \text{if } z > z_{\rm max}. \end{cases}$$

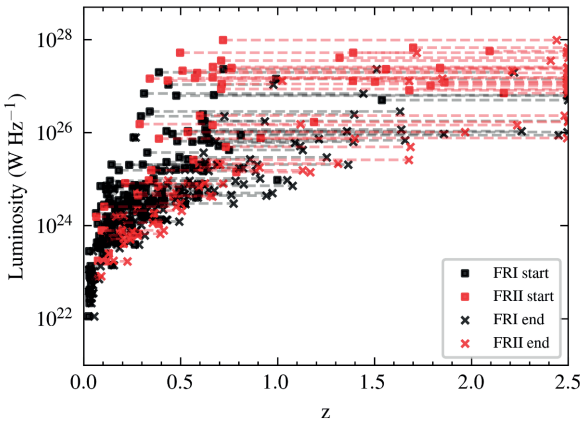


Figure 3.6: Luminosity versus redshift increments with the `redshifting` algorithm for 150 random sources at different luminosities from all sources from M19 and M22 above $L_{150} \sim 10^{22}$ W Hz^{-1}. The starting positions are at the original measured redshifts given by the squared boxes, while the ending positions are the final maximum redshifts a source can be classified (up to $z = 2.5$).

The uncertainties on $z_{\rm max}$ propagate in $V_{\rm max}$ by calculating $\tilde{V}_{\rm max}$ corresponding to every $\tilde{z}_{\rm max}$ with Equation 3.5 and taking the standard deviation of all the obtained values. The final uncertainties of the space densities also include Poisson errors following from Gehrels (1986) and the `LoMorph` classification accuracies from Mingo et al. (2019) (89% for the FRIs and 96% for the FRIIs).

As Figure 3.5 demonstrates how the FRIs are more strongly affected by selection effects, it also shows how with our $V_{\rm max}$ method the FRI space densities will be more strongly corrected as a function of z, compared to the space densities from FRII sources.

3.4.2. Completeness corrections

The $V_{\rm max}$ method enables measuring the space densities for sources within our sample as a function of redshift. We also need to apply a correction that takes into account the sources that are undetected due to the flux density and resolution limits from our observations. This is the (in)completeness factor as defined in Equation 3.5 by $C(S_\nu)$. For this work, this means that we need to estimate how many FRI and FRII sources we are missing as a function of flux density (S_{150}), such that we can better estimate the real space density values. Firstly, we calculate the completeness corrections by generating mock sources based on the sources from our source sample. Secondly, we determine the completeness correction for the sources with sizes below 40″ and above 400″. Combining both corrections gives our final

completeness correction.

3.4.2.1. Correction for $40'' < \theta < 400''$

We find the completeness corrections from our sources as a function of flux density by first generating mock FRI and FRII sources. Those mock sources are created by multiplying the flux density from a source in our catalogue with a random factor loguniformly drawn between 0 and 1 and scaling the pixel brightness in our image with the same factor. We then replace the mock source to the same noise environments as the noise effects are already embedded in the $\sigma_{z_{\max}}$ values. When we do this 200 times per source, we have about ~350k mock sources. Using the classification part from Section 3.3.4, we decide for each case if it is still recoverable as an FRI or FRII source. As the M19 and M22 samples have different sensitivities, we do this for both catalogues separately. With these simulations, we find the flux density incompleteness for sources corresponding to the angular size distribution from the objects in our combined M19 and M22 catalogues.

3.4.2.2. Correction for $\theta < 40''$ and $\theta > 400''$

As explained in Section 3.2.2, we miss sources smaller than $40''$ and larger than $400''$. To quantify the angular size incompleteness in our completeness weights, we use the empirical integral angular size distribution from Windhorst et al. (1990) with the updated parameters fitted to LoTSS-Deep Fields DR1 by Mandal et al. (2021). This is a radio source angular size distribution which has been used in multiple studies and has proven to be a reliable measure (Prandoni et al., 2001; Huynh et al., 2005; Hales et al., 2014; Mahony et al., 2016; Williams et al., 2016; Retana-Montenegro et al., 2018; Prandoni et al., 2018; Mandal et al., 2021; Kondapally et al., 2022). The angular size distribution follows as:

$$\Psi(>\theta) = \exp\left(-\ln 2 \left(\frac{\theta}{\tilde{\theta}}\right)^{0.8}\right), \tag{3.6}$$

where θ is in arcseconds and where $\tilde{\theta}$ is the median angular size which Windhorst et al. derived to be related to the flux density S_{1400} in mJy via

$$\tilde{\theta} = 2 \cdot k \cdot (S_{1400})^m, \tag{3.7}$$

where $m = 0.3 + 0.2 \cdot \exp\left(-S_{1400}^2\right)$ and

$$k = \begin{cases} 7 - 3 \cdot \exp\left(-\frac{S_{1400}}{2}\right) & \text{if } S_{1400} < 4.5, \\ 4 + 3 \cdot \exp\left(-\frac{S_{1400}}{200}\right) & \text{if } S_{1400} \geq 4.5, \end{cases}$$

where we doubled the value of k compared to Mandal et al. (2021), as Kondapally et al. (2022) showed for LoTSS-Deep Fields data the angular size distribution to fit well for AGN with twice the median angular size. By converting Equation 3.7 from $\nu = 1.4$ GHz to $\nu = 150$ MHz, using a spectral index of $\alpha = 0.7$, and combining this with Equation 3.6, we find the angular size distribution as a function of the flux density S_{150}.

The radio luminosity cut below $L_{150} = 10^{24.5}$ W Hz^{-1} (see Section 3.2.2) ensures that the angular size completeness corrections are dominated by extended radio sources. Nonetheless, we cannot distinguish with the angular size distribution from Windhorst et al. the distributions of FRIs and FRIIs. Fortunately, considering the comparison of FRI and FRII physical size distributions by Best (2009), it is reasonable to infer that both morphologies have comparable size distributions. Best (2009) used data from FIRST and NVSS to show that above 40 kpc, the size distributions of FRI and FRII radio galaxies are similar, whereas the smallest source in our sample is 83 kpc and 95% of our sources are larger than 230 kpc. However, the measured angular sizes from FRIs are likely to be more underestimated at higher redshifts, as sources become dimmer as a function of redshift, which affects the detectibility of the full extend of diffuse FRI jets. As a result, the angular size completeness corrections for FRIs might become less accurate for higher redshifts, compared to FRIIs whose sizes are defined by their brightest components. However, to simplify the construction of the RLFs, we rely only on the measured angular sizes to determine the angular size completeness corrections and, therefore, we do not derive a redshift dependency. The effect of this issue on our RLFs is discussed in Section 3.6.2.

3.4.2.3. Final completeness correction

We present our final full completeness corrections (including all corrections discussed above) as a function of flux density in Figure 3.7. We identified a 50% completeness around ~100 mJy. This is also the flux density after which the completeness corrections start to be dominated by the angular size completeness corrections. This becomes evident in the convergence of the completeness for FRI and FRII sources at the upper end of the flux densities, stemming from the assumption of comparable size distribution for both FRIs and FRIIs. The flattening at the end of the completeness curve shows the effects of the 400″ upper angular size threshold that

becomes more dominant at higher flux densities. After 10 Jy, the completeness corrections start to fall off, which matches with the assumption that a notable fraction of bright GRGs are missing due to selection effects (see Section 3.2.2). Below ~ 100 mJy we notice how the better surface brightness sensitivity in LoTSS-Deep Fields DR1, compared to LoTSS DR1, improves the completeness from M22. For M19, we also notice how FRII sources at lower flux densities are more complete compared to FRI sources. This is because the FRI jets tend to be diffuser and more likely to disappear below the detection limit, while an FRII with the same flux density will typically remain classifiable as long as their prominent hotspots remain visible. Because the completeness corrections become extremely large at the lower radio flux density end, we set, similarly to the work by Kondapally et al. (2022), a lower boundary for the completeness corrections by a factor of 10. This threshold ensures that the corrections do not become unreliable large for sources at the lower flux density end of our catalogues.

In Appendix 3.A, we discuss a few tests to validate the RLF construction method and test the reliability of our $V_{\rm max}$ derivations and completeness corrections with a local sample. These tests show that both the $V_{\rm max}$ and the completeness corrections are applied well and we recover a good estimate of the space densities of FRIs and FRIIs.

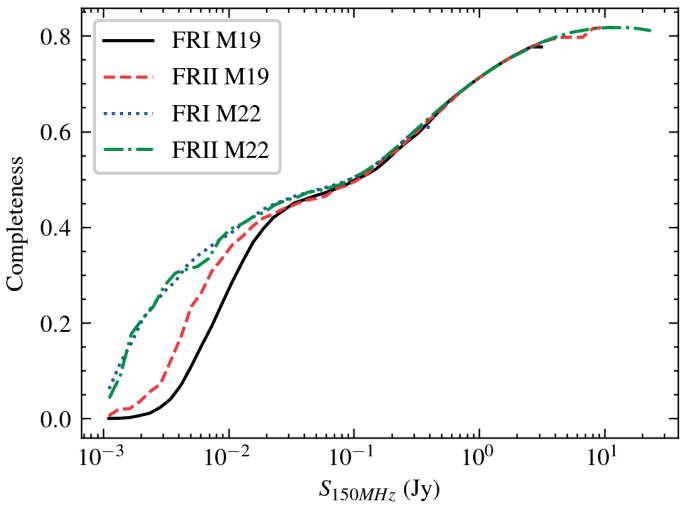


Figure 3.7: Completeness corrections for FRI and FRII sources from M19 and M22.

3.5. Results

Gathering all the ingredients from Sections 3.3 and 3.4, we constructed RLFs to study the FRI and FRII space density evolution between $10^{24.5} \lesssim L_{150} \lesssim 10^{28.5}$ W Hz^{-1} out to $z = 2.5$. In this section, we present our main results, where we look separately at the local RLF and the RLF up to $z = 2.5$.

3.5.1. Local radio luminosity function

The local RLF constructed with the M19 and M22 sources below $z = 0.3$ is shown in Figure 3.8. In this same figure, we also plot the FRI and FRII RLFs from Gendre et al. (2013) and the total 'radio-excess AGN' RLF from Kondapally et al. (2022, hereafter referred to as the total RLAGN RLF). We observe how all our local FRI and FRII space densities are near or below the total RLAGN space densities, which is constructed on the basis of all RLAGN from the LoTSS-Deep Fields DR1 (Kondapally et al., 2021; Duncan et al., 2021; Tasse et al., 2021; Sabater et al., 2021). This is an important requirement, as the FRIs and FRIIs are a subset of the total RLAGN population. The FRI and FRII RLFs from Gendre et al. also agree well with our RLFs, which is another indication that our completeness corrections are working well (see also the tests in Appendix 3.A). The FRIs dominate below the break luminosity ($L_{150} \sim 10^{26}$ W Hz^{-1}), while FRIIs dominate above. This is consistent with prior research (Fanaroff & Riley, 1974; Parma et al., 1996; Ledlow & Owen, 1996; Gendre et al., 2010). Similarly to Gendre et al. (2010, 2013), we find a smooth transition from the FRI to FRII dominance around the original break luminosity, where the FRI RLF drops off more strongly towards the higher radio luminosities compared to the FRII RLF. This smooth transition is also consistent with the FRI and FRII populations observed by FIRST (Capetti et al., 2017a,b).

We can quantify the steepness of the space density drop-off after the break luminosity by fitting a broken power law. To constrain better the power-law, we include the FRI and FRII data points from Gendre et al. for the radio luminosities where our RLFs do not cover (see Figure 3.8). This broken power-law is given by:

$$\rho(L_\nu) = \rho_0 \left(\left(\frac{L_\nu}{\tilde{L}} \right)^{\alpha} + \left(\frac{L_\nu}{\tilde{L}} \right)^{\beta} \right)^{-1}, \tag{3.8}$$

where α and β are respectively the low and high luminosity exponents, ρ_0 is the characteristic space density, and $\tilde{L}$ is the break luminosity. In Table 3.1, we give our best-fit parameters for both the local FRI and FRII fit. We see from these values how the local FRI space densities are declining more strongly towards higher radio luminosities compared to the FRIIs. We also find different values for $\tilde{L}$,

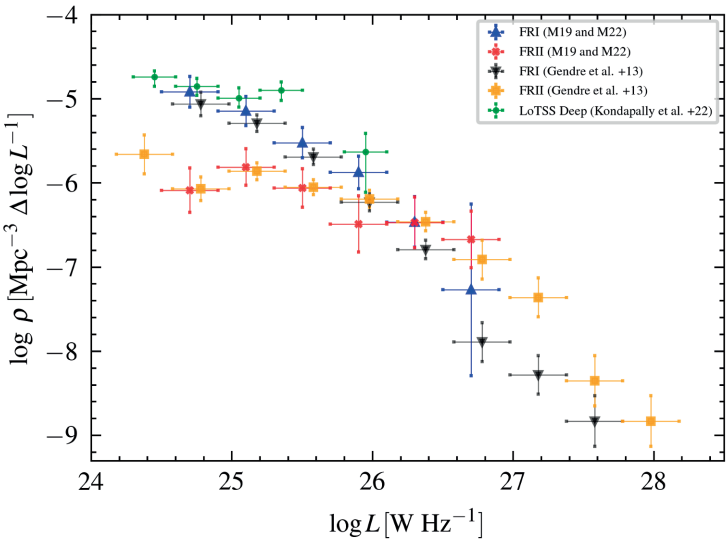


Figure 3.8: Local RLF for FRIs and FRIIs from this paper are plotted in blue and red, respectively. For comparison and to validate that our FRI and FRII space densities do not extend above the total RLAGN population, we added the total local RLAGN RLF from Kondapally et al. (2022) beyond $L_{150} \sim 10^{24}$ W Hz^{-1} (in green). This total RLAGN RLF is constructed based on sources from the LoTSS-Deep Fields DR1. To demonstrate that our results agree with previous work, we also added the FRI and FRII RLF from Gendre et al. (2013) in black and orange, respectively. These points are based on the CoNFIG catalogue (Gendre & Wall, 2008; Gendre et al., 2010), where we converted the luminosity bins to 150 MHz by using a spectral index of $\alpha = 0.7$. The 1σ bars from our RLFs include $z_{\max}$ errors, Poisson errors, completeness corrections errors, and classification errors. All our data points are given in Table 3.2.

which is usually taken to be equal to the break luminosity and determined by visual inspection, as done by Gendre et al. (2010). Despite this different choice, the values for $\tilde{L}$ are still consistent with the break luminosity around $\sim L_{150} \sim 10^{26}$ W Hz^{-1} we find in the literature (e.g. Fanaroff & Riley, 1974; Jackson & Wall, 1999; Willott et al., 2001; Kaiser & Best, 2007; Mingo et al., 2019).

3.5.2. Radio luminosity function up to $z = 2.5$

In Figure 3.9, we see both the FRI and FRII space density as well as the total RLAGN from Kondapally et al. (2022). Similarly to the local RLF in Figure 3.8, we find in Figure 3.9 that the space densities from the FRI and FRII samples are below the total RLAGN RLF. We see for the redshift bins below $z = 0.8$ a break luminosity, where FRIs dominate below and FRIIs above $L_{150} \sim 10^{26}$ W Hz^{-1}.

	ρ_0	α	β	$\tilde{L}$
FRI	-5.5 $\mathrm{Mpc}^{-3}\Delta \log L^{-1}$	0.59	2.0	25.8 W Hz^{-1}
FRII	-6.15 $\mathrm{Mpc}^{-3}\Delta \log L^{-1}$	0.13	1.7	26.4 W Hz^{-1}

Table 3.1: Best fit parameters for the local FRI and FRII fit from Equation 3.8.

Beyond $z = 0.8$ the break luminosity becomes less well-defined. In addition to Figure 3.9, in Figure 3.10 (right panel) we compare the space density evolutions per morphology class for three redshift bins, where the local RLF is fitted with a broken power-law as explained in the previous subsection. By subtracting the space densities from the FRIs and FRIIs in log space and propagating their error bars, we get the FRI/FRII ratio evolution in the left panel of this same figure. We combined all sources between $z = 0.8$ and $z = 2.5$ in one redshift bin because we have only 112 out of the total 1560 sources in this redshift range.

In Figure 3.10, we see the FRII space densities decrease at a lower rate as a function of radio luminosity when we make a comparison the highest with the lowest redshift bin. Above $L_{150} \sim 10^{27}$ W Hz^{-1}, this results in a clear space density enhancement across redshift by $\sim 0.6 - 1$ dex between the local and $z > 0.8$ RLFs. Also Gendre et al. (2010) found for FRIIs a similar space density enhancement above $L_{150} \sim 10^{27}$ W Hz^{-1}, where we converted their luminosities from 1.4 GHz to 150 MHz with a spectral index of $\alpha = 0.7$. They also report space density enhancements for FRIs at those radio luminosities, which also agrees with the mild enhancements that we detect.

In the right panel of Figure 3.10, we detect a space density decline for FRIs from the local ($z < 0.3$) to the highest redshift bin ($0.8 < z < 2.5$) below $L_{150} \sim 10^{26}$ W Hz^{-1}. For FRIIs, we do not find a significant space density decline below this radio luminosity. The space density decline for FRIs is not reported by Gendre et al. (2010). Nonetheless, both our results are still in agreement with each other within the error margin, as Gendre et al. only relied on 7 FRIs with associated redshifts above $z = 0.3$, making their uncertainties large. Despite the FRI space density decline beyond $z = 0.8$, we find the FRI/FRII space density ratios as a function of radio luminosity to remain fairly constant within error bars in the left panel of Figure 3.10.

The different space density evolutions as a function of redshift are also shown in Figure 3.11, where we inverted the RLFs to space density evolutions as a function of redshift for radio luminosity bins. Between $10^{27} \lesssim L_{150} \lesssim 10^{28}$ W Hz^{-1} we find a (mild) space density enhancement for both FRIs and FRIIs, which finds its maximum around $z = 1.6$. The space densities for FRIs and FRIIs do not show any notable redshift evolution between $10^{26} \lesssim L_{150} \lesssim 10^{27}$ W Hz^{-1}. For FRIs, we find a

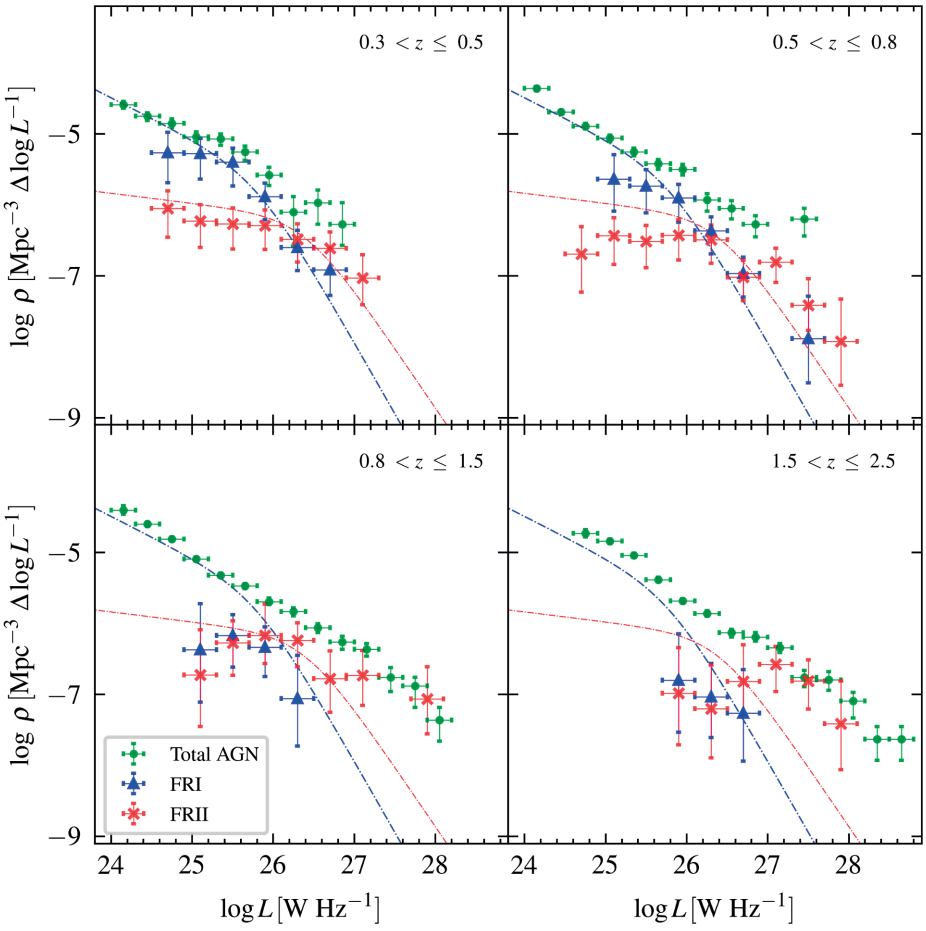


Figure 3.9: RLF for FRI (blue) and FRII (red) sources split in different redshift bins. For comparison and to show that our FRI and FRII space densities do not extend above the total RLAGN population, we added the total RLAGN RLF from Kondapally et al. (2022) beyond $L_{150} \sim 10^{24}$ W Hz^{-1} in green. We also added the local RLFs for both FRIs and FRIIs with dot-dashed lines. We removed radio luminosity bins which only contain a single source. The 1σ bars from our RLFs include $z_{\rm max}$ errors, Poisson errors, completeness corrections errors, and classification errors, as explained in Sections 3.4.1.2 and 3.4.2. Data points are given in Table 3.2.

mild space density decline above $z \sim 0.6$ and between $10^{25} \lesssim L_{150} \lesssim 10^{26}$ W Hz^{-1}, while FRIIs do show for these radio luminosities hints of a mild declining trend above $z \sim 1$. However, this declining FRII space density trend is within error bars less significant than the declining trend of FRIs.

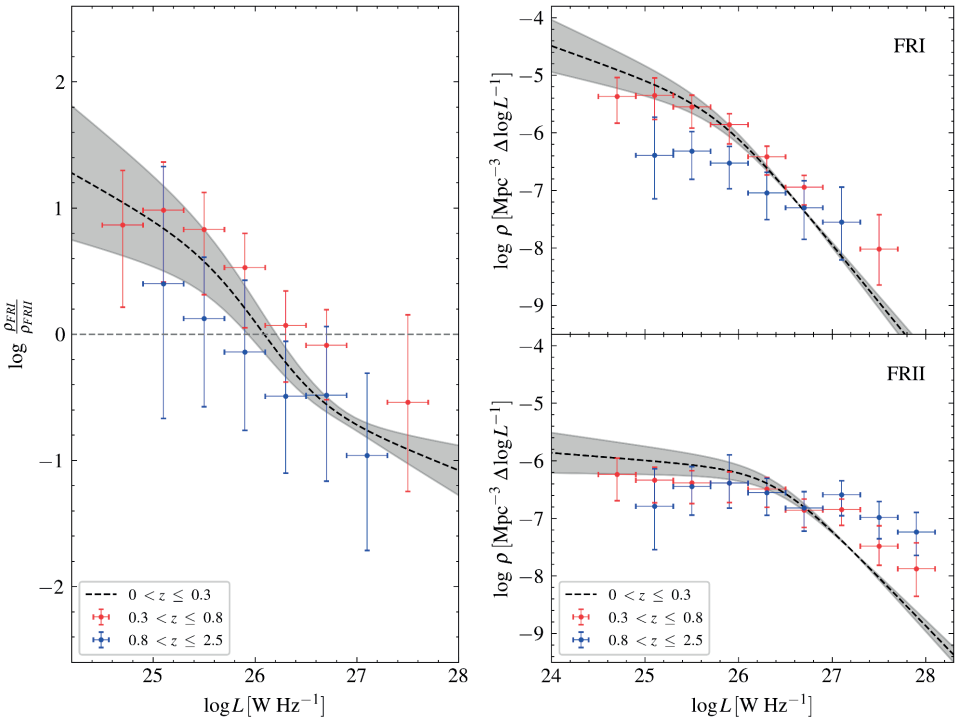


Figure 3.10: RLF ratio evolution. *Left:* Ratio evolution over different redshift bins. *Right:* Separated FRI (above) and FRII (below) space density evolution plots corresponding to the ratio evolution plot in the left panel. In both figures, we added the Gendre et al. (2013) data points for the higher end of the luminosity range, such that we can further constrain the local RLF, which was fitted with a broken power law from Equation 3.8 and corresponding values from Table 3.1, where we used a 1σ confidence interval. We removed radio luminosity bins which only contain a single source. The 1σ bars from our RLFs include $z_{\max}$ errors, Poisson errors, completeness corrections errors, and classification errors, as explained in Sections 3.4.1.2 and 3.4.2.

3.6. Discussion

With the arrival of larger FRI/FRII catalogues sensitive to lower flux densities and better redshift measurements, we have in the previous sections been able to construct RLFs up to $z = 2.5$ based on the M19 and M22 catalogues. We will in this section discuss the effects of energy losses on our RLFs, how the total RLAGN RLF compares to our FRI and FRII RLFs, along with the remaining selection biases, and how we can interpret the space density evolution that we find for FRIs and FRIIs.

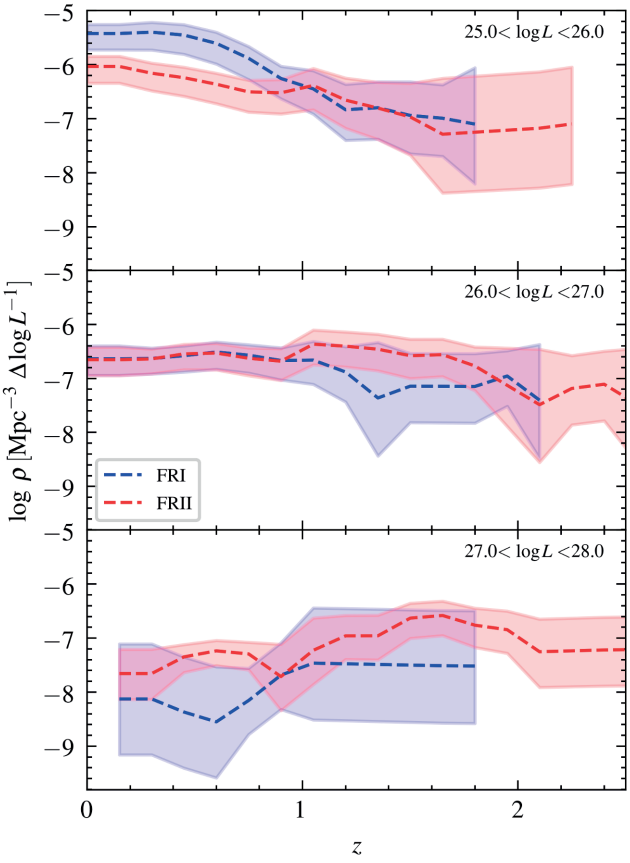


Figure 3.11: Space density evolution as a function of redshift (z) for 1 dex luminosity bins. This is constructed with sources from M19 and M22. The 1σ bars from our RLFs include $z_{\rm max}$ errors, Poisson errors, completeness corrections errors, and classification errors, as explained in Sections 3.4.1.2 and 3.4.2.

3.6.1. Energy losses and the RLF

At higher redshifts, IC losses for each source become more important, as more electrons will be scattered by CMB photons when the energy density of the CMB increases. These IC losses are suggested to be one of the primary mechanisms behind the $\alpha - z$ relation (Klamer et al., 2006; Ghisellini et al., 2014; Morabito & Harwood, 2018; Sweijen et al., 2022a), which associates steeper redshifts with higher spectral indices (Tielens et al., 1979; De Breuck et al., 2000; Miley & De Breuck, 2008). Because we are interested in the redshift evolution of FRIs and FRIIs, it is important to discuss how these energy losses affect the interpretation of our RLFs.

The impact of the IC losses on the measured surface brightness of a source is mainly contingent upon the magnetic field strength within the lobes and the equivalent magnetic field associated with the CMB (Harwood et al., 2013). Due to their different morphologies, we would anticipate the IC losses to affect FRIs and FRIIs disproportionately, as this effect can alter both the size and detectability of FRIs, whereas for FRIIs the detectability will likely not change because compact hot spots at the ends of the jets are relatively unaffected. As a result, we expect that when we would take IC losses into account, the $z_{\rm max}$ and the completeness corrections will change more significantly for FRIs than for FRIIs. Similarly, we could also consider other losses, such as the effects of synchrotron losses on the measured radio flux densities of FRIs and FRIIs (Myers & Spangler, 1985; Alexander & Leahy, 1987; Jamrozy et al., 2008; Harwood et al., 2013; Sweijen et al., 2022a). In addition, a more distant universe is also denser, which leads to greater confinement of the radio lobes and lower adiabatic losses of sources at higher redshifts compared to sources in a local universe (e.g. Barthel & Arnaud, 1996). Nevertheless, incorporating corrections for all the above-mentioned energy losses would violate our initial assumption that the radio luminosity from a source remains constant as a function of redshift while calculating $V_{\rm max}$ (see Section 3.3.2). Therefore, for the interpretation of the comparison of FRIs and FRIIs in our RLFs it is important to be aware that these are measured radio luminosity functions and not energy loss-corrected radio luminosity functions.

Fortunately, the FRI/FRII ratio RLF is less affected by those losses, as part of their effect on the $z_{\rm max}$ and completeness corrections are cancelled out. As a result, this makes the physical interpretation more comparable to the hypothetical energy-loss-adjusted RLFs. The left panel of Figure 3.10 supports this comparability, as the three redshift bins are the same within the uncertainties.

3.6.2. Comparing FRI/FRII RLFs

It is well known that FRIs and FRIIs make up a significant fraction of the total RLAGN population (Sadler, 2016; Baldi et al., 2018). In this subsection, we will discuss the comparison of our FRI and FRII RLFs with those from Gendre et al. (2010, 2013) and Kondapally et al. (2022). For the total RLF we exclusively use the total RLAGN RLF from Kondapally et al., as they demonstrated consistency with previous total RLAGN RLFs in the literature.

Figure 3.8 shows how local the FRI space densities closely follow the total RLAGN space densities up to about $L_{150} \sim 10^{26}$ W Hz^{-1}, after which FRIIs become the dominant morphology population. This similarity between the FRI and total RLAGN space densities is due to the fact that the fraction of compact sources increases towards the lower radio luminosities and we only consider sources above

$L_{150} \sim 10^{24.5}$ W Hz^{-1} (e.g. Baldi et al., 2015, 2018; Capetti et al., 2020). It is reassuring to find the local RLF from Gendre et al. (2013) to be similar to ours, as this indicates that the local RLFs constructed in this paper and by Gendre et al. are not biased by the data or RLF construction methods. In Figure 3.9, we observe that the space densities of the FRI and total RLAGN space densities continue to show consistent similarity up to a similar break luminosity and $z = 0.8$. Examination of the FRI evolution above $z = 0.3$ could not be accurately done by Gendre et al. (2010), due to the fact they had in their sample only seven FRIs associated with redshifts above $z = 0.3$ (see also Figure 6 in Gendre et al., 2010). This is over 100 times fewer FRIs compared to what we have available from the combined M19 and M22 catalogues, comprising 877 FRIs above $z = 0.3$, with accurate redshifts (Duncan et al., 2019, 2021).

Beyond $z = 0.8$, we do not find a well-defined break luminosity above which FRIIs dominate over FRIs. However, we do find in Figure 3.9 above $L_{150} \sim 10^{26}$ W Hz^{-1} the FRIIs to be a dominant morphology type when we compare the total RLAGN space densities with the FRII space densities. This demonstrates that up to $z = 2.5$, the FRIIs dominate the RLAGN population at the high radio luminosity end. Below $L_{150} \sim 10^{26}$ W Hz^{-1} and beyond $z = 0.8$ we detect a prominent offset in the order of 1 to 1.5 dex between the total RLAGN and the FRI and FRII space densities. This is likely related to remaining unaccounted selection effects. Some of this might be a result of energy loss effects that we did not take into account, which more strongly affect FRI detections (see Section 3.6.1). Also, the probable underestimation of the completeness corrections for FRIs due to the difficulty of accurately measuring their sizes in the higher redshift bins could play a role (see Section 3.4.2.2). Moreover, Kondapally et al. (2022) discusses how the space densities of star-forming low-excitation radio galaxies (LERGs) increase over redshift and dominate the LERG space densities after $z \sim 1$, while the space densities of quiescent LERGs decreases over redshift. Given that low-power FRIs and FRIIs are predominantly associated with LERGs (Mingo et al., 2022), there is a possible selection bias in our completeness corrections that is unaccounted for if some of those star-forming LERGs might indeed turn out to be FRIs or FRIIs with small jets that are difficult to detect at a resolution of $6''$. We also argue that the difficulty of detecting star-forming radio galaxies with radio jets in our sample is likely a more significant issue for FRIs, as their smaller weaker jets are more difficult to detect compared to FRIIs. Other missing morphology types such as hybrid or double-double sources are expected to be rare (Harwood et al., 2020; Mingo et al., 2019, 2022) and therefore will not have a significant contribution to remaining selection effects. Thus, although the redshift simulations help us to correct for observational biases when deriving $z_{\max}$ values (see Section 3.4.1.1) and

completeness corrections (see Section 3.4.2.1), it is possible that residual resolution and surface brightness limits could still reside in selection effects that we are not yet able to fully correct for.

3.6.3. Space density enhancements

Space density enhancements of the total RLAGN RLF have been well-measured for a long time (e.g. Dunlop & Peacock, 1990). Multiple studies have found space density enhancements for bright FRIs and FRIIs (Snellen & Best, 2001; Willott et al., 2001; Jamrozy, 2004; Rigby et al., 2008; Gendre et al., 2010). Given our increased number of sources above $z = 0.3$ compared to previous studies, it is valuable to re-examine the space density enhancements and declines with our RLFs and look specifically at how these relate to our understanding of radio galaxy evolution.

The space density enhancement from low to high redshift that we find for FRIs and FRIIs above $L_{150} \sim 10^{27}$ W Hz^{-1} could simply be related to the higher gas availability in the earlier universe, increasing the probability for RLAGN to become more powerful. Although taking into account energy loss effects in our RLFs is complex and beyond the scope of this paper (see Section 3.6.1), it is worth mentioning that if the net energy loss effects are lower in the earlier denser universe (e.g. significant lower adiabatic losses) compared to the local universe, we could expect to measure even more powerful sources at higher redshifts. We are not well-enough constrained for bright FRIs to find a notable difference between the space density enhancements of FRIs and FRIIs above $L_{150} \sim 10^{27}$ W Hz^{-1} (see Figures 3.10 and 3.11). Nevertheless, a recent simulation demonstrated how radio sources were more likely to live in less rich environments at $z = 2$ compared to $z = 0$ (Thomas & Davé, 2022), which is in favour of FRII morphologies (Croston et al., 2019). More data at the high radio luminosity end would help to better constrain the FRI and FRII space density enhancements and test if powerful FRIIs have indeed a stronger space density enhancement than powerful FRIs.

The space density decline of FRIs below $L_{150} \sim 10^{26}$ W Hz^{-1} is likely related to unaccounted selection effects, as we discussed in Section 3.6.2 to explain the observed space density offsets between the local and higher redshift bins. However, despite the fact that the remaining selection effects have a stronger effect on the space densities from FRIs than FRIIs, we find the FRI/FRII space density ratio overall radio luminosities to be remarkably similar when we compare the three redshift bins in the left panel of Figure 3.10. If this holds for larger and more comprehensive sample sizes, resulting in reduced Poisson errors, it could suggest that the jet-disruption of FRIs is not primarily influenced by environmental factors, as those evolve over redshift. Instead, disruption of the jets from the FRI parent population would more likely be associated with events happening in close proximity

to, or within, the parent radio galaxy.

3.6.4. Future prospects

We demonstrate in this paper how detection limits affect selection biases and how it is possible to partly correct for these if we want to construct reliable RLFs from RLAGN. In order to enhance the significance of our findings, we need more objects from more sensitive radio maps along with wider area observations. The $40''$ angular size cut that we needed to apply, due to the angular size selection from Mingo et al. (2019), removes a large fraction of the small FRIs and FRIIs. Especially sources at the lowest radio flux densities are the most affected. The angular size cut is due to the $6''$ resolution from the LoTSS radio maps that were used to extract the FRI and FRII sources from. So, to collect a reliable sample below $40''$, we need to improve the resolution.

Recent efforts have been ongoing to increase the sky coverage with LoTSS DR2 at $6''$ (Shimwell et al., 2022), which will increase the number of objects in M19 by a factor of about 13. This data release is based on radio maps at the same resolution and a similar sensitivity as LoTSS DR1. This will therefore only help to reduce the uncertainties at the higher radio luminosity end and up to $z = 0.8$. Fortunately, work has also been done to bring LOFAR's resolution down to $0.3''$ by including all baselines up to ~2000 km (Varenius et al., 2015, 2016; Harris et al., 2019; Morabito et al., 2022a; Sweijen et al., 2022c). This resolution will help us to collect sources that are unresolved at $6''$, such that we can lower the $40''$ angular size cut. An improved resolution comes with the unavoidable loss of surface brightness sensitivity. Sweijen et al. (2022c) reported from $6''$to $0.3''$an estimated 60% detection loss of sources that were unresolved at $6''$. As FRIs are on average less bright and have by definition diffuser jets than FRIIs, this will surely have a more negative effect on the detections of FRIs compared to FRIIs. To ensure that enough small FRIs will be detected, it becomes essential to improve the sensitivity by processing and calibrating multiple observations of the same fields (de Jong et al. in prep.) and to complement these high-resolution radio maps with intermediate resolutions around for example ~$1''$(Ye et al., 2024).

3.7. Conclusions

We presented in this paper RLFs of FRI and FRII morphologies up to $z = 2.5$ and beyond $L_{150} \sim 10^{24.5}$ W Hz^{-1}, by utilizing the M19 and M22 catalogues. We corrected for redshift effects and the incompleteness of our sample by using the `redshifting` algorithm. This algorithm gave us a reliable estimate of the maximum distance an FRI and FRII source can still be classified. In particular, our RLFs for

FRIs are an improvement above $z = 0.3$, as we have over 100 times more available FRIs with associated redshifts above this redshift, compared to previous studies (Gendre et al., 2010, 2013). These RLFs also provide us with continuing evidence of the evolution of FRI and FRII sources.

In our RLFs, we do not detect any sharp transitions between the FRI and FRII morphologies as a function of radio luminosity or redshift. We find a space density enhancement from low to high redshift for FRI and FRII sources beyond $L_{150} \sim 10^{27}$ W Hz^{-1}, which might be explained by a higher gas availability in the earlier universe, increasing the likelihood for FRIs and FRIIs to become powerful. Also, the net energy losses at higher redshifts could potentially increase the measured radio luminosities at higher redshifts if the adiabatic losses are significantly lower compared to lower redshifts. However, to test this assertion we need a more detailed analysis of the different energy loss effects (e.g. IC and synchrotron losses), which is beyond the scope of this paper. At the low radio luminosity end, we tentatively identify a declining trend of the FRI space densities with redshift. This is likely related to remaining selection effects such as from the underestimation of distant FRI angular sizes at higher redshifts and the difficulty in detecting star-forming FRIs with small jets, which have been proven to be more prominent at higher redshifts. Although there are significant uncertainties represented by the large error bars, the evolution of the FRI/FRII ratio that we derive suggests that the FRI morphology is primarily a result of the disruption of jets on scales originating within or close to the host galaxy, rather than jet-disruption due to large-scale environmental factors.

The potential residual selection biases in our results highlight the necessity to further develop the calibration and imaging pipeline of LOFAR data with baselines up to 2000 km, such that it will be possible to incorporate radio sources at smaller angular scales and lower radio luminosities. This will eventually help to more precisely model the FRI and FRII sources above $z = 0.3$ towards lower radio luminosities, which can further enhance our understanding of radio jet evolution from RLAGN and the link between these jets and their environment and host galaxies.

3.A. Testing the RLF reliability

Relocating sources to higher redshifts reduces their flux densities and apparent sizes, which makes it more difficult to detect and classify FRI and FRII sources. Figure 3.5 shows how FRIs and FRIIs are affected differently by these effects. To test whether the $V_{\rm max}$ method and our completeness corrections are taking the effect, due to changes in flux density and resolution, correctly into account, we compare an FRI and FRII sample below $z < 0.4$ with the same sample shifted to $z = 1.5$. This is done by using the algorithm from Section 3.3.2 and applying the same completeness

correction process discussed in Section 3.4.2. If our $V_{\rm max}$ method and completeness corrections are well-applied, we expect the derived space densities of the original $z < 0.4$ sample and of the same sample shifted to $1.5 < z < 2.5$ to agree with each other, as the simulated sample has larger space density corrections and smaller $V_{\rm max}$ values. The left panel from Figure 3.12 has a sample of 478 FRIs before relocating them to $z = 1.5$. After relocating there are only 105 sources that could be classified. Although the number reduces with a factor 4, we find that the radio luminosity functions to still agree with each other. In the right panel of this same figure we also compare the RLF ratios before and after relocating and also find a strong agreement. We have also experimented with other redshift bins (smaller and larger) and find similar agreements. This gives us additional confidence that both the $V_{\rm max}$ and the completeness corrections are applied well.

To test the completeness corrections derived with the angular size distribution (see Section 3.4.2), we applied a larger angular size cut at 80″ and compared the space densities with the already present 40″ cut. The comparison for sources below $z = 0.5$ is shown in Figure 3.13 and shows how (even though the number count is reduced by nearly a factor of 2) the completeness corrections correct remarkably well for this effect and make both RLFs agree with each other. We also compared this for other angular size cuts and found similar consistencies. This demonstrates the reliability of correcting our space densities with the angular size distributions from Windhorst et al. (1990) with updated fitting parameters from Mandal et al. (2021) on a local sample.

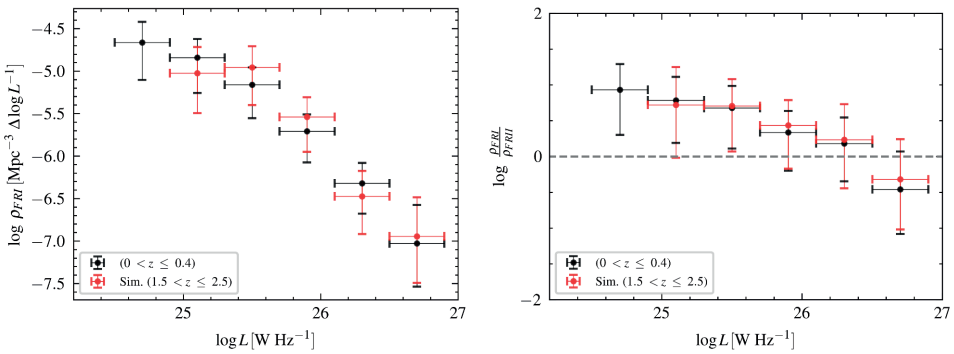


Figure 3.12: RLF space density evolution difference between the original and simulated sample. *Left:* RLF for FRI sample before and after applying the `redshifting` algorithm from $z < 0.4$ to $z = 1.5$. *Right:* RLF ratio for a sample of FRI and FRII sources before and after applying the `redshifting` algorithm from $z < 0.4$ to $z = 1.5$. Error bars from our RLFs include $z_{\rm max}$ errors, Poisson errors, completeness corrections errors, and classification errors.

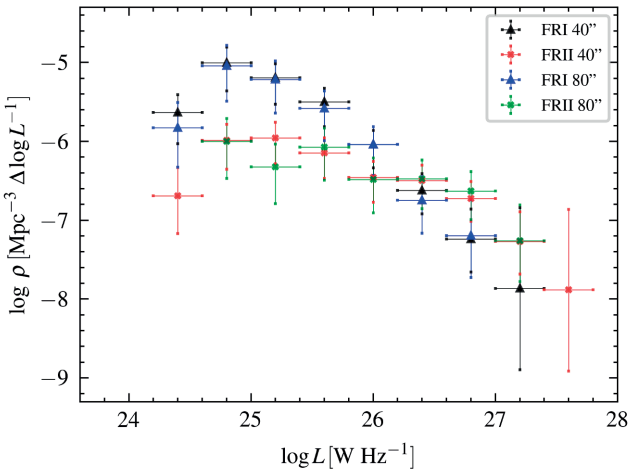


Figure 3.13: Comparison between a 40″ and 80″ source cut below $z = 0.5$. Both RLFs are corrected with the completeness corrections from Section 3.4.2. The angular size completeness correction is applied for respectively the 40″ and 80″ angular sizes. This figure contains all sources in the redshift bin $0 < z < 1.0$. The 40″cut selection has 865 sources, while the 80″ cut has 405 sources left. The 1σ error bars contain $z_{\max}$ errors, Poisson errors, completeness corrections errors, and classification errors.

$\log L$ (W Hz^{-1})	$0 < z < 0.3$		$0.3 < z < 0.5$		$0.5 < z < 0.8$		$0.8 < z < 1.5$		$1.5 < z < 2.5$	
	FRI	**FRII**	**FRI**	**FRII**	**FRI**	**FRII**	**FRI**	**FRII**	**FRI**	**FRII**
24.7	$-4.92^{+0.18}_{-0.33}$ (114)	$-6.09^{+0.26}_{-0.39}$ (9)	$-5.26^{+0.24}_{-0.42}$ (94)	$-6.05^{+0.24}_{-0.41}$ (21)	$-6.32^{+0.32}_{-0.5}$ (13)	$-6.69^{+0.38}_{-0.54}$ (7)	-	-	-	-
25.1	$-5.15^{+0.17}_{-0.32}$ (86)	$-5.81^{+0.22}_{-0.34}$ (16)	$-5.28^{+0.2}_{-0.36}$ (146)	$-6.23^{+0.23}_{-0.37}$ (18)	$-5.64^{+0.26}_{-0.45}$ (84)	$-6.43^{+0.25}_{-0.41}$ (24)	$-6.37^{+0.64}_{-0.74}$ (2)	$-6.72^{+0.64}_{-0.73}$ (2)	-	-
25.5	$-5.52^{+0.18}_{-0.3}$ (40)	$-6.06^{+0.23}_{-0.33}$ (12)	$-5.4^{+0.19}_{-0.34}$ (130)	$-6.27^{+0.22}_{-0.36}$ (18)	$-5.73^{+0.21}_{-0.38}$ (133)	$-6.51^{+0.22}_{-0.37}$ (26)	$-6.17^{+0.29}_{-0.44}$ (10)	$-6.27^{+0.31}_{-0.46}$ (8)	-	-
25.9	$-5.87^{+0.19}_{-0.29}$ (21)	$-6.49^{+0.33}_{-0.4}$ (5)	$-5.88^{+0.19}_{-0.33}$ (46)	$-6.29^{+0.22}_{-0.34}$ (19)	$-5.9^{+0.19}_{-0.34}$ (119)	$-6.43^{+0.21}_{-0.35}$ (35)	$-6.33^{+0.28}_{-0.41}$ (9)	$-6.17^{+0.25}_{-0.4}$ (13)	$-6.8^{+0.64}_{-0.74}$ (2)	$-6.98^{+0.64}_{-0.73}$ (2)
26.3	$-6.46^{+0.3}_{-0.36}$ (6)	$-6.47^{+0.3}_{-0.35}$ (6)	$-6.6^{+0.24}_{-0.33}$ (11)	$-6.49^{+0.22}_{-0.32}$ (14)	$-6.36^{+0.19}_{-0.33}$ (46)	$-6.49^{+0.2}_{-0.33}$ (34)	$-7.06^{+0.61}_{-0.67}$ (2)	$-6.23^{+0.25}_{-0.37}$ (13)	$-7.04^{+0.47}_{-0.57}$ (3)	$-7.2^{+0.62}_{-0.69}$ (2)
26.7	$-7.27^{+1.02}_{-1.03}$ (1)	$-6.67^{+0.34}_{-0.38}$ (5)	$-6.91^{+0.3}_{-0.36}$ (6)	$-6.61^{+0.23}_{-0.3}$ (12)	$-6.97^{+0.23}_{-0.33}$ (13)	$-7.02^{+0.23}_{-0.33}$ (12)	$-7.37^{+1.03}_{-1.06}$ (1)	$-6.78^{+0.39}_{-0.48}$ (4)	$-7.26^{+0.61}_{-0.68}$ (2)	$-6.82^{+0.36}_{-0.47}$ (5)
27.1	-	-	$-7.73^{+1.02}_{-1.04}$ (1)	$-7.03^{+0.33}_{-0.38}$ (5)	-	$-6.8^{+0.19}_{-0.29}$ (22)	$-7.44^{+1.02}_{-1.05}$ (1)	$-6.73^{+0.34}_{-0.43}$ (5)	$-7.63^{+1.03}_{-1.06}$ (1)	$-6.58^{+0.26}_{-0.38}$ (11)
27.5	-	-	-	$-7.75^{+1.02}_{-1.03}$ (1)	$-7.88^{+0.6}_{-0.63}$ (2)	$-7.42^{+0.3}_{-0.35}$ (6)	-	$-7.51^{+1.02}_{-1.04}$ (1)	-	$-6.81^{+0.3}_{-0.4}$ (7)
27.9	-	-	-	$-7.76^{+1.02}_{-1.03}$ (1)	-	$-7.93^{+0.6}_{-0.62}$ (2)	-	$-7.06^{+0.45}_{-0.49}$ (3)	-	$-7.41^{+0.6}_{-0.65}$ (2)
28.3	-	-	-	$-7.76^{+1.02}_{-1.03}$ (1)	-	-	-	-	-	$-7.78^{+1.02}_{-1.04}$ (1)

Table 3.2: RLF space density values in Mpc$^{-3}\Delta \log L^{-1}$ from Figures 3.8 and 3.9. The source count is given between brackets below each space density value.

Acknowledgements

This publication is part of the project CORTEX (NWA.1160.18.316) of the research programme NWA-ORC which is (partly) financed by the Dutch Research Council (NWO). This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-1287. RK and PNB are grateful for support from the UK STFC via grant ST/V000594/1. BM acknowledges support from the Science and Technology Facilities Council (STFC) under grants ST/T000295/1 and ST/X001164/1. RJvW acknowledges support from the ERC Starting Grant ClusterWeb 804208. This work was supported by the Medical Research Council [MR/T042842/1]. LOFAR data products were provided by the LOFAR Surveys Key Science project (LSKSP; `https://lofar-surveys.org/`*) and were derived from observations with the International LOFAR Telescope (ILT). LOFAR (van Haarlem et al., 2013) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and which are collectively operated by the ILT foundation under a joint scientific policy. The efforts of the LSKSP have benefited from funding from the European Research Council, NOVA, NWO, CNRS-INSU, the SURF Co-operative, the UK Science and Technology Funding Council and the Jülich Supercomputing Centre. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version.*

4

Into the depths: Unveiling ELAIS-N1 with LOFAR's deepest sub-arcsecond wide-field images

J.M.G.H.J. de Jong, R.J. van Weeren, F. Sweijen, J.B.R. Oonk, T.W. Shimwell, A.R. Offringa, L.K. Morabito, H.J.A. Röttgering, R. Kondapally, E.L. Escott, P.N. Best, M. Bondi, H. Ye

Astronomy & Astrophysics, Volume 689, id.A80, 32 pp.

Abstract

We present the deepest wide-field 115-166 MHz image at sub-arcsecond resolution spanning an area of 2.5°×2.5° centred at the ELAIS-N1 deep field. To achieve this, we improved the direction-independent (DI) and direction-dependent (DD) calibrations for the International LOw Frequency ARray (LOFAR) Telescope. This enhancement enabled us to efficiently process 32 hrs of data from four different 8-hrs observations using the high-band antennas (HBAs) of all 52 stations, covering baselines up to approximately 2,000 km across Europe. The DI calibration was improved by using an accurate sky model and refining the series of calibration steps on the in-field calibrator, while the DD calibration was improved by adopting a more automated approach for selecting the DD calibrators and inspecting the self-calibration on these sources. For our brightest calibrators, we also added an additional round of self-calibration for the Dutch core and remote stations in order to refine the solutions for shorter baselines. To complement our highest resolution at 0.3″, we also made intermediate resolution wide-field images at 0.6″ and 1.2″. Our resulting wide-field images achieve a central noise level of 14 μJy beam^{-1} at 0.3″, doubling the depth and uncovering four times more objects than the Lockman Hole deep field image at comparable resolution but with only 8 hrs of data. Compared to LOFAR imaging without the international stations, we note that due to the increased collecting area and the absence of confusion noise, we reached a point-source sensitivity comparable to a 500-hrs ELAIS-N1 6″ image with 16 times less observing time. Importantly, we have found that the computing costs for the same amount of data are almost halved (to about 139,000 CPU hrs per 8 hrs of data) compared to previous efforts, though they remain high. Our work underscores the value and feasibility of exploiting all Dutch and international LOFAR stations to make deep wide-field images at sub-arcsecond resolution.

4.1. Introduction

The International LOw Frequency ARray (LOFAR) Telescope (ILT) is an interferometer uniquely designed to measure low-frequency radio waves between 10 and 80 MHz with the low-band antennas (LBAs) and between 110 and 240 MHz with the high-band antennas (HBAs) (van Haarlem et al., 2013). With its baselines extending up to ~2,000 km coupled with a degree-scale field of view, it can thus produce wide-field images at sub-arcsecond resolution. Nevertheless, reducing data from all 38 Dutch and 14 international stations of LOFAR for wide-field imaging is nontrivial, as it requires a carefully optimised calibration strategy to correct for various corrupting effects on the data and extensive computing facilities to handle the substantial data volumes and perform the final imaging (Sweijen et al., 2022c; Ye et al., 2024).

In recent years, significant efforts have been devoted to automatically calibrating and imaging observations with the Dutch HBA stations located in the Netherlands. This has led to the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al., 2017, 2019, 2022; Williams et al., 2019) and the LoTSS-Deep Fields (Kondapally et al., 2021; Duncan et al., 2021; Tasse et al., 2021; Sabater et al., 2021; Best et al., 2023; Bondi et al., 2024), which have provided wide-field images of the northern sky at 144 MHz and 6″ resolution. Despite the fact that these works discovered many new radio sources at the lowest frequencies, approximately 90% of these sources at 6″ remain unresolved at 144 MHz. This becomes an issue when, for instance, one aims to examine the detailed dynamics of bright radio-loud active galactic nuclei (RLAGN) (e.g. Mahatma et al., 2023), derive source size distributions at the smallest angular scales (e.g. Sweijen et al., in prep.), separate radio emission from (radio-quiet) AGN and star formation (e.g. Morabito et al., 2022b), or when the 6″ resolution limit introduces selection effects in the study of the cosmic evolution of resolved RLAGN (e.g. de Jong et al., 2024). This, among other scientific objectives, underscores the scientific value of the sub-arcsecond wide-field imaging capabilities of LOFAR.

Calibrating data at low frequencies is challenging due to direction-dependent effects (DDEs), which are variations of data corruption across the field of view. At low frequencies, most DDEs are posed by the ionosphere, resulting in propagation delays of radio waves (Intema et al., 2009; Smirnov, 2011b; van Weeren et al., 2016b; Tasse et al., 2018). Correcting these factors inadequately results in image fidelity issues due to calibration artefacts that extend from arcsecond up to arcminute scales. These effects are destructive for the quality of the high-resolution images if not properly corrected. Incorporating data from all international LOFAR stations during calibration makes the data reduction more complicated, as extra phase delays are induced by the fact that the international stations have independent clocks (Mora-

bito et al., 2022a). Moreover, the fact that the availability of bright sources reduces towards higher resolutions complicates the calibration strategy, as this heavily relies on (self-)calibration of sources with a high S/N on all baselines. On top of this is the substantial volume of visibility data that needs to be processed. A typical LOFAR observation with an integration time of 8 hrs and a time and frequency resolution of 1 sec and 12.21 kHz is in the order of 16 TB, which can be reduced to 4 TB with Dysco compression (Offringa, 2016). To process LOFAR data, it is therefore essential to have access to machines with enough computational power and with extensive storage capacities.

Early pioneering works have demonstrated how to utilise all HBA antennas from LOFAR's international stations to produce images at sub-arcsecond resolutions (e.g. Varenius et al., 2015, 2016; Ramírez-Olivencia et al., 2018; Harris et al., 2019). Subsequent efforts by Morabito et al. (2022a) standardised and partly automated the calibration and imaging process with the international stations, which resulted in the first version of the LOFAR Very Long Baseline Interferometry (VLBI) calibration workflow.[1] The value of their work is directly evident through the large number of studies that have already utilised their workflow (e.g. Sweijen et al., 2022b, 2023; Bonnassieux et al., 2022; Timmerman et al., 2022b,a; Harwood et al., 2022; Kukreti et al., 2022; Morabito et al., 2022b; Mahatma et al., 2023; Venkattu et al., 2023). During the same time, Sweijen et al. (2022c) extended this strategy to perform wide-field imaging and produced with 8 hrs of LOFAR data from the Lockman Hole the first 6.6 deg^2 wide-field image at a resolution of $0.30'' \times 0.38''$ at 144 MHz with a sensitivity down to 25 μJy beam^{-1}. This image, produced with a computational cost of 250,000 CPU hrs, captured in one snapshot 2,483 high-resolution sources, each with peak intensities five times greater than their local RMS. These types of wide-field images contain approximately 10 billion pixels, which makes imaging the most dominant part of the total computational costs. Another recent study by Ye et al. (2024) adopted a similar calibration approach to Sweijen et al. (2022c) but with the aim to make an intermediate resolution wide-field image of the ELAIS-N1 deep field at $1.2'' \times 2''$. This resolution serves as a scientifically valuable intermediary that improves the $6''$ resolution from LoTSS and recovers extended emission that is lost at the finer $0.3''$ resolution, such as from low-excitation radio galaxies (LERGs) (Ye et al., 2024). Since imaging represents the main computational bottleneck for the complete data processing pipeline, Ye et al. (2024) achieved a total computing time speedup of nearly a factor five compared to Sweijen et al. (2022c).

Even though one of the primary objectives of achieving higher resolutions is to resolve more sources, high-resolution images are less suitable for studying low surface brightness structures, as these are more likely to be resolved out. For in-

[1] `https://github.com/LOFAR-VLBI`

stance, Sweijen et al. (2022c) showed that only 40% of the sources that are detected and unresolved at 6″ are detected in their corresponding 8-hrs radio map at 0.3″, of which 11% of these are resolved at 0.3″. The number of 6″ counterparts at 1.2″ doubles, as shown for ELAIS-N1 by Ye et al. (2024). Hence, in order to recover more resolved sources at higher resolutions, it is essential to enable the production of deeper images through the use of multiple 8-hrs observations of the same field and to get more information out of the data by making images at intermediate resolutions (between 0.3″ and 6″) as well. This approach is further supported by the fact that confusion noise limits the sensitivity obtainable by deep wide-field imaging using only the Dutch stations, as was demonstrated by Sabater et al. (2021) in their imaging of ELAIS-N1 with 163.7 hrs of LOFAR observations. They reached a best noise level of approximately ~17 μJy beam^{-1}, which is expected to be achievable with about ten times less LOFAR observing time when including both the Dutch and international stations.

We aim in this paper to produce the deepest wide-field images currently available at sub-arcsecond resolution at 140 MHz (115-166 MHz) by jointly calibrating four LOFAR observations, totalling 32 hrs, of the ELAIS-N1 deep field. Building upon the work from Morabito et al. (2022a), Sweijen et al. (2022c), and Ye et al. (2024), we refined the direction-independent (DI) calibration steps and improved the direction-dependent (DD) calibrator selection. This enabled us to obtain the final merged calibration solutions for Dutch and international LOFAR stations, which are required for facet-based imaging at (sub-)arcsecond resolutions. Utilising the calibrated data, we produced wide-field images at 0.3″, 0.6″, and 1.2″ resolution. This allowed us to make source catalogues at different resolutions and thereby analyse source detections across different resolutions and sensitivities.[2]

In Section 4.2, we discuss the details of our four LOFAR observations, which leads into a detailed discussion of the calibration process in Section 4.3. Following calibration, we address the imaging process in Section 6.4.5 and then detail the creation of the associated source catalogues in Section 4.5. Our discussion extends to evaluating the quality of our image and catalogue outputs in Section 4.6, and we finish with our conclusions in Section 4.7.

4.2. Data description

The area covered by ELAIS-N1 has been studied in the optical (e.g. McMahon et al., 2001; Aihara et al., 2018), infrared (e.g. Lawrence et al., 2007; Mauduit et al., 2012), ultraviolet (e.g. Martin et al., 2005), X-rays (e.g. Manners et al., 2003), and radio (e.g. Ciliegi et al., 1999; Sirothia et al., 2009; Croft et al., 2013; Ocran et al., 2020).

[2]All data products are published at `https://lofar-surveys.org/hd-en1.html`.

Observation ID	L686962	L769393	L798074	L816272
Project	LT10_012	LT10_012	LT14_003	LT14_003
Calibrator	3C 295 (L686958)	3C 295 (L769389)	3C 295 (L798082)	3C 295 (L816280)
Observation date	26-11-2018	24-05-2020	14-11-2020	13-5-2021
Pointing centres	16:11:00, +54.57.00	16:11:00, +54.57.00	16:11:00 +55.00.00	16:11:00 +55.00.00
Integration time	8 hrs	8 hrs	8 hrs	8 hrs
Frequency range	120-166 MHz	120-166 MHz	115-164 MHz	115-164 MHz
Stations (International)	51 (13)	51 (13)	50 (12)	52 (14)

Table 4.1: Metadata from the four ELAIS-N1 observations used in this paper.

This extensive multi-wavelength coverage has made ELAIS-N1 an invaluable field for extra-galactic science and it was therefore selected as one of the LOFAR Deep Fields (Sabater et al., 2021; Kondapally et al., 2021; Best et al., 2023).

In order to make the deepest sub-arcsecond resolution wide-field radio map of this field with LOFAR, we selected four 8-hrs LOFAR observations of ELAIS-N1 by examining calibration solutions of calibrator sources 3C 295 or 3C 48 that were already observed for 10 minutes before or after 16 different available ELAIS-N1 observations stored in the LOFAR long-term archive (LTA). This calibration step is an important part of the entire calibration (as discussed in Section 4.3.1) and provides a computationally cheap way to assess the ionospheric conditions and the quality of the data (as discussed in Section 4.3.1). Although we could select more than these 4 observations, it is important to stress that the compute costs for calibrating and imaging data at sub-arcsecond resolutions are expensive and limit us to selecting more than 4 observations (as highlighted by Sweijen et al. (2022c) and by us in Section 4.4.3). Our selected observations are part of two different observing projects (LT10_012 and LT14_003, PI: P.N. Best) and were retrieved from the LTA.[3]

We provide a description of our selected observations in Table 5.1. All four observations have 3C 295 as the primary calibrator. The pointing centres of two observations are 0.03° offset from the other two, which necessitates a phase-shift correction to a common right ascension (RA) of 16:11:00 and declination (DEC) of +54.57.00 to enable combined imaging (see Section 6.4.5). Prior to the storage of the observations L798074 and L816272 on the LTA, their data was averaged from a time resolution of 1 sec to 2 sec. The pre-averaging leads to additional time-smearing effects on the data. This was unfortunately only noticed after doing most of the calibration discussed in Section 4.3 and thus we kept the data in our final images. Whilst the time smearing cannot be completely mitigated (see Figure 4.1), we reduce the impact during calibration by flagging the baselines that are most severely affected (see Section 6.4).

The observations we have used have variations in the stations used. The observation with ID L816272 has the largest number of stations (52), as it includes

[3] lta.lofar.eu

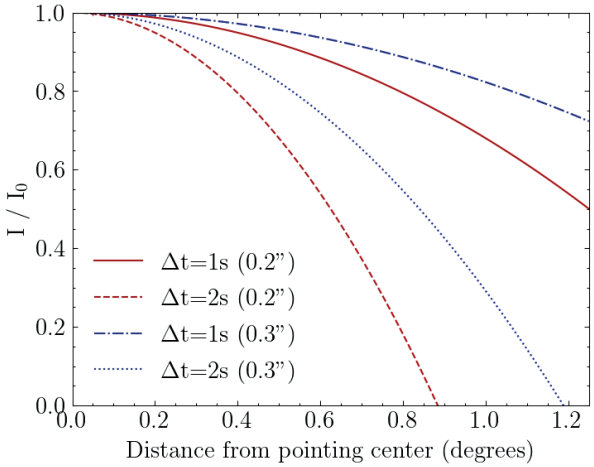


Figure 4.1: Intensity I as a function of distance from the pointing centre due to a combination of bandwidth and time smearing over the original intensity at the pointing centre I_0. We used the smearing formulas according to Bridle & Schwab (1999) with a central frequency of 140 MHz and a bandwidth of 12.21 kHz. This plot includes the smearing from the longest baseline (1980 km) between the LOFAR stations in Birr (Ireland) and Łazy (Poland), reaching a resolution of about $0.2''$ (red). We also added I/I_0 for our target resolution of $0.3''$ (blue), corresponding to a baseline of 1470 km. This figure shows both the smearing for the 1 sec and 2 sec pre-averaged datasets.

also the Latvian station that only recently became operational (Vrublevskis et al., 2020). This adds more baselines longer than ~1700 km with the stations located in Ireland and France (see Table 4.2). Observation IDs `L686962` and `L769393` have the same stations as `L816272` but without the Latvian station. Observation `L798074` includes the Latvian station but misses one German and the Polish station in Łazy, leading to the absence of the longest LOFAR baseline (see Table 4.2). The different combinations of LOFAR stations result in different *uv*-coverages, as displayed in Figure 4.2. The *uv* sampling gaps between 80 and 180 km are due to the sparsity of LOFAR stations between the Dutch remote and German stations.

4.3. Calibration

Our calibration strategy of all our observations builds upon the procedures described in Morabito et al. (2022a) and Sweijen et al. (2022c), where we further refined parts of their calibration strategy. Sweijen et al. (2022c) averaged their data to a time resolution of 2 sec. Given that half of our data is averaged to 2 sec while the remainder is at a 1-second resolution, it follows that our data volume is about 6 times larger than the data from Sweijen et al. (2022c). This introduces additional challenges

	Łazy (PL611HBA)	**Birr** (IE613HBA)	**Bałdy** (PL612HBA)	**Nançay** (FR606HBA)	**Borówiec** (PL610HBA)
Irbene (LV614HBA)	X	1930 km	X	1735 km	X
Birr (IE613HBA)	1980 km	X	1880 km	X	1679 km

Table 4.2: Five longest baselines between LOFAR stations according to the Euclidean distance of the Earth-centered coordinates of these stations. The station IDs are given between brackets.

regarding storage and computational demands, leading to our decision to utilise a high-throughput compute cluster named Spider for our full data processing.[4] Spider enables us to run many of our jobs embarrassingly parallel, which reduces the wall-clock time of our full data processing.

In the following subsections, we discuss the calibration workflow starting with downloading the data to arriving at the final DI-and DD-corrected solutions necessary for imaging. We will also highlight the improvements we have made compared to previous work.

4.3.1. Initial Dutch calibration

The first calibration steps focus on calibrating the *uv*-data of the Dutch stations. This follows the standard procedure, similar to LoTSS (Shimwell et al., 2017, 2019, 2022), but with the goal to pre-process our data up to the stage where we can start with the calibration of the international stations (Morabito et al., 2022a; Sweijen et al., 2022c). The steps, as described in this section, are summarised in the workflow from Figure 4.3.

After downloading the data from the LTA, we ran the standard data reduction pipeline from `Prefactor`[5] on our four separate datasets (van Weeren et al., 2016b; de Gasperin et al., 2019b). This pipeline starts with a calibration of our primary calibrator 3C 295. `Prefactor` corrects for all stations the phase differences between XX and YY polarisations, and derives constant clock offsets between the stations, and the bandpass. The pipeline proceeds with the target pipeline where the goal is to correct the Dutch stations of the target data of ELAIS-N1 for DI effects. This starts by transferring the calibrator solutions to the target data and removing the international stations to reduce the data volume. The pipeline involves flagging

[4] `https://doc.spider.surfsara.nl`

[5] Predecessor of the LOFAR Initial Calibration (LINC) pipeline.

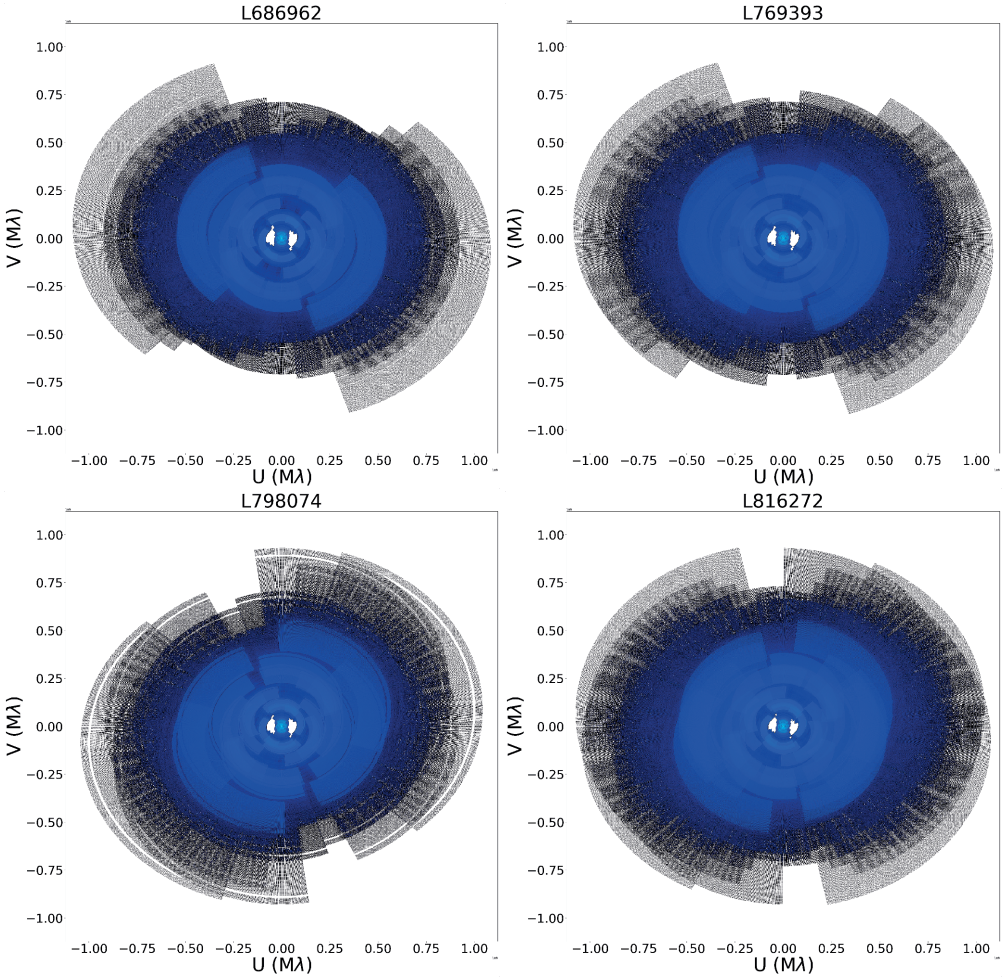


Figure 4.2: *uv*-coverage of all four LOFAR observations utilised in this paper. These define the shape of the dirty beam. The *uv*-coverages in this plot include flagging and are plotted with conjugate *uv* points. They also include the full frequency bandwidth, which produces the radial extent. These figures are made with the Python library `shadems`.

bad data and problematic stations, finding Faraday corrections with `RMextract` (Mevius, 2018), and phase calibration against a sky model from the TIFR Giant Metrewave Radio Telescope (GMRT) Sky Survey (TGSS, Intema et al., 2017). This procedure results in the first DI solutions for all Dutch stations. We utilised the solution inspection plots from `LoSoTo`[6] to conduct a first assessment of the quality of our observations and to notify whether there were substantial parts of the data

[6] `https://github.com/revoltek/losoto/`

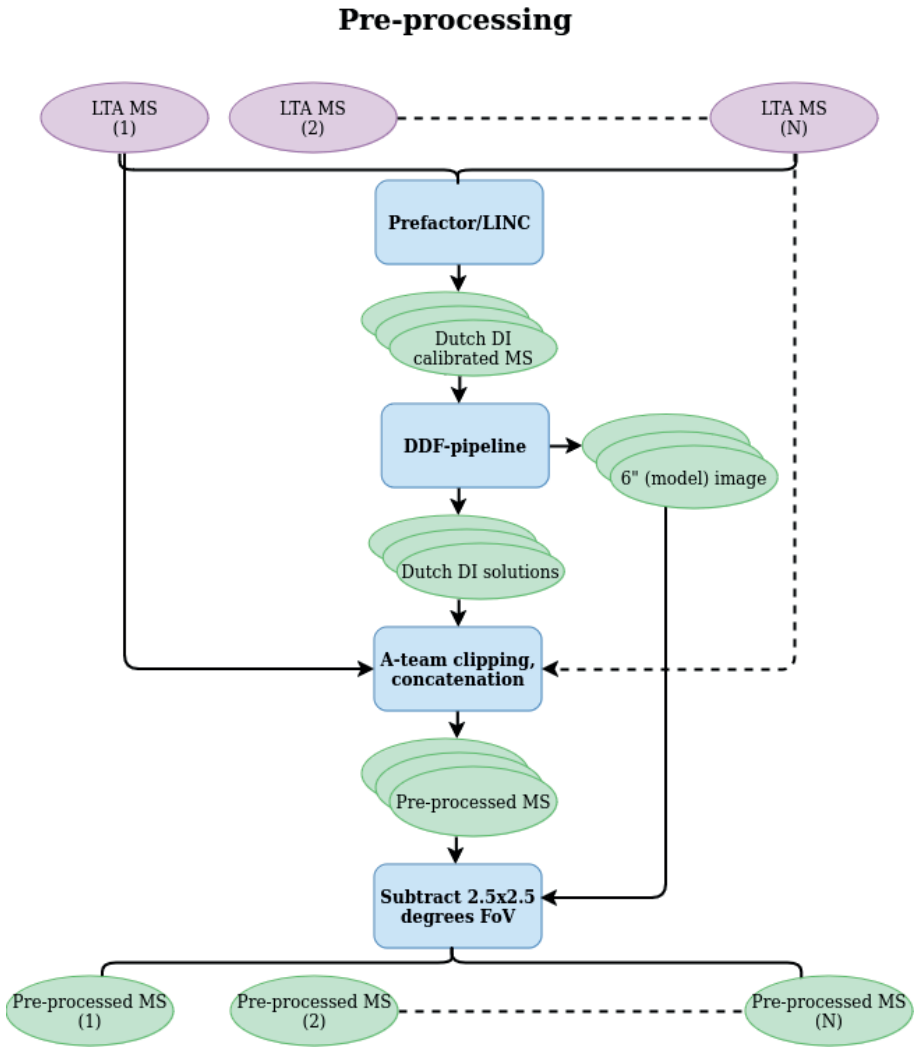


Figure 4.3: Workflow corresponding to the calibration steps explained in Section 4.3.1 for the general case with N observations. The workflow starts with the uv-data pulled from the LTA and ends with pre-calibrated uv-data, ready for calibrating the international stations for N different observations of the same field. Purple ovals are input data, blue boxes are operations on the data, and green ovals are output data. Stacked ovals imply that there are output products for each observation. Dashed lines indicate the presence of numerous observations that can run for this workflow in parallel. For a description of the calibration operations, we refer to Table 4.3.

flagged or entire stations removed.

Using the output from the `Prefactor` target pipeline, we also ran the `DDF`-pipeline[7] to obtain DI and DD corrections, DDE-corrected images, and DDE-corrected models for the Dutch core and remote stations (Shimwell et al., 2019; Tasse et al., 2021). This pipeline uses `KillMS`[8] (Tasse, 2014a,b; Smirnov & Tasse, 2015) to derive phase and amplitude corrections, which are applied during imaging with `DDFacet`[9] (Tasse et al., 2018). The resulting 6″ wide-field images of each of our 8-hrs observations were used to assess the quality of the corrections on the data from the Dutch stations and to gauge the calibratability of the ionosphere during each of our selected observations.

Following the strategy from Morabito et al. (2022a), we prepare our data for international DI calibration by transferring the Dutch DI calibration solutions, predicting and flagging (part of) the response of bright off-axis so-called 'A-team' sources (Cassiopeia A, Cygnus A, Taurus A, and Virgo A), and perform concatenation of datasets into subbands of 1.95 MHz. Unlike when processing just Dutch stations, we do not perform any averaging. Given that the full width at half maximum (FWHM) is smaller at sub-arcsecond resolutions, due to the size difference between Dutch and international LOFAR stations, we adopt a narrower field of view compared to the 6″ resolution. We made therefore use of the `DDF`-pipeline models and solutions to subtract sources outside a box of 2.5 × 2.5 deg^2 centred on the pointing centre. This box size sets the field of view of our final image products. The subtraction step suppresses artefacts induced by sources outside this field of view. After performing these steps, we have a final total data volume of ~12 TB after compression (Offringa, 2016). This total consists of 4 TB for the two datasets with 1-sec resolution and 2 TB for the two datasets with 2-sec resolution. The data is now prepared for calibration with international stations using DI.

4.3.2. Direction independent calibration of full array

After obtaining the pre-calibrated data using existing pipelines, as described in the previous subsection, we proceed with the initial DI calibration of the international stations. This step is challenging as there are fewer suitable calibrators available with enough S/N compared to observations at lower resolutions (Morabito et al., 2022a; Jackson et al., 2022). Finding the best fitting calibration strategy remains partly empirical and therefore needs additional attention. In Figure 4.4, we illustrate the workflow starting with the pre-calibrated data, as detailed in Section 4.3.1, and concluding with the final DI calibrated data, discussed in this subsection.

[7] `https://github.com/mhardcastle/ddf-pipeline`
[8] `https://github.com/saopicc/killMS`
[9] `https://github.com/saopicc/DDFacet`

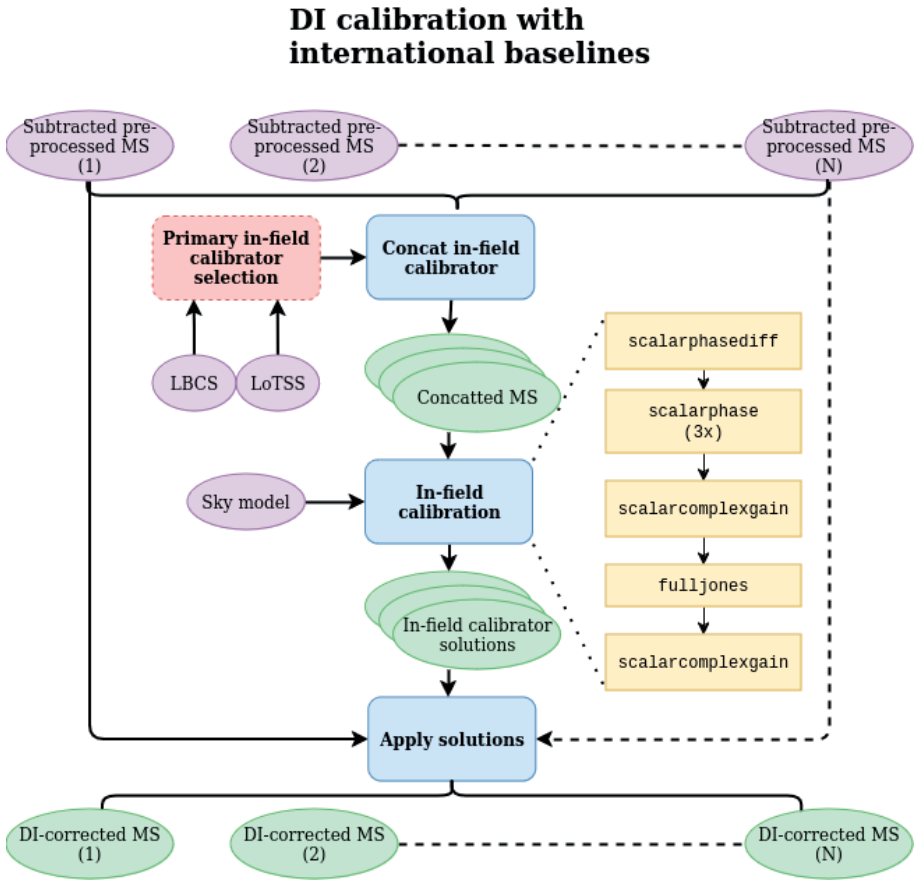


Figure 4.4: Workflow corresponding to the calibration steps explained in Section 4.3.2 for the general case with N observations. The workflow starts with pre-processed uv-data and ends with DI-calibrated uv-data for N different observations of the same field. These steps follow after the workflow in Figure 4.3. Purple ovals are input data, blue boxes are operations on the data, red boxes are data filters, yellow boxes are calibration steps, and green ovals are output data. Stacked ovals imply that there are output products for each observation. Dashed lines indicate the presence of numerous observations that can run in parallel. For a description of the calibration operations, we refer to Table 4.3.

4.3.2.1. Direction independent calibrator selection

An important step in the DI calibration is the selection of a suitable primary in-field calibrator. Not every bright source is a good primary in-field calibrator (Jackson et al., 2016, 2022). Proxies for good DI calibrators are:

- **S/N**: The source must exhibit high S/N on the longest baselines, ensuring

sufficient signal to calibrate the phases and amplitudes of the data from the international stations. The source should ideally be one of the brightest within the field of view, having a peak intensity of at least ~25 mJy beam^{-1} at 0.3″.

- ☐ **Position**: The primary in-field calibrator needs to be well within the FHWM of the international station beam to avoid too much attenuation due to the primary beam. Therefore, it is desirable to have a source located within 1° of the pointing centre.

- ☐ **Polarisation (optionally)**: If the information is available, it is an advantage to select an unpolarised in-field calibrator, as this allows for polarisation calibration on the in-field calibrator; we explain this further in Section 4.3.2.3.

Identifying the best in-field calibrator is essential as bad amplitude or phase corrections will be largely irreversible. This is due to the higher time and frequency resolution that we use when calibrating the primary in-field calibrator compared to the time and frequency resolution used when correcting DDEs, as we later discuss in Section 6.4.

Fortunately, we already knew from the calibration of ELAIS-N1 by Ye et al. (2024) which source satisfied the in-field calibrator selection criteria above. For their selection, they used the Long-Baseline Calibrator Survey (LBCS, Jackson et al., 2022), and selected the Seyfert 2 galaxy identified by ICRFJ160607.6+552135 (Charlot et al., 2020; Sexton et al., 2022). This source, with a compact flux density of ~0.28 Jy at 140 MHz, is located about 0.8° away from the ELAIS-N1 pointing centre. Moreover, there is no evidence to suggest that this source is polarised (Herrera Ruiz et al., 2021; Callingham et al., 2023).

4.3.2.2. Sky model

For the calibration of the primary in-field calibrator, we constructed a point source sky model, as our source does not show any structure at sub-arcsecond scale. In order to determine the spectral index for our calibrator, we used the measured flux densities from observations by the NRAO VLA Sky Survey at 1.4 GHz (NVSS; Condon et al., 1998), the GMRT at 610 MHz (Garn et al., 2008), the Westerbork Northern Sky Survey at 325 MHz (WENSS; Rengelink et al., 1997), LoTSS DR2 at 144 MHz (Shimwell et al., 2022), and the 6C and 7C survey at 151 MHz (Vollmer et al., 2010). We found the flux density to turn over between WENSS and LOFAR HBA frequencies, which led us to decide to better characterise the spectrum by processing an LBA observation at 54 MHz from the ELAIS-N1 field using the `LiLF`[10] calibration pipeline (de Gasperin et al., 2018, 2019a, 2020). We have also added

[10] `https://github.com/revoltek/LiLF`

the flux density limit from the Very Large Array Low-frequency Sky Survey Redux (VLSSr; Lane et al., 2014). This supports the accuracy of our fitted spectrum. With the flux densities and frequencies, we fitted a second-order logarithmic polynomial

$$\log S(\nu) = \log S_0 + c_0 \log\left(\frac{\nu}{\nu_0}\right) + c_1 \log\left(\frac{\nu}{\nu_0}\right)^2,$$

where S is the flux density as a function of frequency ν. Using ν_0 as the reference frequency at 141 MHz, we found $\log S_0 = 2.45$, $c_0 = 1.11$ and $c_1 = -1.13$. This gave the fit shown in Figure 4.5. With these results, we obtained a flux density at 141 MHz of 0.28 Jy and the following spectral index as a function of frequency:

$$\alpha = \frac{\delta \log S(\nu)}{\delta \log \nu} = -2.26 \log\left(\frac{\nu}{\nu_0}\right) + 1.11.$$

Based on Charlot et al. (2020), we also have the coordinates of our in-field calibrator with a positional precision for the RA and DEC of dRA=1.88 mas and dDEC=1.83 mas respectively. We used this information together with the source spectral index as input for our point source sky model. Our sky model ensures an accurate astrometry and flux density scale (as we later demonstrate in Sections 4.6.3 and 4.6.4).

4.3.2.3. In-field calibration

Before performing any calibration on our selected in-field calibrator, we first phase-shifted the visibilities to the position of our calibrator source, after which we averaged the data down to 488 KHz and 32 sec, which decreases the data volume by a factor 640 or 1280 (depending on the original 2 sec or 1 sec data resolutions). This was followed by a primary beam correction at the position of the in-field calibrator. Applying the primary beam after averaging helps reduce computational time and is justified by the fact that the beam only varies very slowly as a function of time and frequency. We express the operations to prepare our data for calibration mathematically by using the radio interferometry measurement equation (RIME; Hamaker et al., 1996; Smirnov, 2011a,b) as follows:

$$V_{\rm IF} = \mathbf{B}_{\rm IF} \langle \mathbf{P}_{\rm IF} V_{\rm pre} \mathbf{P}_{\rm IF}^H \rangle \mathbf{B}_{\rm IF}^H,$$

where $V_{\rm pre}$ are the visibilities with pre-applied Dutch solutions and the subtracted $6''$ model outside our $2.5° \times 2.5°$ field of view. $\mathbf{B}_{\rm IF}$ is the beam correction in the direction of the in-field calibrator, $\mathbf{P}_{\rm IF}$ is the phase-shift to the position of the in-field calibrator, $V_{\rm IF}$ the corresponding visibilities centred on the in-field calibrator

as starting point of the calibration. The angular brackets represent the averaging over time and frequency, while H denotes the conjugate transposed matrix.

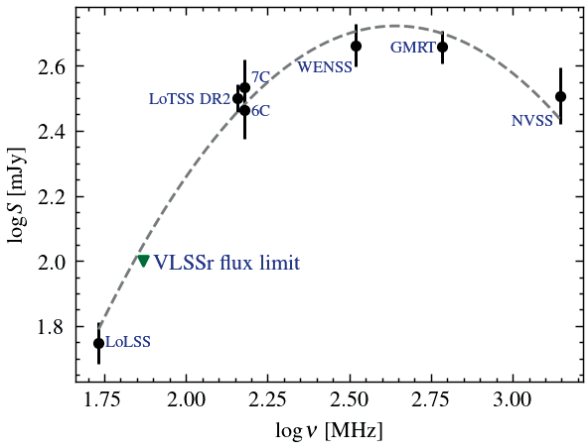


Figure 4.5: Fit of the radio spectrum of our primary in-field calibrator using data from WENSS at 330 MHz, the GMRT at 610 MHz, NVSS at 1.4 GHz, LoTSS DR2 at 144 MHz, the 6C and 7C surveys at 151 MHz, and the LoLSS image at 54 MHz constructed by us. We added the green downward triangle from the VLSSr flux density limit to illustrate the accuracy of the spectrum at lower frequencies.

For the DI calibration of our in-field calibrator, we used `facetselfcal`,[11] which utilises the Default Preprocessing Pipeline (`DP3`, van Diepen et al. (2018a); Dijkema et al. (2023)) and `WSClean` (Offringa et al., 2014) to perform (self-)calibration on a source. This calibration algorithm allowed us to derive the best phase and amplitude solutions on station level through the minimisation of the difference between our sky model and the input visibilities. `facetselfcal` uses an "iterative-perturbative" approach, where after each calibration step the model is adjusted with the new solutions before going to the next step. This procedure gives us full flexibility to incorporate our own calibration strategy to correct for different effects on different time, frequency, and antenna selections, on our data. The calibration steps used in this paper are described in Table 4.3. To allow ourselves to tailor the magnitude of phase corrections as a function of time and frequency for different subsets of antennas, we split below the `scalarphase` calibration up into three separate iterations (`scalarphase` I, II, and III), wherein each iteration we reset the solutions for a set of antennas to phase 0 and amplitude 1 values after running the calibration operation. After experimenting with various solution intervals, smoothness constraints, and calibration steps from `facetselfcal`, we found that the strategy described below

[11] `https://github.com/rvweeren/lofar_facet_selfcal`

Operation name	Description
`scalarphase`	Solving for phase errors as a function of time and frequency, independent of polarisation.
`scalarcomplexgain`	Solving for phase and amplitude errors as a function of time and frequency, independent of polarisation.
`fulljones`	Solving for phase and amplitude errors as a function of time and frequency for both diagonal and cross-hand polarisations simultaneously.
`scalarphasediff`	Derive a diagonal phase correction to eliminate any phase difference present between circular RR and LL (or linear XX and YY) polarisations as a function of time and frequency.

Table 4.3: Description of the calibration operations used in this paper. The names originate from the operation names used by `DP3` (van Diepen et al., 2018a; Dijkema et al., 2023) and `facetselfcal` (van Weeren et al., 2021).

performed best on our in-field calibrator source. This strategy is illustrated by a selection of solution plots from different LOFAR stations in Figures 4.6 and 4.7.

1. `scalarphasediff`: Our in-field calibrator is unpolarised (Tremblay et al., 2016; Herrera Ruiz et al., 2021; Callingham et al., 2023). The absence of a signal in Stokes V polarisation enables us to employ `scalarphasediff` calibration in circular polarisation basis to correct for differential Faraday rotation after converting our data polarisation basis from linear to circular. We constrain the Dutch stations for this step to have the same solutions, as the effect of differential Faraday rotation is negligible on shorter baselines. We found a suitable solution interval for this step to be 8 min and the frequency to be best constrained by a smoothness kernel of 10 MHz. The varying calibration solutions for the international stations are illustrated in the first row of Figure 4.6.

2. `scalarphase I`: After having corrected the RR and LL polarisation phase difference, we derive polarisation-independent corrections for phase errors with the `scalarphase` solve. In the first `scalarphase` iteration, we solve for 'fast' phase variations for the international stations by taking a solution interval of 32 sec and a small frequency smoothness kernel of 1.25 MHz. These are the smallest solution interval and frequency smoothness constraints, as we expect the largest phase variability across the longest baselines. The reset option setting the phase solutions to 0 and amplitude solutions to 1 for the Dutch core and remote stations results in only solutions for the international stations. On the second row in Figure 4.6 we see how the solutions corresponding to this step are wrapping

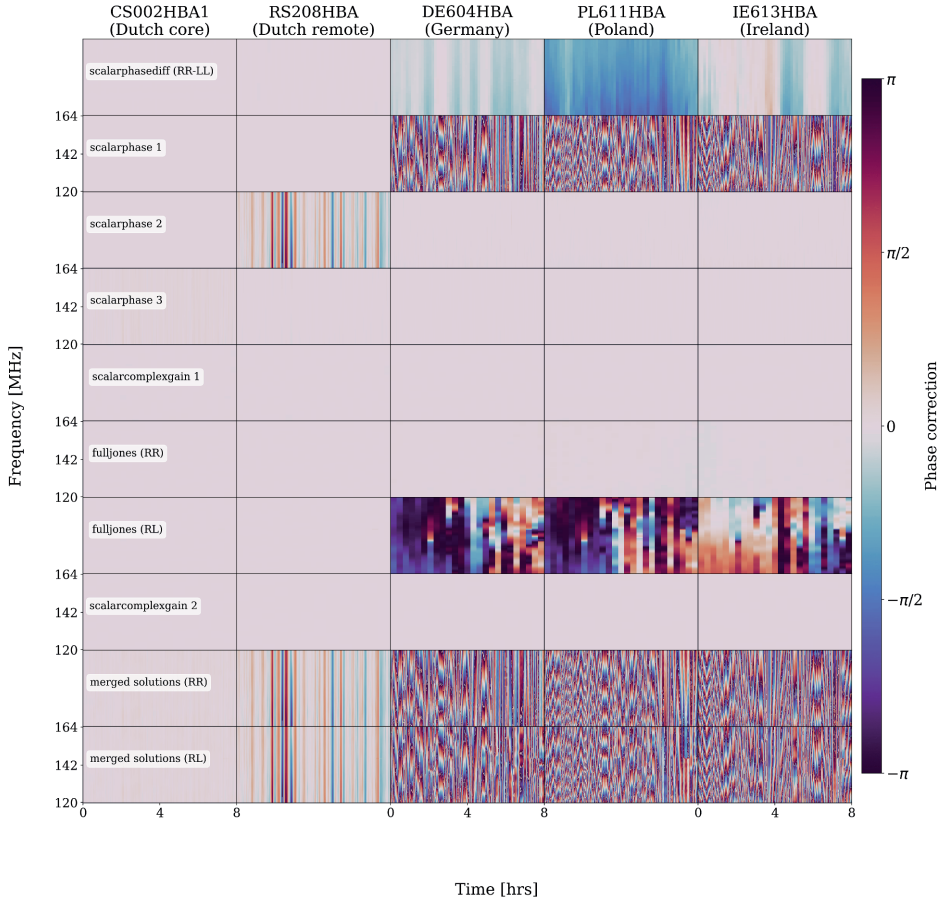


Figure 4.6: Phase calibration solution plots corresponding to the different calibration steps (*rows*) and different stations, given by their station IDs (*columns*), for calibrating the primary in-field calibrator. These solutions are relative to the CS001HBA0 Dutch core station. For the full-Jones corrections, we only show the RR and RL solutions. The solutions on the last two rows show how these solutions are combined into a final merged solutions. It is important to note that the `scalarcomplexgain` and `fulljones` corrections have small phase corrections for RR (and LL) polarisations, due to the fact that these are already corrected in the previous steps. However, the same `scalarcomplexgain` steps do correct significantly for amplitudes (see Figure 4.7) and the `fulljones` step for the RL (and LR) polarisations.

fast from $-\pi$ to π radians for the international stations.

3. `scalarphase II`: In the second `scalarphase` iteration, we solve again for 'fast'

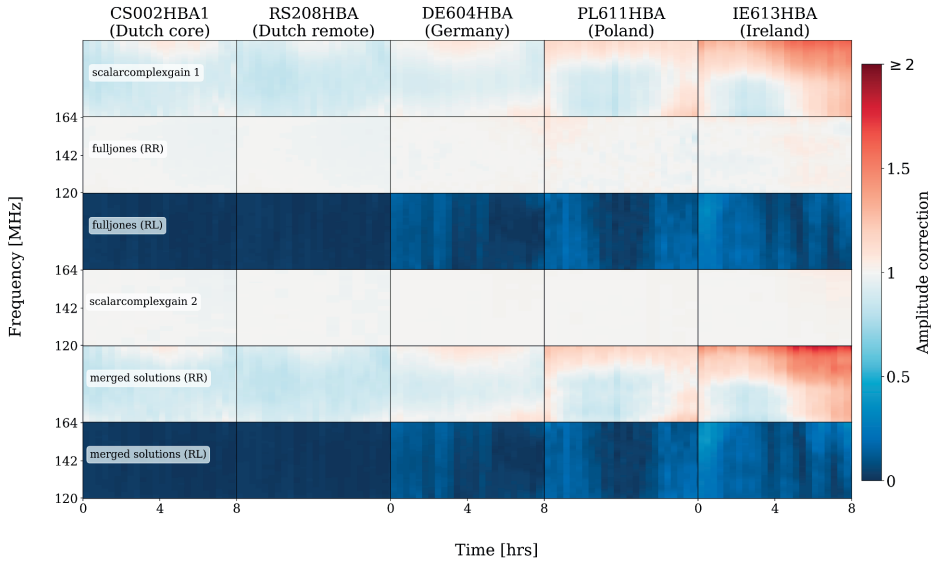


Figure 4.7: Amplitude calibration solution plots corresponding to the different calibration steps (*rows*) and different stations, given by their station IDs (*columns*), for calibrating the primary in-field calibrator. For the full-Jones corrections, we only show the RR and RL solutions. The solutions on the last two rows show how these solutions are combined into final merged solutions.

phase changes with a solution interval of 32 sec. However, we now include the Dutch remote stations by only resetting the solutions for the Dutch core stations to phase solutions equal to 0 and amplitude solutions equal to 1 after the solve. Compared to the previous `scalarphase` solve, we found a larger frequency constraint with a smoothness kernel of 10 MHz to work best. The solutions are most significant for remote stations because the phases for the international stations are already corrected, as is illustrated on the third row in Figure 4.6.

4. `scalarphase III`: In the third `scalarphase` iteration we solve for 'slow' phase changes for all stations, including the Dutch core stations, by taking a solution interval of 20 min and without using a reset of solutions. With a smoothness kernel of 20 MHz we use a larger frequency constraint compared to the other two `scalarphase` iterations. The Dutch core stations observe a similar ionosphere and were already corrected for DI effects (see Section 4.3.1). This results in small corrections between these stations, as illustrated on the fourth row in Figure 4.6.

5. **`scalarcomplexgain` I**: After correcting for phase errors, we also incorporate polarisation-independent phase and amplitude corrections by doing a 'slow' `scalarcomplexg` solve with a solution interval of 20 min. We constrained the frequency axis here by a smoothness kernel of 7.5 MHz. On the fifth row in Figure 4.6 we see that the phase corrections are negligible, due to the phase corrections from the previous iterations. The amplitude corrections on the first row of Figure 4.7 are most significant for the more distant international stations.

6. **`fulljones`**: After having corrected for phases and amplitudes for the diagonal RR and LL polarisation directions, we also correct with a full-Jones correction for leakage in the RL and LR cross-hands. As we have already applied full-Jones DI corrections for the Dutch stations (see Section 4.3.1), we expect the leakage of Dutch stations to be similar. Hence, we constrained these stations to have the same value. This also boosts the calibration signal at these stations. We opt for solution intervals of 20 min and constrain the frequencies with a smoothness frequency kernel of 5 MHz. In Figures 4.6 and 4.7, we find the most significant corrections for the off-axis polarisations of the international stations.

7. **`scalarcomplexgain` II**: Finally, we performed an additional final round of scalar corrections, by using a slow `scalarcomplexgain` solve set to a solution interval of 40 min and a frequency smoothness constraint of 7.5 MHz. This step serves as a final verification to ensure the stability of the solutions. In Figures 4.6 (eighth row) and 4.7 (fourth row), we see that the corrections are minor compared to the solutions from the previous calibration steps. This confirms the reliability of the iterative calibration up to the full-Jones calibration.

Throughout the calibration, we ignored baselines with a length smaller than 20,000 times the wavelength (λ), by setting a constraint on the `uvmin` parameter. This is to prevent possible issues related to having an incomplete sky model. This `uvmin` corresponds to a largest angular scale (LAS) of ~10″ at 140 MHz. Each calibration step returned an `h5parm` solution table. We merged all solutions derived for each of the four observations to obtain four final solution tables with phase and amplitude corrections. If we let $\mathbf{G}_{\rm DI}$ represent the final solutions after merging all solutions, we express the RIME equation to obtain the final DI calibrated visibilities on the $V_{\rm pre}$ visibilities with pre-applied Dutch solutions as

$$V_{\rm DI} = \mathbf{B}_{\rm IF}^{-1}\mathbf{G}_{\rm DI}\mathbf{B}_{\rm IF}V_{\rm pre}\mathbf{B}_{\rm IF}^{H}\mathbf{G}_{\rm DI}^{H}\mathbf{B}_{\rm IF}^{H-1}, \tag{4.1}$$

where $\mathbf{B}_{\rm IF}^{-1}$ is the inverse beam correction from the centre of the in-field calibrator back to the pointing centre of the ELAIS-N1 observation.

It is important to stress that the order of merging solutions is essential, as our `scalarphasediff` and `fulljones` corrections do not commute. This implies that we need to merge the solutions in the order of the steps we have iteratively solved for. Similarly, the order of applying the beam corrections ($\mathbf{B}_{\rm IF}$ and $\mathbf{B}_{\rm IF}^{-1}$) and the full-Jones solutions from ($\mathbf{G}_{\rm DI}$) in Equation 4.1 are not commutative either, due to the fact that they are typically not simultaneously diagonalisable Jones matrices (Smirnov, 2011a,b).

4.3.3. Direction-dependent calibration of full array

The ionosphere and errors in the beam model introduce DDEs that corrupt the 'real' visibilities across the field of view. These are not corrected by the DI calibration, as they depend on the direction of the calibration. We, therefore, divided the sky area up into smaller facets by selecting and calibrating for bright compact secondary calibrators distributed across ELAIS-N1 (van Weeren et al., 2016b; Williams et al., 2016). The main challenge in the selection is that from the best existing radio images, we only have source information available at $6''$ resolution, while we need to find compact calibrators that have enough S/N to calibrate at $0.3''$ resolution. It is therefore vital, after the initial selection and performing self-calibration on the DD calibrators, to examine both the calibration solutions and the resulting images to ensure that we have selected good calibrators with good calibration solutions. The workflow discussed in this subsection is illustrated in the diagram in Figure 4.8 for the general case of N observations.

4.3.3.1. Direction-dependent calibrator selection

To initiate the search for compact sources, we used the ELAIS-N1 deep-field catalogue constructed from a $6''$ resolution LOFAR HBA map (Kondapally et al., 2021; Sabater et al., 2021). From this catalogue, we selected a sample of 86 sources with peak intensities above 25 mJy beam^{-1} inside our $2.5°\times2.5°$ field of view. To investigate whether these sources may be good calibrators, the sources were first all split off by phase-shifting the DI corrected visibilities from Equation 4.1. We averaged the phase-shifted data down to 32 sec and 390.56 kHz, which decreased the data volume by a factor 512 or 1024 (depending on the original 2 sec or 1 sec data resolutions). The averaging also reduced the effects from other nearby sources, without introducing smearing effects in our calibrator data. The full procedure can be expressed as

$$\begin{aligned} V_{{\rm S},n} &= \mathbf{B}_{{\rm S},n}\langle\mathbf{P}_{{\rm S},n}\mathbf{B}_{\rm IF}^{-1}\mathbf{G}_{\rm DI}\mathbf{B}_{\rm IF}V_{\rm pre}\mathbf{B}_{\rm IF}^{H}\mathbf{G}_{\rm DI}^{H}\mathbf{B}_{\rm IF}^{H-1}\mathbf{P}_{{\rm S},n}^{H}\rangle\mathbf{B}_{{\rm S},n}^{H} \\ &= \mathbf{B}_{{\rm S},n}\langle\mathbf{P}_{{\rm S},n}\mathbf{V}_{\rm DI}\mathbf{P}_{{\rm S},n}^{H}\rangle\mathbf{B}_{{\rm S},n}^{H}, \end{aligned}$$

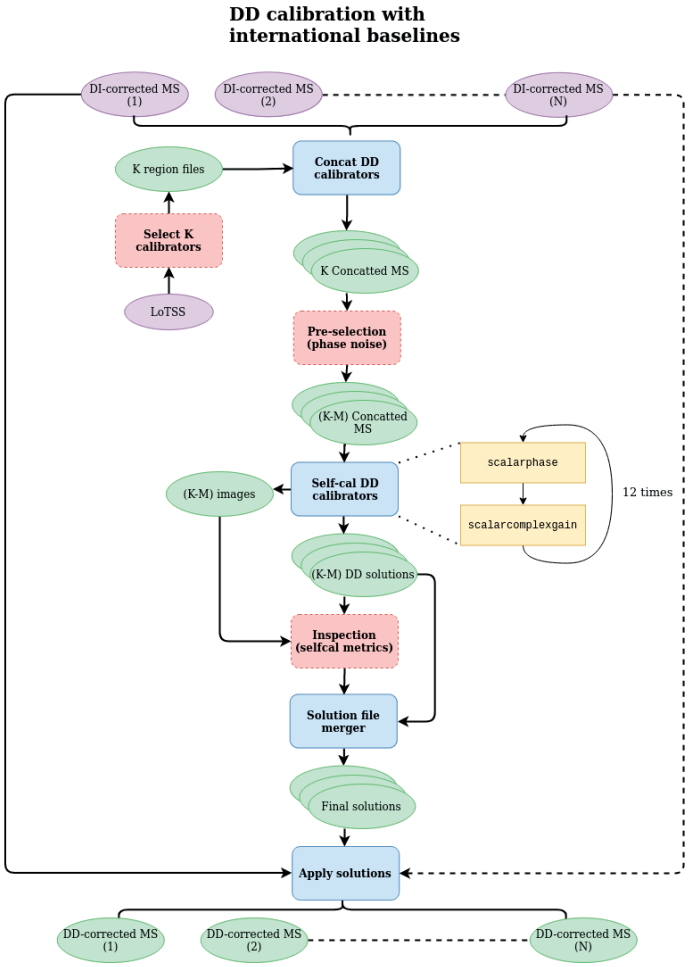


Figure 4.8: Workflow corresponding to the calibration steps explained in Section 6.4 for the general case with N observations. The workflow starts with DI-corrected uv-data and ends with DD-corrected uv-data for N different observations of the same field. These steps follow after the workflow in Figure 4.4. In the first source selection, based on the brightest sources from the LoTSS catalogue, we select K sources, after which M of these are filtered out during the phase noise selection metric (see Section 4.3.3.1). This leaves us with $(K - M)$ solutions for each of the N observations. We note that the `scalarcomplexgain` is only optionally triggered in `facetselfcal` for brighter sources (See van Weeren et al., 2021). Purple ovals are input data, blue boxes are operations on the data, red boxes are data filters, yellow boxes are calibration steps, and green ovals are output data. Stacked ovals imply that there are output products for each observation. Dashed lines indicate the presence of numerous observations that can run in parallel. For a description of the calibration operations, we refer to Table 4.3.

where $V_{\mathrm{S},n}$ are the visibilities after applying the DI corrections, phase-shifting and beam corrections in the direction of source n, and where we substituted Equation 4.1 on the second line.

We expect a significant fraction of the selected sources to be resolved out at $0.3''$ resolution. If we were to run self-calibration naively on all 86 candidates and visually examine the results, it would not only cost extensive manual inspection time but it would also be computationally demanding. We therefore came up with a computationally cheap but reliable metric to identify which of our 86 candidate sources have enough S/N at the longest baselines, as we outline below.

For the selection metric we use the fact that circularly polarised sources are very rare at low frequencies, as Callingham et al. (2023) found in their $20''$ V-LoTSS survey at 144 MHz only 68 circularly polarised sources across 5634 deg^2. One of their detections appears within our field of view (Callingham et al., 2021) but is not in our list of candidate calibrators. Considering that Callingham et al. (2023) reports a completeness above 1 mJy and all our ELAIS-N1 calibrators have flux densities exceeding 25 mJy, it is reasonable to assume that none of our 86 calibrators are strongly circularly polarised. This implies that differences between corrections on RR and LL polarisations of our calibrators are attributed to the amount of noise on the solutions (ignoring polarisation leakage variations and the small effect of DD differential Faraday rotation, see further below). Therefore, calibrators with a high S/N will have more similar phase corrections on both RR and LL polarisations, whereas diffuse sources with a low S/N will exhibit noisier phase corrections.

To quantify the differences in RR and LL polarisations we first run one round of `scalarphasediff` calibration with `facetselfcal` (see Table 4.3). To make the assessment consistent when comparing different sources, we use a fixed solution interval of 10 min. Since we are interested in the amount of S/N at the longest baselines, we only consider the `scalarphasediff` solutions from the Dutch and German stations. From the obtained solutions we take the discrete difference along the frequency axis to account for small differential Faraday rotation, after which we use the circular standard deviation as a measure of the phase noise. The circular standard deviation serves as an alternative to the traditional standard deviation to account for phase wrapping (e.g. Mardia, 1972). The formula for the circular standard deviation is given by

$$\sigma_c = \sqrt{-2 \ln R},$$

where R is the mean resultant length given by

$$R = \sqrt{\bar{C}^2 + \bar{S}^2},$$

with the mean cosine angles

$$\bar{C} = \frac{1}{N}\sum_{i=1}^{N}\cos\theta_i$$

and the mean sine angles

$$\bar{S} = \frac{1}{N}\sum_{i=1}^{N}\sin\theta_i$$

for N phase solutions (θ). To test this metric and find a rejection threshold for the standard deviation, we applied this first on 40 sources from one of our four observations. With this sample, we empirically found sources below a circular standard deviation of 2.3 rad to be sufficient for self-calibration with the international LOFAR stations. Upon rejecting sources above this threshold from our initial 86 sources, we were left with 30 candidate sources. One of the 30 sources was less than 0.1° away from a neighbouring selected calibrator, which made us decide to pick the source with the lowest circular standard deviation value. This procedure narrowed our selection down to 29 candidates.

This phase noise metric takes ~1 CPU hr for each source, which includes averaging down the *uv*-data to calculate the circular standard deviation. By implementing this metric we reduced the number of self-calibration runs by about a factor of 3, compared to running self-calibration on all 86 original candidates. This decreases the total computing time in our case of using 4 observations by ~18,000 CPU hrs. While this is a small fraction of the total computing costs (see Section 4.4.3), it does remove a large part of the visual inspection when doing this fully automated (see also Appendix 4.A).

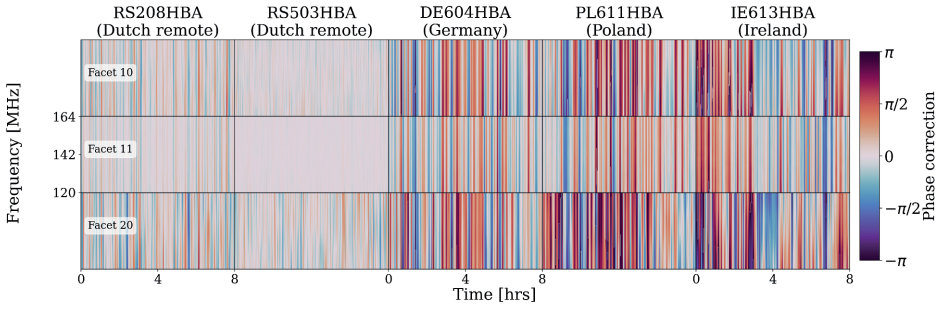


Figure 4.9: Merged phase calibration solution plots corresponding to the different facets (*rows*) and different stations, given by their station IDs (*columns*). These solutions are relative to the CS001HBA0 Dutch core station. The facets have the DD solutions from their corresponding calibrator, as depicted in Figure 4.14.

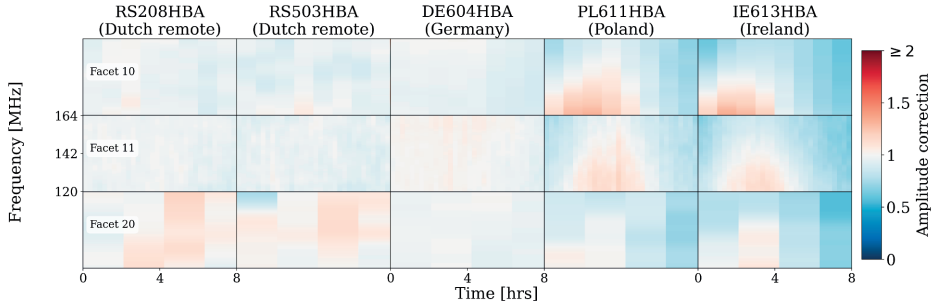


Figure 4.10: Merged amplitude calibration solution plots corresponding to the different facets (*rows*) and different stations, given by their station IDs (*columns*). The facets have the DD solutions from their corresponding calibrator, as depicted in Figure 4.14.

4.3.3.2. Self-calibration

For the remaining sources, we carried out up to 12 rounds of self-calibration by employing the `auto` option in `facetselfcal`. This calibration step is essential to calibrate for the ionospheric differences across the field of view. The number of cycles was set based on experience (Sweijen et al., 2022c; Ye et al., 2024), as it has consistently been shown to achieve convergence for good calibrators. The `auto` setting automatically adjusts the solution intervals, and frequency smoothness constraints, among other parameters, based on the available flux density from the source, as discussed in van Weeren et al. (2021). As we also have two observations that were pre-averaged by a factor of two in time (see Section 4.2), we utilised an additional setting, `flagtimesmeared`, that flags visibilities where the amplitude reduction due to time smearing is more than a factor of two. This is especially important at the longest baselines and for calibrators more distant from the pointing centre, as these suffer most from smearing (as we see in Figure 4.1 and later in Section 4.6.2). Before the first calibration, we let `facetselfcal` also apply a phase-up of the Dutch core stations into a superstation in the centre of the Dutch array. This suppresses the signal from nearby sources on shorter baselines. To further tune the calibration for structures on small angular scales, we apply the same *uv*-cut at 20,000λ as used for the DI calibration of the in-field calibrator (corresponding to a LAS of ~10″). The phase-up also reduces the data volume from each measurement by about 80% (Morabito et al., 2022a), which therefore speeds up the self-calibration significantly. The `auto` setting performs calibration during the first four rounds for phases by applying a `scalarphase` solve, while the subsequent eight rounds might, depending on the available S/N from the calibrator, apply `scalarcomplexgain` calibration to find amplitude corrections as well (van Weeren et al., 2021). To make sure that the flux density scale, after applying amplitude corrections, is not

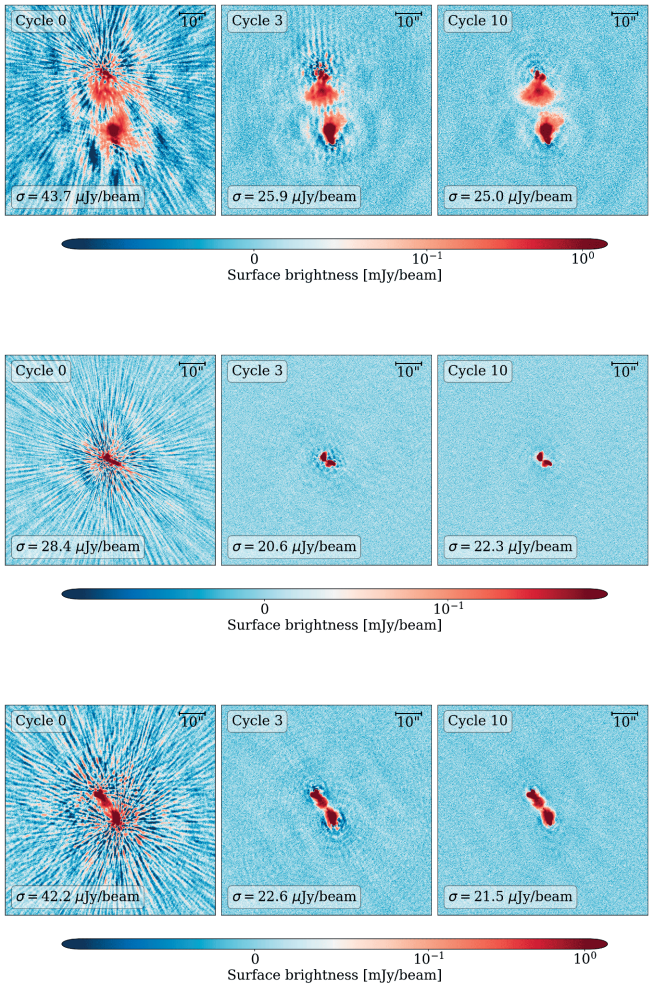


Figure 4.11: Three examples of self-calibration from our selected DD calibrators. Cycle 0 is the first image with only DI solutions applied from the in-field calibrator (see Section 4.3.2). Cycle 3 corresponds to the self-calibration image after 3 rounds of `scalarphase` calibration. After this cycle, `scalarcomplexgain` calibration is added. This also calibrates for amplitude errors. Cycle 10 shows the result after the 10th self-calibration round. In some cases, the RMS noise (given by σ in the figures) goes slightly up when comparing cycles 3 and 10. This is due to the introduction of amplitude corrections, which can cause slight increases or decreases in the overall local RMS values when for instance larger corrections for stations closer to the Dutch core are obtained. The angular size scale is indicated in the right top corner.

drifting, we normalise the global amplitude corrections over all antennas and our four observations to one.

The final phase and amplitude solutions for a selection of stations from three selected DD calibrators are given in Figures 4.9 and 4.10. This shows that we allow phases to have shorter solution intervals because we anticipate these to vary more rapidly over time than amplitudes. From the amplitude solutions we also see the solution interval size differences, which is because `facetselfcal` ensures that there is enough S/N on the longest baselines. To illustrate the self-calibration image quality, we show in Figure 4.11 three examples of sources with self-calibration cycles 0 (no correction), 3 (phase correction only), and 10 (phase and amplitude corrections). These demonstrate how the phase and amplitude corrections have improved the image fidelity.

4.3.3.3. Direction-dependent calibration inspection

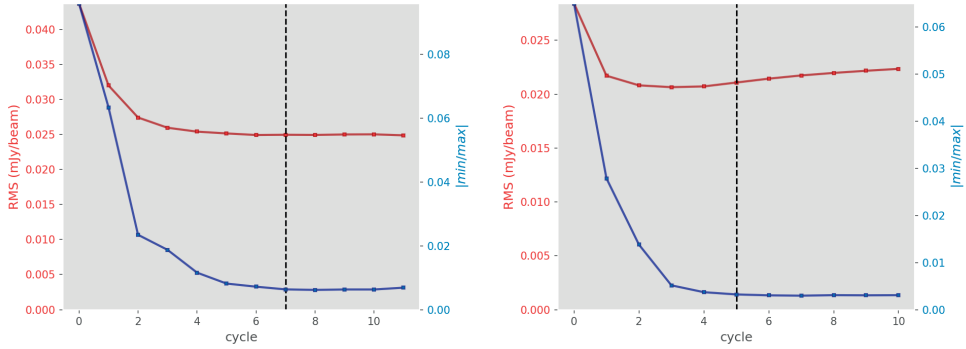


Figure 4.12: Self-calibration image stability for two different sources. The red line displays the progression of the RMS over self-calibration cycles, while the blue line represents the dynamic range (absolute min/max pixel) over self-calibration cycles. The black dashed line is the best calibration cycle according to a combined assessment of the solution and image stability. *Left panel*: This example corresponds to the self-calibration cycles of the source in the first row of Figure 4.11. *Right panel*: This example corresponds to the self-calibration cycles of the source in the second row of Figure 4.11.

Although self-calibration is well-established (e.g. Cornwell & Fomalont, 1999) and `facetselfcal` has proven to be reliable to calibrate our best candidate DD calibrators (e.g. van Weeren et al., 2021; de Jong et al., 2022; Sweijen et al., 2022c; Ye et al., 2024), it is essential to perform a final quality control on the self-calibration output products, as was done by Ye et al. (2024). This ensures our phase noise selection metric discussed in Section 4.3.3.1 did not include any false-positives and the calibration algorithm from `facetselfcal` performed as expected.

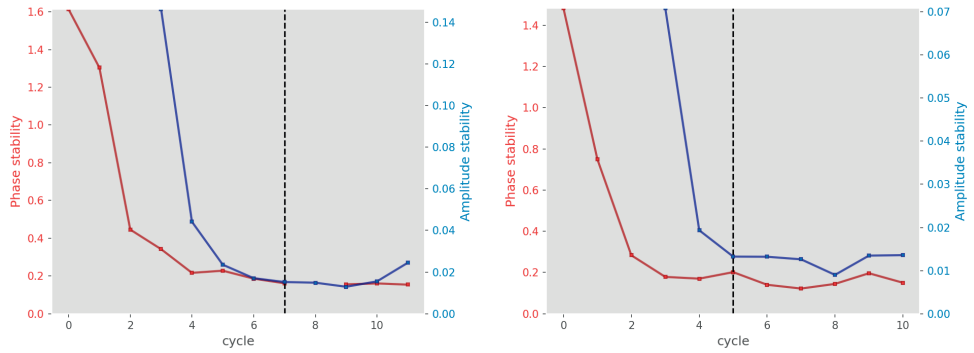


Figure 4.13: Self-calibration solution stability for two different sources (corresponding to the sources from Figure 4.12). The red line displays the circular standard deviation of the phase solution difference between the current and previous self-calibration cycle for each time and frequency value for each station. The blue line gives the standard deviation of the amplitude ratio of each time and frequency value for each station between the current and previous self-calibration cycle. We note that amplitude solves are only optionally triggered from cycle 3 onward, when the S/N is deemed sufficient by metrics from `facetselfcal`. The black dashed line corresponds to the selected calibration cycle, based on a combined assessment of the solution and image stability.

We inspect for each self-calibration cycle of each source the following characteristics:

- **RMS noise**: We expect for self-calibration improvements on calibrators with compact emission the RMS to be significantly lower compared to the image that is only calibrated with DI solutions (cycle 0 in Figure 4.11). In the left panel of Figure 4.12 we have an example of a stable improvement of the RMS. The source corresponding to the curve in the right panel has an RMS that is going up after cycle 4, which is due to the effects of the `scalarcomplexgain` calibration lifting the amplitude values. Although the increase is minor, this example demonstrates that relying exclusively on the image RMS to assess the self-calibration quality is insufficient.

- **Dynamic range**: Since the RMS does not fully convey the quality of the image, we also evaluate the dynamic range of the self-calibration images. We define the dynamic range in the figures as the absolute value of the minimum pixel value divided by the maximum pixel value. We expect for image improvements the most negative pixels to get closer to 0, which improves our measure of the dynamic range. Both cases in Figure 4.12 corresponding to two of our DD calibrator sources show dynamic range improvements.

- **Solution stability**: Another important metric is the stability of the solutions

over self-calibration cycle, as we expect the solutions to converge over self-calibration cycles. To assess this, we subtract the phase solutions for each time and frequency solution value between two consecutive self-calibration cycles and take the circular standard deviation as a measure of the stability. Similarly, to examine the behaviour of the amplitude solutions, we calculate the standard deviation of the ratio between solutions from two consecutive self-calibration cycles. Both measures should for stable self-calibration converge to small values, depending only on the solution noise. This converging behaviour is illustrated in Figure 4.13 for the same sources as in Figure 4.12.

Using these metrics, we can quickly assess both the quality of self-calibration and the best self-calibration cycle (see Appendix 4.A for automatic approaches). This aligns with the ad-hoc calibrator selection criteria implemented by Ye et al. (2024). Among our 29 self-calibrated sources we did not find any diverging behaviour. These results also reassured us that the phase noise selection metric, discussed in Section 4.3.3.1, did not select false-positive candidates.

During testing of our selection metrics, we did also run self-calibration on some of the sources that were above our phase noise selection threshold (see Section 4.3.3.1). Although sources with scores close to our selection threshold did slightly improve after self-calibration, were the corrections too small to sufficiently improve the DDEs within a facet. In Appendix 4.B we discuss two examples with strong divergent calibration behaviour. These illustrate the effectiveness of performing an additional inspection of the image and solution quality.

4.3.3.4. Facet layout

After selecting the best self-calibration cycles of each source, we merge all solutions from each direction together into one multi-direction solution file for each of our observations, stored in `HDF5` format (Folk et al., 2011). We also add phase and amplitude solutions of 0 and 1, respectively, from our primary in-field calibrator, as we did not have to apply self-calibration on this source after doing the thorough calibration, which includes DD calibration, as discussed in Section 4.3.2.3. The positions of the 30 calibrator sources (1 primary in-field calibrator source and the 29 DD calibrator sources) determine our facet layout through a Voronoi tessellation. This assigns each point in our field of view to the solutions of their nearest calibrator source. Across each facet, we assume that calibration solutions are constant (Schwab, 1984; van Weeren et al., 2016b).

In Figure 4.14 we show an $1.2''$ DI image from one of our observations, which we produced with `WSClean`[12] (Offringa et al., 2014), after applying the solutions

[12] `https://gitlab.com/aroffringa/wsclean`

from our DI calibrator. On top of this image, we projected the Voronoi tessellation corresponding to our 30 calibrators. The figure illustrates the successful correction of DDEs near the DI calibrator (indicated by the yellow star) and highlights how strong the DDEs are around our selected calibrator. Imaging including our final DD solutions is discussed in Section 6.4.5.

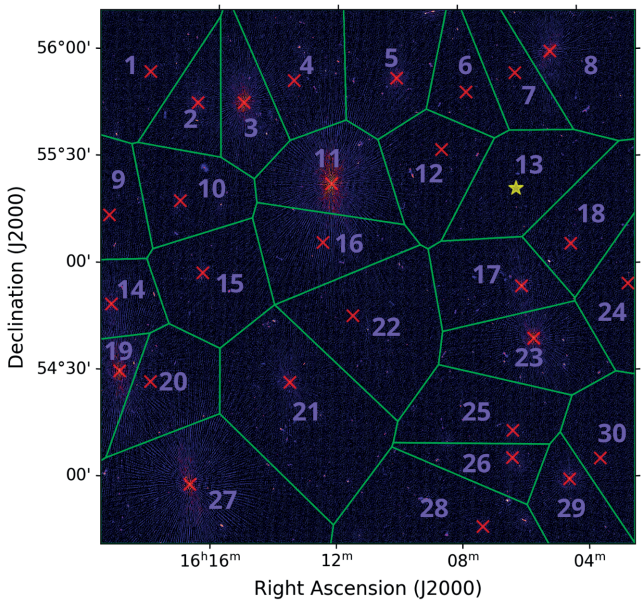


Figure 4.14: Final facet-layout on top of the 1.2″ DI image, created by applying DI corrections on data from one of our observations. The yellow star indicates the position of the primary in-field calibrator, the red crosses correspond to the DD calibrators, and the green boundaries show the Voronoi tessellation corresponding to these calibrators. The numbers are used for reference throughout this paper.

4.3.3.5. Refining Dutch calibration

Our calibration strategy is specifically designed to enhance the calibration solutions for international stations, by incorporating the phase-up of the Dutch core stations and excluding baselines shorter than 20,000λ (see Section 4.3.3.2). However, after applying DD calibration corrections, we observed strong artefacts from 18 of our 30 calibrators when we created facet images at 6″. This arises from bright sources outside the facet boundaries that affect the calibration of shorter baselines during DD self-calibration. Adding subtraction of sources near our calibrators before DD calibration, using for instance the 6″ models, would have been too expensive, as this brings an additional cost of ~800,000 CPU hrs. However, as we are using subtraction

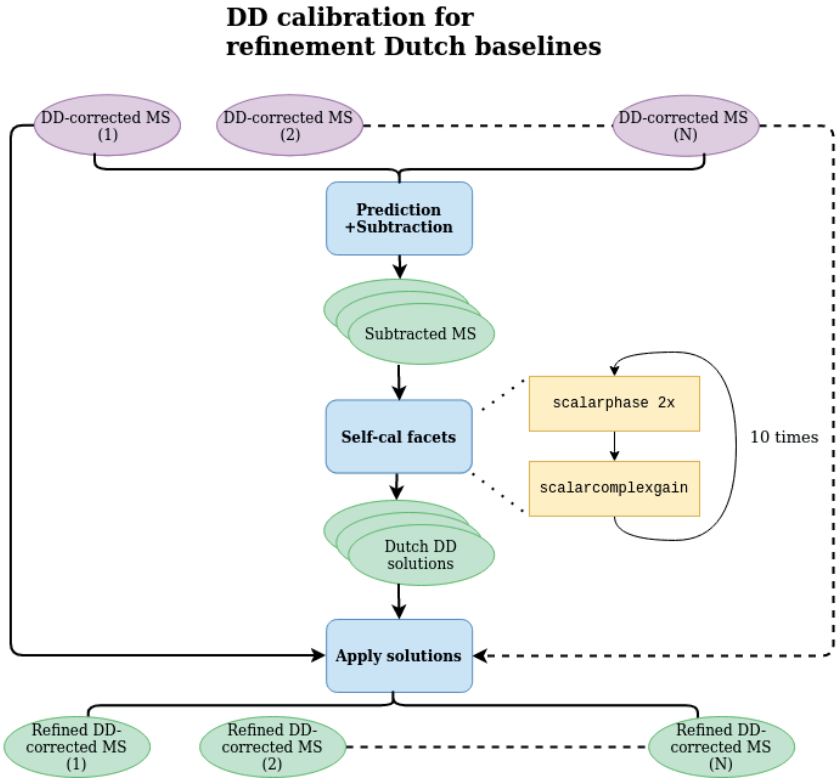


Figure 4.15: Workflow corresponding to the calibration steps explained in Section 4.3.3.5 for the general case with N observations. The workflow starts with the DD-corrected uv-data and ends with the refined DD-corrected uv-data for N different observations of the same field. These steps follow after the workflow in Figure 4.8. The prediction and subtraction steps are in more detail explained in Section 4.4.1. Purple ovals are input data, blue boxes are operations on the data, red boxes are data filters, yellow boxes are calibration steps, and green ovals are output data. Stacked ovals imply that there are output products for each observation. Dashed lines indicate the presence of numerous observations that can run in parallel. For a description of the calibration operations, we refer to Table 4.3.

around our facets before imaging (see Section 4.4.1), we could still further refine the calibration solutions for the shorter baselines, as we outline below. The workflow for refining the solutions corresponding to the Dutch LOFAR stations, discussed in this subsection, is also presented in the diagram in Figure 4.15 for the general case of N observations.

To suppress calibration issues introduced by bright sources beyond the facet boundaries, we subtracted sources from our visibility data that correspond to sources outside our facets, using image models at $1.2''$. These model images were produced

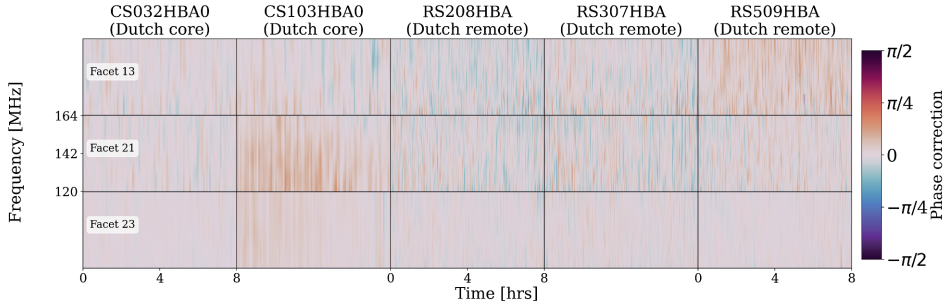


Figure 4.16: Merged phase calibration solution plots corresponding to the different facets (*rows*) and different stations, given by their station IDs (*columns*). These solutions are relative to the CS001HBA0 Dutch core station. The facets have the DD solutions from their corresponding calibrator, as depicted in Figure 4.14.

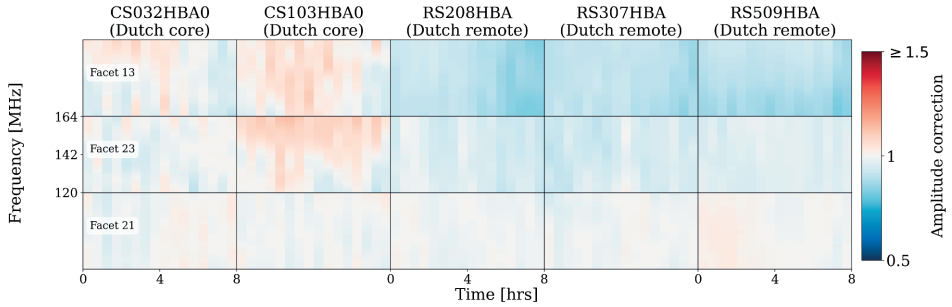


Figure 4.17: Merged amplitude calibration solution plots corresponding to the different facets (*rows*) and different stations, given by their station IDs (*columns*). The facets have the DD solutions from their corresponding calibrator, as depicted in Figure 4.14.

for each of our four observations with the merged DD solutions discussed in Section 4.3.3.2. As the subtraction of sources is part of our imaging procedure, we refer to Section 4.4.1 for further discussion about this process. After subtraction, we conducted extra rounds of self-calibration on the entire facet with only the Dutch core and remote stations, using a *uv*-cut of 750λ (corresponding to a LAS of ~$275''$). Since we calibrate at $6''$ and our facets are smaller than the entire wide-field image, we are allowed to average our data to 20 sec and 244 kHz and apply on top of this additional averaging based on the facet size. With the resulting *uv*-data sets we found 10 rounds of self-calibration to be sufficient to reach convergence. After experimenting with different settings, we found for each self-calibration cycle the following steps to work best (see Table 4.3 for the operation description):

1. `scalarphase` **I**: We start by solving for 'fast' phase variations for the Dutch remote stations, by applying `scalarphase` corrections with a solution interval

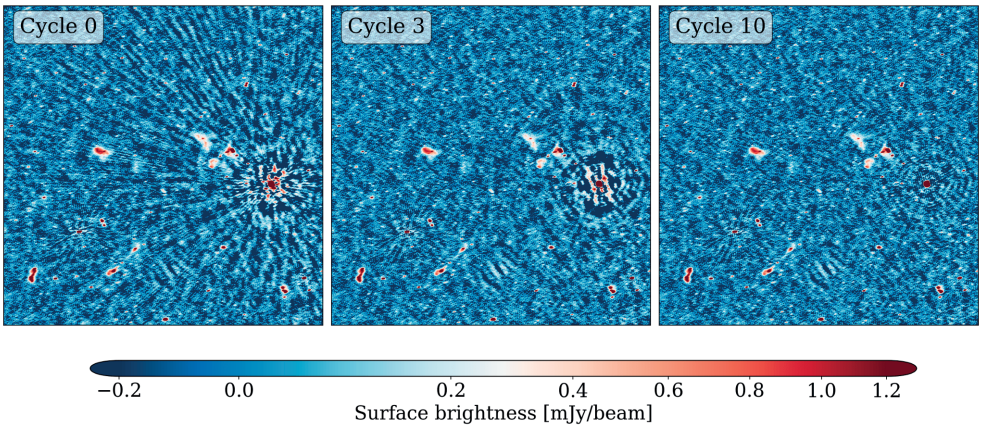


Figure 4.18: Example of the image improvements after applying extra self-calibration with only Dutch core and remote stations at 6″ on facet 17 (see Figure 4.14) in a 0.4°×0.4° cutout image. Cycle 0 has only DI solutions applied, cycle 3 had the first round of `scalarcomplexgain` solve where both phases and amplitudes are calibrated, while cycle 10 shows the result after the final round of self-calibration.

of 1 min and a frequency smoothness kernel of 10 MHz for the Dutch remote stations. This is because the remote stations have faster phase variations.

2. **`scalarphase` II**: To solve for the slower varying phases for the Dutch core stations, we then employed `scalarphase` corrections with solution intervals of 5 min and a larger frequency smoothness kernel of 20 MHz.

3. `scalarcomplexgain`: While the first two self-calibration cycles only have `scalarphase` corrections, we introduce in the third self-calibration also `scalarcomplexgain` calibration to correct for scalar amplitude effects as well. This step solves with a solution interval of 30 min and a frequency smoothness kernel of 15 MHz.

The final merged solutions from three different facets for different Dutch stations across the Netherlands are shown in Figures 4.16 and 4.17. Figure 4.18 demonstrates for 1 facet the significant image improvements at 6″ resolutions. We merged the resulting Dutch core and remote solutions back into our full merged solutions for all stations that we obtained after DD calibration (see Section 4.3.3.2). To compare how these new solutions improve the image quality at 3 different resolutions, we show for the same facet in Figure 4.19 the image quality.

This additional step was implemented after we already completed the full imag-

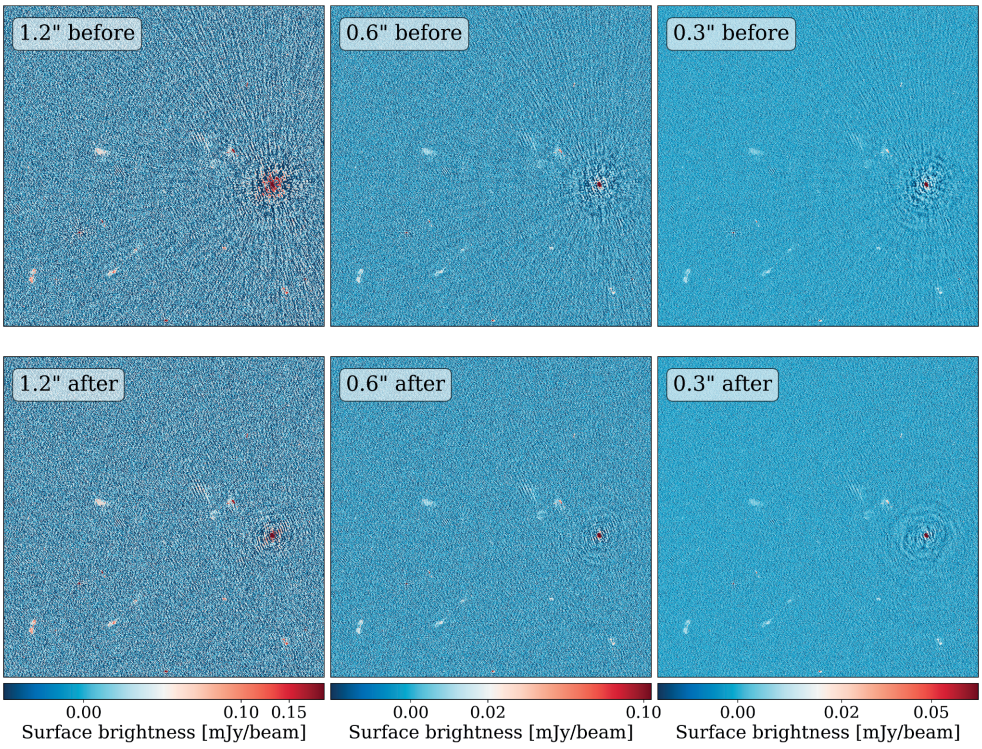


Figure 4.19: Image quality changes across four resolutions after applying extra self-calibration with only Dutch core and remote stations on the calibrator from facet 17 (see Figure 4.14) in a 0.4°×0.4° cutout image. The top row displays part of the facet images before adding the extra Dutch solutions and the bottom row displays the results after adding the extra Dutch solutions.

ing of all facets at all resolutions (see Section 6.4.5). Since the imaging procedure is computationally expensive and increases with the image size and the size of our *uv*-data, we opted to only use these solution refinements for the imaging at 1.2″ resolution and just for the 5 facets that we found visually to be most affected by the Dutch solution issues at 0.6″ and 0.3″ resolutions (such as the facet from Figure 4.19). This is also motivated by the fact that the Dutch core and remote stations have the most short-baselines, which implies that the calibration issues for those stations reduce towards higher resolutions.

	0.3″			0.6″			1.2″		
Facet	**Resolution** (arcsec^2)	**RMS** (μJy beam^{-1})	ρ (degree^{-2})	**Resolution** (arcsec^2)	**RMS** (μJy beam^{-1})	ρ (degree^{-2})	**Resolution** (arcsec^2)	**RMS** (μJy beam^{-1})	ρ (degree^{-2})
1	0.36×0.45	29	668	0.61×0.67	37	797	1.25×2.74	60	668
2	0.35×0.42	21	1278	0.58×0.63	29	1388	1.09×2.03	56	943
3	0.34×0.40	20	1824	0.56×0.61	27	1824	1.05×1.71	50	1450
4	0.34×0.41	20	1491	0.57×0.62	28	1336	1.06×1.81	49	1052
5	0.34×0.41	19	1594	0.57×0.62	26	1403	1.05×1.78	51	989
6	0.34×0.41	19	1596	0.57×0.62	27	1500	1.06×1.87	47	1093
7	0.35×0.42	22	928	0.58×0.63	30	877	1.06×1.97	54	577
8	0.36×0.43	28	718	0.62×0.65	36	778	1.15×2.51	62	692
9	0.35×0.43	22	875	0.58×0.64	30	967	1.08×2.11	55	611
10	0.34×0.41	18	1615	0.56×0.61	25	1456	1.04×1.68	47	1023
11	0.33×0.39	15	2118	0.55×0.59	22	1764	1.02×1.45	48	900
12	0.33×0.39	16	2281	0.55×0.60	22	1811	1.02×1.52	41	1057
13	0.34×0.40	17	1843	0.56×0.61	25	1740	1.03×1.63	44	1214
14	0.34×0.42	21	1435	0.57×0.62	28	1373	1.05×1.78	56	1071
15	0.34×0.40	16	1860	0.55×0.60	22	1571	1.02×1.53	41	1062
16	0.33×0.38	14	2071	0.54×0.59	21	1678	1.00×1.40	40	912
17	0.33×0.39	15	1983	0.55×0.59	21	1721	1.01×1.48	39	1132
18	0.34×0.41	20	1269	0.57×0.62	27	1218	1.04×1.78	49	860
19	0.35×0.43	25	1115	0.59×0.64	33	1158	1.09×2.14	58	974
20	0.35×0.42	21	1203	0.58×0.63	28	1239	1.06×1.92	49	848
21	0.34×0.40	17	1942	0.55×0.60	24	1733	1.02×1.54	43	1152
22	0.33×0.38	15	2315	0.55×0.59	21	1966	1.00×1.41	41	1145
23	0.33×0.40	17	1826	0.56×0.61	24	1756	1.03×1.60	44	1178
24	0.35×0.42	21	863	0.58×0.63	30	1052	1.05×1.94	52	821
25	0.34×0.40	19	1573	0.56×0.61	26	1433	1.03×1.63	51	948
26	0.35×0.42	20	1365	0.57×0.63	27	1455	1.07×2.01	54	997
27	0.35×0.43	28	712	0.60×0.64	37	785	1.12×2.32	63	630
28	0.35×0.42	24	1044	0.58×0.63	32	1121	1.09×2.11	58	742
29	0.36×0.45	28	679	0.56×0.61	38	658	1.28×2.82	71	679
30	0.35×0.44	28	734	0.60×0.66	36	824	1.14×2.50	61	657

Table 4.4: Resolutions, RMS noises, and source densities of each individual facet. The facet numbers correspond to the numbers in Figure 4.14. The source density (ρ) is based on the catalogues after cleaning our source detections, as discussed in Section 4.5.

4.4. Wide-field imaging

Following the completion of the calibration procedures and collecting our merged phase and amplitude solutions, we performed wide-field imaging to obtain our final image products. In this section, we discuss the imaging method, show parts of our imaging results, and discuss the computational costs of this step in the process compared to calibration.

4.4.1. Method

We employed `WSClean` Version 3.3 to produce the wide-field images. This imager has, using the `wgridder` module (Arras et al., 2021; Ye et al., 2022), a facet-based imaging mode that enables wide-field imaging with solutions for different facets. While this option has proven to be fast and reliable for making large wide-field images corrected for DD effects (e.g. de Jong et al., 2022; Ye et al., 2024), with the large data volumes the computational demands for 0.3″-imaging are so high, at this resolution, it would take over four months to make at this resolution images up to 10^{10} pixels directly with four observations (see Section 4.4.3). For an 1.2″ resolution wide-field image, it only takes up to four days for an 8 hrs dataset. So, we decided to

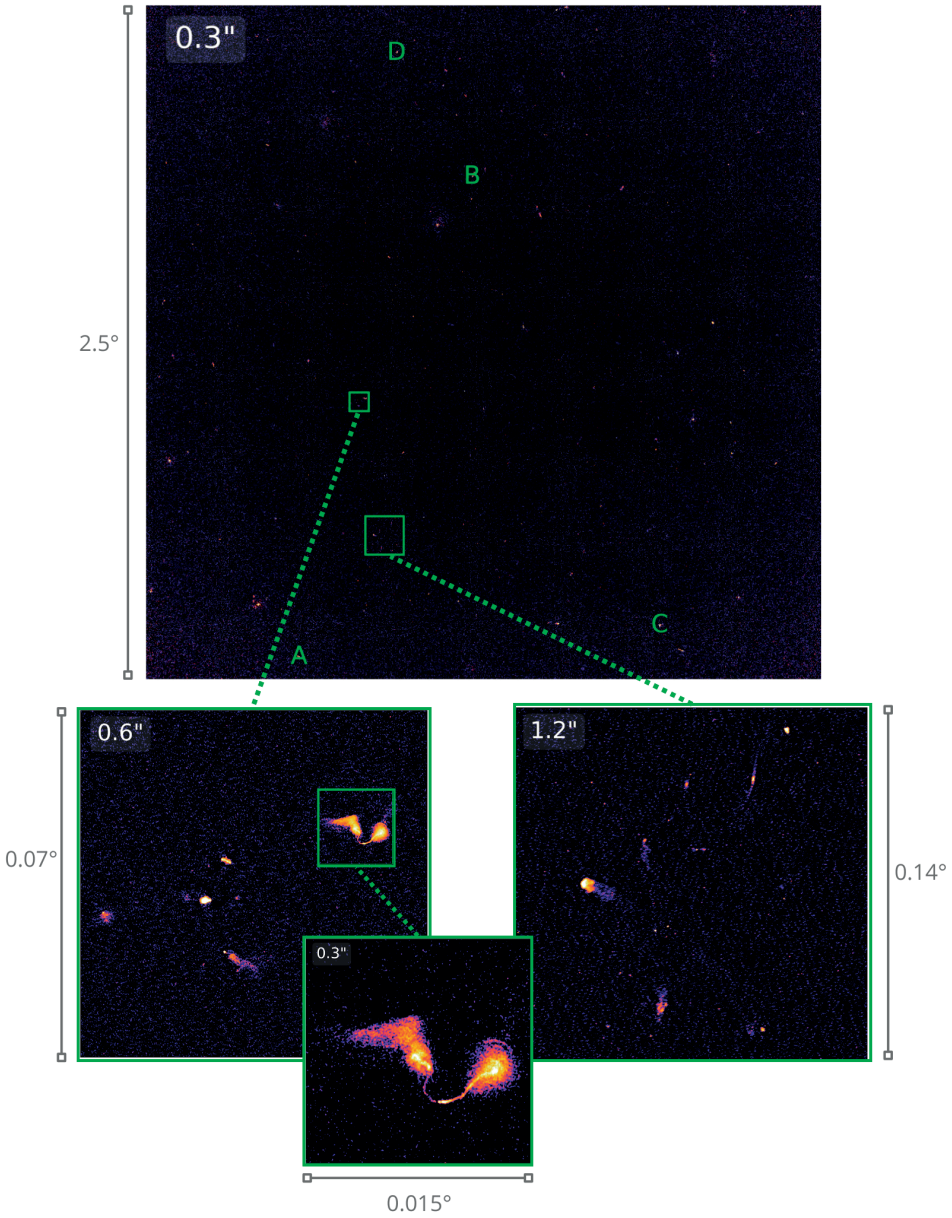


Figure 4.20: Our final 0.3″ wide-field image centred on RA=242.75° and DEC=54.95° with cutouts at 0.3″, 0.6″, and 0.6″ from selected areas. The letters correspond to the selection of sources in Figure 4.21.

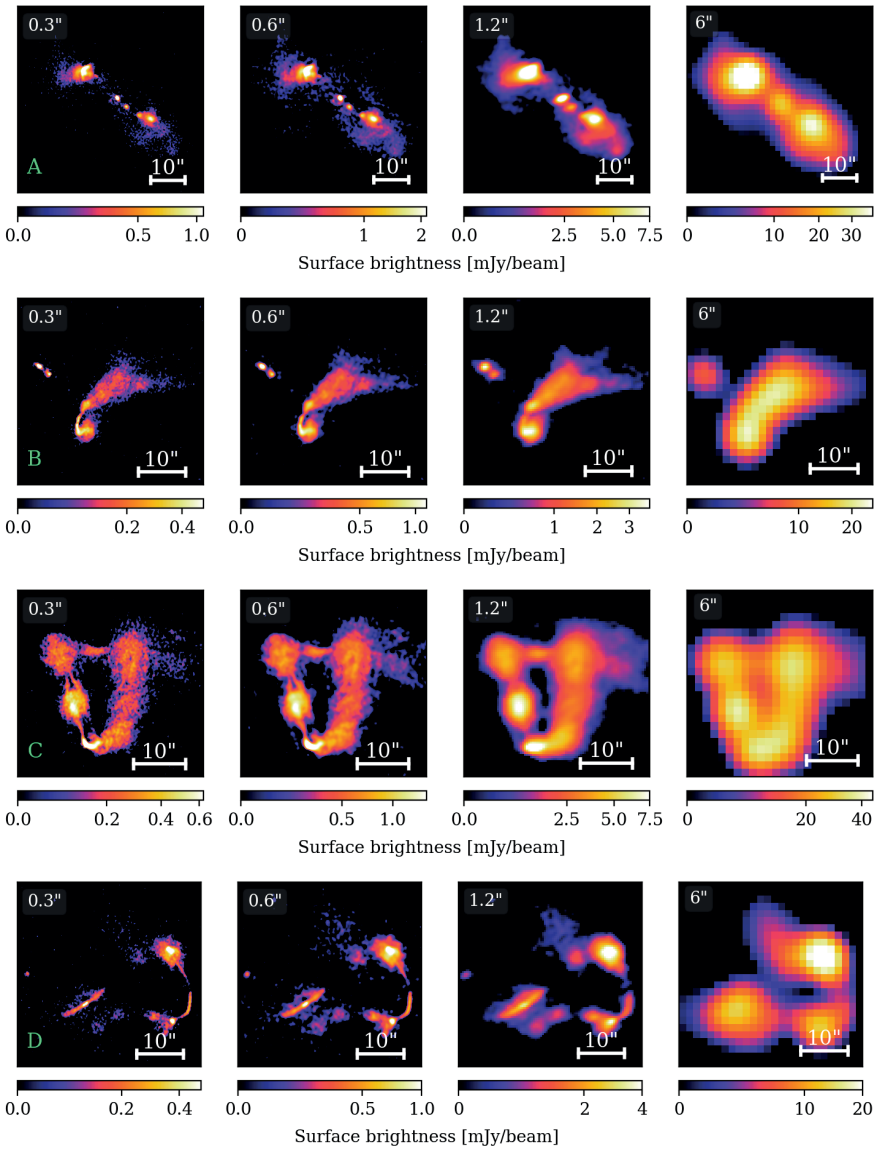


Figure 4.21: Different radio galaxies (*rows*) across resolution (*columns*) for a selection of cutouts in the wide-field images produced in this paper. The green letters correspond to the letters depicted in the wide-field image from Figure 4.20. The 0.3″, 0.6″, and 1.2″ images are produced by us, the 6″ images are from the wide-field image of ELAIS-N1 produced by Shimwell et al. (in prep.). The angular size scale is indicated in the right lower corner.

only utilise the facet-based imaging for making wide-field images at 1.2″ resolution for each of the four individual observations, as this gives us an image to assess the quality of our fully calibrated data across the entire field of view for each observation. Additionally, the model images from `WSClean` were essential for imaging our facets separately. This involved predicting and subtracting data outside each facet, a process we describe in detail below.

By making images of our facets separately, we allowed ourselves to average the visibilities without introducing smearing effects. The averaging factors in both time and frequency are determined by the facet size and vary from 3 to 7. This speeds up the imaging, compared to the original 1 sec (or 2 sec) and 12.21 kHz resolutions of the datasets before averaging. We note that the datasets that were originally averaged to 2 sec will have smaller time-averaging factors. To remove emission outside a facet, we first derived model visibilities corresponding to the sky outside each facet by utilising the model images from the 1.2″ resolution radio maps corresponding to each observation. To achieve this, we masked the facet in the model image and used this masked model image to predict visibilities with applied DD solutions in `WSClean`. This yields the model data visibilities that we then subtracted from the original DI corrected visibilities. As this process can be done in parallel over frequency blocks, we did the prediction and subtraction for smaller frequency sub-band for each of our observations, which helped us reduce the processing wall-clock time by a factor 16 (see Section 4.4.3). We note that this does not reduce the total CPU time. After phase-shifting the subtracted data to the centre of the facet, applying the solutions from the DD calibrator, and accounting for the beam at the facet's centre, we averaged the data for each observation before proceeding with the final imaging using `WSClean`.

We imaged each facet with all observations together, using a Briggs weighting of −1.5 (Briggs, 1995), a minimum *uv*-value of 80λ (corresponding to a LAS of ∼43′), pixel sizes of 0.1″, 0.2″, and 0.4″, and corresponding Gaussian tapers of 0.3″, 0.6″, and 1.2″. For efficient deep cleaning and to better recover extended diffuse emission, we apply 'auto' masking, multi-scale deconvolution, and an RMS box equal to 50 times the synthesised beam size (Cornwell, 2008; Offringa & Smirnov, 2017). `WSClean` ends by applying a final full primary beam correction to correctly account for the attenuation of the primary beam.

4.4.2. Facets and mosaics

Table 4.4 gives the resolutions, RMS noise, and source density of each individual facet. We reach a best RMS noise value of 14 μJy beam^{-1} for the 0.3″ facets, 21 μJy beam^{-1} for the 0.6″ facets, and 39 μJy beam^{-1} for the 1.2″ facets near the pointing centre (see Section 4.6.1 for a more detailed RMS noise analysis). This

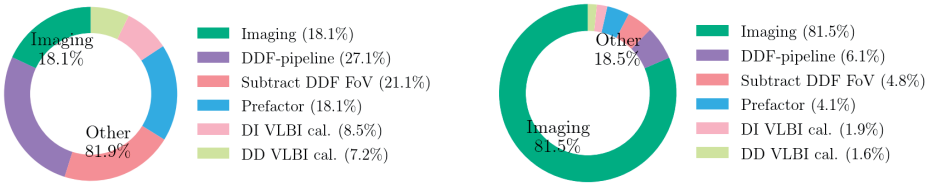


Figure 4.22: Pie plots depicting the fraction of CPU hrs of each of the major calibration and imaging steps for 1.2″-imaging (*upper panel*) and 0.3″-imaging (*lower panel*). The 1.2″ processing was done for one observation with the facet-mode from `WSClean` in about 7,000 CPU hrs, and the 0.3″-imaging was done for one observation using the predict-subtract method for one observation in about 139,000 CPU hrs. These numbers scale roughly linearly with the number of observations. The 'Subtract DDF FoV' includes the subtraction of sources outside the 2.5°×2.5° field of view from the last paragraph of Section 4.3.1. The 'DI VLBI calibration' includes the DI calibration from Section 4.3.2. The 'DD VLBI calibration' includes the DD calibration selection and self-calibration from Section 6.4. To highlight the computational costs for imaging compared to the other data reduction steps combined, we indicate in the figure 'Imaging' and 'Other' in the pie plots.

is about twice as deep as the sensitivities reported by Sweijen et al. (2022c) at 0.3″ and Ye et al. (2024) at 1.2″, who utilised data with four times less integration time. This aligns with the expected behaviour from the radiometer equation, which states that sensitivity improves as the square root of the integration time (Kraus, 1966). We find our best resolutions near the pointing centre, with resolutions of 0.33″×0.38″ for the 0.3″ facets, 0.54″×0.59″ for the 0.6″ facets, 1.00″×1.40″ for the 1.2″ facets. The stronger elongation of the synthesised beam for the 1.2″ target resolution is due to the sparsity of LOFAR stations between 80 and 180 km (see Figure 4.2).

To make wide-field images, we convert the individual resolutions from the facets to one common resolution. Due to issues with one of our computing nodes, we lost 5 of our residual and model images of the 0.3″ facets. We were therefore only able to convolve our 0.3″ map to a common resolution equal to the facet with the lowest resolution, using CASA `imsmooth` (The CASA Team et al., 2022). This gives us a resolution of 0.36″×0.45″. Having all model and residual images available for the other resolutions, we were able to restore these maps with `WSClean` to a common resolution of 0.58″×0.62″ and 1.0″×1.5″.

After mosaicing the individual facets, we have our wide-field images for all three resolutions with image sizes of 90,000×90,000, 45,000×45,000, and 22,500×22,500 pixels for the 0.36″×0.45″, 0.58″×0.62″ and 1.0″×1.5″ respectively. In Figure 4.20, we present one of our wide-field images with a few cutouts of areas at different resolutions. In order to assess visually the quality and level of detail across various resolutions, we showcase selected cutouts from our radio maps at resolutions of

$0.36'' \times 0.45''$, $0.58'' \times 0.62''$ and $1.0'' \times 1.5''$, and $6''$ in Figure 4.21, where for the $6''$ counterparts we utilised the deep wide-field image recently created by Shimwell et al. (in prep.), who used more than 500 hrs of LOFAR data. These selected cutouts reveal the structural details at the higher resolutions, notably evident in the lobes of radio galaxies, while the lower resolutions highlight the diffuse emission from these same sources. during the rest of this paper, we use our individual facet images, as these have the best-fitted resolutions and depths. We therefore continue to refer to these images by $0.3''$, $0.6''$, and $1.2''$ resolutions.

4.4.3. Computing costs

Making wide-field images in the order of 10^9 to 10^{10} pixels requires significant computing resources. For processing our data we utilised AMD EPYC 7551 and AMD EPYC 7702P processor nodes with each 60 cores and with 0.5 TB and 1 TB RAM. Although the predict-subtract method before imaging of the individual facets (see Section 4.4.1) incurs a large computational cost, accounting for approximately 76% of the total imaging costs, we managed to reduce the wall-clock time by a factor 16 through additional parallelisation by splitting our total frequency bandwidth in smaller blocks and performing the prediction and subtraction step for each block separately. The final imaging costs for four observations total 550,000 CPU hrs, which brings us to a total of about 680,000 CPU hrs for full data processing including calibration. With the large data volumes, we found a linear relationship between data volume and computing costs. Using this linear relationship, we find an improvement of about a factor 2 compared to Sweijen et al. (2022c), who worked with only a sixth of the data size we processed (taking into account that they pre-averaged their 8 hrs dataset by a factor 2 in time). This speedup is primarily due to a combination of software enhancements in `WSClean` and the optimisation of our software containers for the appropriate hardware. While we observe an improvement in processing speed, it is notable that when mapping the sky at the highest resolution the imaging expenses account for 81.5% of our overall processing costs, as depicted in the lower panel of Figure 4.22, which is slightly higher compared to the 76% reported by Sweijen et al. (2022c). This difference could be due to a combination of different numbers of facets and software improvements that have affected parts of the pipeline differently than other parts. That computational demands for $0.3''$-imaging are predominantly driven by imaging, highlights that full data reduction speedups need development for this step. For creating the $0.6''$ resolution wide-field image, we averaged our data to 2 sec and 24.42 kHz, before imaging and changed the final imaging parameters. We could similarly for the facets at $1.2''$ resolution average again by a factor two (4 sec and 48.84 kHz) compared to the $0.6''$ resolution. The averaging makes only the final imaging after the prediction and subtraction 4

(0.6″) or 16 (1.2″) times faster, compared to 0.3″-imaging.

As a part of the prediction and subtraction step for the 0.3″ and 0.6″ resolution imaging, we made wide-field model images at 1.2″ for each of our four observations. The computing costs for this step, using the `WSClean` facet-based imaging mode, required 7,000 CPU hrs per observation. This is almost two times faster than what Ye et al. (2024) reported for wide-field imaging. The improvement is again due to a combination of recent software improvements and the different computing nodes they used for imaging. In the upper panel of Figure 4.22 we display the imaging costs for 1.2″-imaging. Comparing this with the plot corresponding to 0.3″-imaging from the right panel, it is evident that 1.2″-imaging with the facet-mode from `WSClean` significantly reduces the weight of the imaging step on the complete data reduction workflow. The reduced computational costs, relative to sub-arcsecond imaging, make imaging at 1.2″ resolution an interesting intermediate resolution for specific science goals or surveys, as was also highlighted by Ye et al. (2024). The scientific benefits of the different resolutions are discussed further in Section 4.6.5.

4.5. Cataloguing

We constructed catalogues with radio sources for all three of our image resolutions (0.3″, 0.6″, and 1.2″) by employing `PyBDSF`[13] on our individual facets (Mohan & Rafferty, 2015). All parameter settings that we modified from the default settings are displayed in Table 4.5. The `rms_box` sets the sliding box parameters for calculating the RMS and mean flux density per beam over the entire image. With the `rms_box_bright` parameter we enable `PyBDSF` to more effectively increase the noise in regions of artefacts, by using a smaller box around brighter sources. The `group_tol` parameter groups Gaussian components fitted by `PyBDSF`. We opted to use the value 10 for this parameter, as this value has been often adopted for source detections at the same or similar frequencies (e.g. Williams et al., 2019; Ocran et al., 2020; Sabater et al., 2021; Ye et al., 2024). `PyBDSF` detected with these settings 24251 objects for our 0.3″ resolution radio map, 14099 objects for our 0.6″ resolution radio map, and 10229 objects for our 1.2″ resolution radio map.

Although we set the `group_tol` higher than the default value to enhance the association of components, we find by eye unassociated detections by `PyBDSF` that are part of the same physical object. We therefore decided to apply additional source association. Automated source association approaches at 6″, using for instance a convolutional neural network, have shown to provide a similar accuracy to what visual inspection by astronomers would obtain (Mostert et al., 2022). Since we have radio maps at higher resolutions, these models require extra training and

[13] `https://pybdsf.readthedocs.io`

Parameter	Value
rms_box	(120, 15)
rms_box_bright	(40, 10)
adaptive_rms_box	True
atrous_do	True
group_tol	10.0

Table 4.5: PyBDSF settings modified from the default values. We refer to the documentation for a full description of these parameters.

testing, introducing additional challenges that are beyond the scope of this paper. Fortunately, our field of view is confined to 2.5°×2.5°, limiting the number of large resolved sources. We therefore decided to visually inspect all clusters of sources within a distance of 50″ from each other and associate components that are likely part of the same source. For guidance, we used images at each of the four different resolutions, which allowed us, in case of doubt, to make comparisons during the association of extended objects. After this visual inspection, we were left with 22804, 13119, and 9577 sources at 0.3″, 0.6″, and 1.2″ respectively.

Following component associations, we employed the `shapely`[14] Python package to automatically find the integrated flux densities of the visually associated components. With this package, we automatically drew polygons enveloping all islands from each source and summed the pixels within to obtain the integrated flux density. For the final central source position, we applied, similar to `PyBDSF`, moment analysis (Hu, 1962), which calculates the weighted mean of the brightness distribution, thus providing a robust measure of the centroid of extended sources. We carried out an additional visual inspection of the sources that we identified by eye as having a bad polygon fitting. Consequently, this led us to manually calculate the integrated flux density of 67 sources. To illustrate our method, we show in Figure 4.23 two examples of extended sources that have been subjected to our fitting procedure.

To ensure the reliability of the sources in our final catalogue, we decided to use a S/N threshold at 5σ, which implies that we reject all sources with a peak intensity below 5 times the local RMS. This is reported by `PyBDSF` in the `Isl_rms` column. After removing these sources we had 13058, 10241, 6997 sources at respectively 0.3″, 0.6″, and 1.2″. The remaining sources were cross-matched with the catalogue from the 6″ ELAIS-N1 LOFAR HBA map by Shimwell et al. (in prep.), where we reject sources from our catalogue that do not have a cross-match within 6″. Their map has a minimal sensitivity of 11 μJy beam^{-1} and is, therefore, deeper than our images. This additional selection step ensured the reliability of our catalogue

[14] `https://shapely.readthedocs.io`

content and left us with final source counts of 9203, 8567, 5872 sources at 0.3″, 0.6″, and 1.2″ respectively.

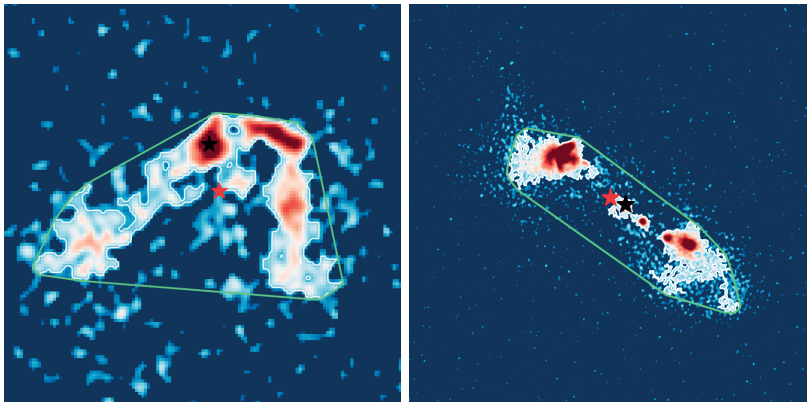


Figure 4.23: Examples of resolved sources where the integrated flux density and source position were obtained by using an enveloping polygon (light green) circumventing all source islands. Similar to `PyBDSF`, we require each polygon to have a peak intensity exceeding 5 times the local RMS. The final source position, determined using moment analysis, is marked by the red star. The black star is at the position of the peak intensity of the source within the polygon boundary.

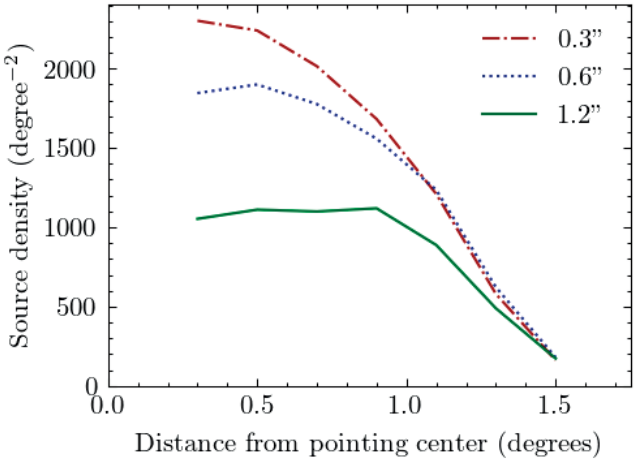


Figure 4.24: Source density as a function of distance from the pointing centre in degrees for all three resolutions. This figure was constructed by evaluating the median source density in bins of 0.2°.

Going about two times deeper than Sweijen et al. (2022c) and Ye et al. (2024), we find respectively 4 and 2.5 times more objects at the same resolutions. To

illustrate the source distribution across our different facets and resolutions, we give in Table 4.4 the source densities across our 30 facets and plot in Figure 4.24 the source density as a function of distance from the pointing centre. In Section 4.6.5 we further discuss the different sources detected at different resolutions.

4.6. Discussion

We have created the deepest (sub-)arcsecond wide-field images of ELAIS-N1 at 140MHz, accomplished by processing together four 8 hrs observations including all available LOFAR stations. In this section, we do additional analysis of the image and source detection quality.

4.6.1. RMS noise

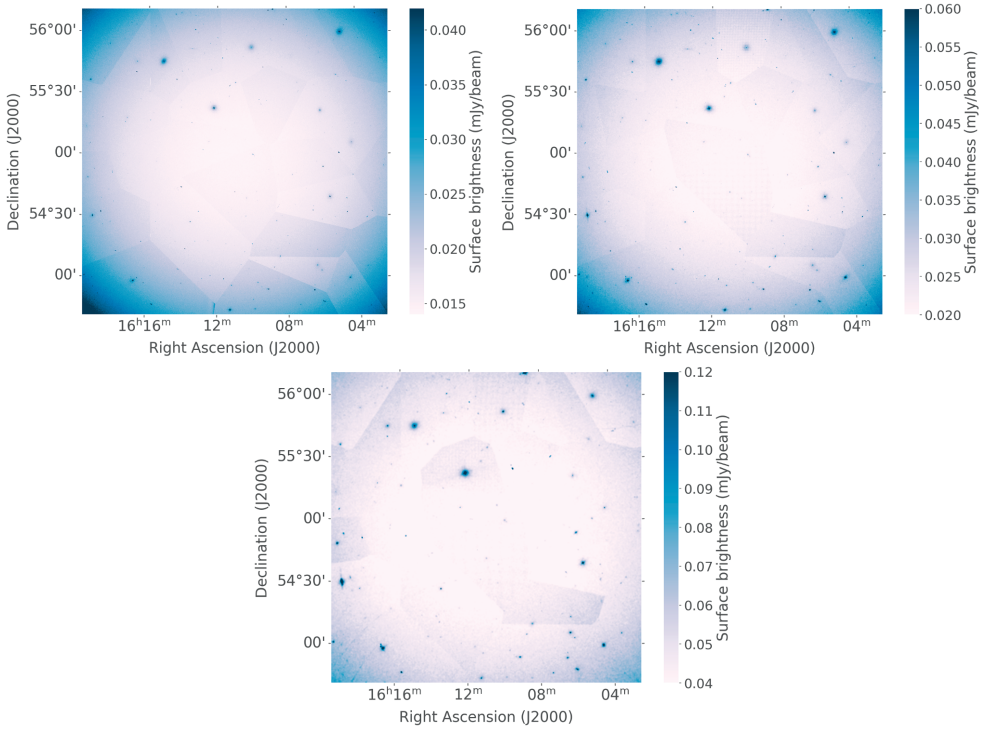


Figure 4.25: RMS maps for different resolutions. *Upper left:* 0.3″. *Upper right:* 0.6″. *Lower:* 1.2″. These maps are made with `PyBDSF` and scaled between one and three times the best RMS noise in the image (see Figure 4.26).

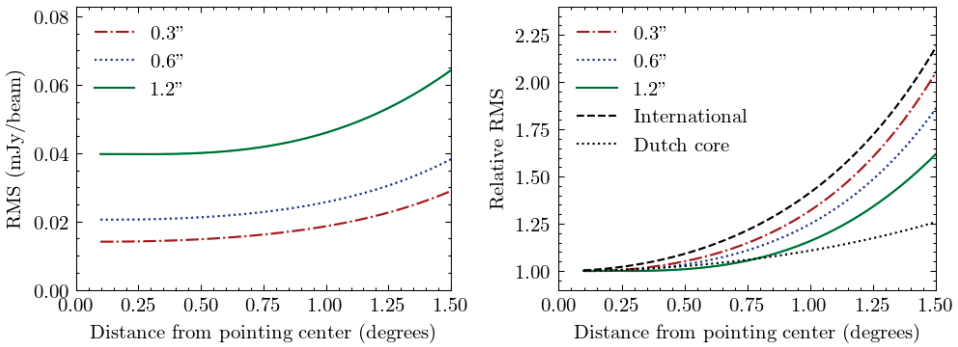


Figure 4.26: Median RMS noise as a function of distance from the pointing centre in degrees for all three of our resolutions. The median RMS noise values are calculated in bins of 0.06° and fitted by a second-order polynomial. *Left panel:* Absolute RMS. *Right panel:* RMS divided by the RMS at the pointing centre, where we added the inverse primary beam intensity for an international and Dutch core station.

In Section 4.4.2 we touched upon the RMS noise across different facets. This variation is also notable in the RMS noise map from `PyBDSF`, as shown in Figure 4.25. The higher RMS noise values at lower resolutions are due to a combination of tapering and Briggs weighting, which we set for all three resolutions at -1.5 (see Section 4.4.1). Given that we have not optimised the Briggs weighting for all resolutions, this may particularly negatively affect the RMS noise at $1.2''$ resolution, as this resolution is most susceptible to the *uv* sampling gaps between 80 and 180 km (see Figure 4.2). However, optimising the Briggs weighting parameter is with the current computing costs too expensive (see Section 4.4.3). The RMS noise offsets between facets in Figure 4.25 are attributed to solution quality differences across the DD calibrators, with each facet having its own set of DD solutions. The peaks in the local RMS noise at the positions of bright sources are due to remaining DDEs that are not completely removed around the brightest sources. The sensitivity variation due to the attenuation from the primary beam is also a key contributor to the RMS noise increases towards the edge of the field. This is demonstrated with the median RMS noise as a function of distance in the left panel of Figure 4.26.

In the right panel of Figure 4.26 we compare the shape of the RMS curves as a function of distance from the pointing centre for the $0.3''$, $0.6''$, and $1.2''$ resolution. The larger steepness of the relative RMS noise for higher resolution reflects the primary beams of the stations used at that particular resolution. This explains why the source densities converge for the three resolutions at the edges of the field, as shown in Figure 4.24 and Table 4.4. In the right panel of Figure 4.26 we plot for comparison also the inverse primary beam intensity ($I_{\rm P}$) for an international and a Dutch core station as a function of distance (θ) from the pointing centre, using a

simple Gaussian model given by

$$I_{\rm P} = \exp\left(-4\ln 2\frac{\theta^2}{\rm FWHM^2}\right),$$

where the FWHM is the full width at half maximum of the synthesised beam and determined by

$$\rm FWHM = \alpha \cdot \frac{\lambda}{D}.$$

In this formula λ represents the wavelength (corresponding to 140 MHz) and D signifies the diameter of the stations, where we used a diameter of 30.75 meters for a Dutch core station, and 56.5 for the international station (van Haarlem et al., 2013). The value for α varies based on the station layout and additional tapering, being 1.02 for a perfect circular aperture (Napier, 1999), and 1.3 for LOFAR.[15] As anticipated, we observe that the relative RMS noise across all resolutions falls between the primary beam intensities of the international and Dutch core stations.

4.6.2. Smearing

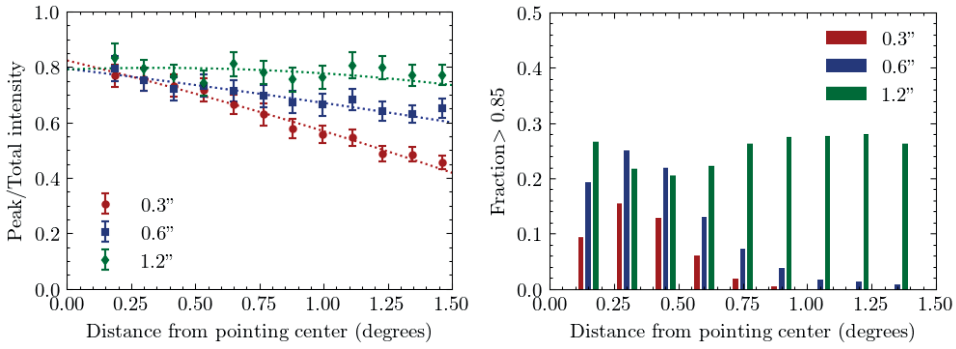


Figure 4.27: Smearing across distance from the pointing centre. *Left panel:* Ratio of peak over integrated flux density as a function of the distance from the pointing centre for the 3 different resolutions considered in this paper. The plot is made by taking the median ratio for several distance bins. The error bars are based on the errors on the peak and integrated flux densities measured by `PyBDSF`. *Right panel:* Fraction of sources with peak over integrated flux densities above 0.85 for different distance bins. To enhance the reliability of our source sample, we considered for both panels only sources fitted by a single Gaussian by `PyBDSF` and a peak intensity at least 15 times above the local RMS noise.

Bandwidth and time smearing are important effects that degrade the quality of

[15] `https://science.astron.nl/telescopes/lofar/lofar-system-overview/observing-modes/lofar-imaging-capabilities-and-sensitivity`

our radio maps and cannot be corrected by calibration. These effects cause sources to appear smeared or elongated in the radial direction in the case of bandwidth smearing and tangentially in the case of time smearing with respect to the pointing centre. This makes accounting for smearing effects essential for accurately measuring source sizes, morphologies, and peak intensities. Since smearing becomes more pronounced closer to the edge of the field of view, it contributes to the decrease in source detections in Figure 4.24.

An effective measure for assessing smearing levels is to compare the peak intensity with the integrated flux density of unresolved sources across the field. This is because smearing affects the peak intensity much more than the integrated flux density, as the integrated flux density only reduces when flux disappears below the detection threshold, while the peak intensity always decreases. For an ideal, unsmeared point source, the ratio of the peak intensity over the integrated flux density should be one. In the left panel of Figure 4.27, we illustrate the variation of this ratio with distance from the pointing centre, by selecting sources with peak intensities at least 15 times above their local RMS and fitted by a single Gaussian, across our three resolution images. The figure demonstrates a noticeable reduction in this ratio towards the edge of our field of view, most pronounced at the highest resolution. These declining trends are due to the inevitable effects of smearing, which in our case are intensified by the fact that half of our observations were stored in the LTA with a factor of two extra time averaging. Although we find the peak over integrated flux densities for some sources to be close to one near the pointing centre, the sources with much lower ratios push the median peak over integrated flux density down to below 0.8 in our distance bins. Part of the reason why our ratios are not at unity is due to our source selection where we are plotting all sources that are fitted by `PyBDSF` by a single Gaussian, while a source might still be a partly resolved source. Therefore, to quantify the number of sources with a peak over integrated flux density closer to 1, we show in the right panel of Figure 4.27 the fraction of sources with a peak over integrated flux density above 0.85. Similar to the left panel, we find only the $0.3''$ and $0.6''$ to have a declining trend from the pointing centre towards the edge of the field, which implies that smearing is negligible at $1.2''$.

4.6.3. Astrometry

We used an in-field calibrator source in our sky model of which the position was known with milliarcsecond-level accuracy (see Section 4.3.2.2). To evaluate the quality of the final astrometry from our radio maps, we conduct a comparison between our catalogue and the $6''$ catalogue from the LOFAR deep field DR1 (Kondapally et al., 2021; Sabater et al., 2021). The wide-field image behind the $6''$ catalogue

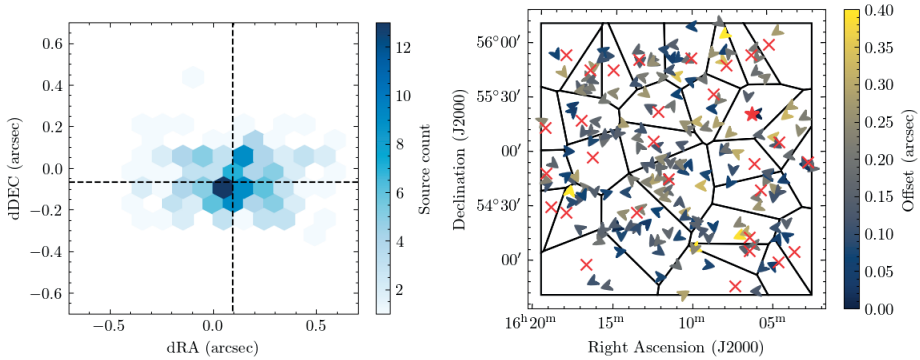


Figure 4.28: Position offsets between our 0.3″ radio source detections and optical counterparts according to the 6″ catalogue from Kondapally et al. (2021), where we used selection filters based on the accuracy of the associations at 6″ as described in the text. *Left panel:* Two-dimensional hex-bin histogram with the RA/DEC offsets (dRA and dDEC) between the 0.3″ catalogue and the optical counterparts. The median offsets are given by the black dashed line at dRA=0.094 ± 0.093″ and dDEC=-0.067 ± 0.064″. *Right panel:* Position offsets of the selected 192 sources at 0.3″ resolution, where the colour bar corresponds to their absolute offsets and the arrow direction to their directional offset. We added the positions of the DD calibrator with red crosses and the in-field calibrator with a red star. Facets are given by black contours.

has a sensitivity up to 20μJy beam^{-1}. The source detections have been associated with sources across multiple wavelengths, enabling an astrometric reference frame. The multi-wavelength data includes the Hyper-Suprime-Cam Subaru Strategic Program (HSC-SSP) survey (Aihara et al., 2018), Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Kaiser et al., 2010), the UK Infrared Deep Sky Survey Deep Extra-galactic Survey (UKIDSS-DXS Lawrence et al., 2007), the Spitzer Adaptation of the Red-sequence Cluster Survey (SpARCS; Wilson et al., 2009; Muzzin et al., 2009), the Space Infrared Telescope Facility (SIRTF) Wide-Area Infrared Extragalactic Survey (SWIRE; Lonsdale et al., 2003) and the Spitzer Extragalactic Representative Volume Survey (SERVS; Mauduit et al., 2012). The multi-wavelength catalogues were generated by combining imaging datasets from all of these surveys across optical-infrared. The astrometry for the optical datasets used was calibrated to Two-Micron All Sky Survey (2MASS Skrutskie et al., 2006). Kondapally et al. (2021) compared the astrometry between their generated catalogues to publicly available catalogues, finding typical offsets of around 0.1″ to 0.2″.

We first associate for each of our sources at 0.3″ a nearest radio counterpart from the 6″ catalogues from Kondapally et al. (2021), where we allowed a maximum distance of 6″. Since point sources are most accurately cross-matched, we only

select sources fitted by PyBDSF with a single Gaussian with major and minor axes less than 1.25 the size of the synthesised beam. To ensure that we cross-match sources that are with high certainty detected in both radio observations, we select from both catalogues sources with peak intensities 15 times larger than the local RMS. As our goal is to determine the astrometry using the optical cross-matches from Kondapally et al. (2021), we use flags in their catalogue to exclude sources in regions around bright stars, as the positions of the optical counterparts of our radio galaxies might be affected due to close proximity to these bright stars. We also use a threshold on the likelihood ratio (LR), as determined by Kondapally et al. (2021). This value indicates the ratio of the probability that a galaxy is a correct cross-match against being an incorrect cross-match. The LR is an often used statistical value to assess the quality of the counterpart cross-matching (e.g. de Ruiter et al., 1977; Sutherland & Saunders, 1992; Smith et al., 2011; McAlpine et al., 2012). We opt for selecting the top 30% of sources with the highest LR scores, as this leaves us with a sufficient number of sources (192) with the best optical-radio cross-matches. With this sample, we find median offsets of dRA$=0.094\pm0.093''$ and dDEC$=-0.067\pm0.064''$ between our $0.3''$ radio positions and the optical positions selected from Kondapally et al. (2021) catalogue. These offsets are within the astrometric uncertainty of the optical positions reported by Kondapally et al. (2021).

The scatter on the positional offsets is illustrated in the left panel of Figure 4.28. A positional 'random' scatter around the median value is expected because, due to the possible complex morphological nature of radio sources, there is no guarantee that the brightest point of a radio source aligns exactly with the position of an optical host. However, calibration errors can lead to systematic offsets. Since each facet has its own calibration solutions (see Section 6.4), we assess systematic offsets in the right panel of Figure 4.28, by plotting the distribution of the selected radio-optical associations with absolute offsets and offset directions across our 30 facets. While most facets do not show any preferred positional offset direction, we only find for facets 17, 22, and 23 hints of a positional offset direction in the +RA direction (see Figure 4.14 for the corresponding facet numbers). However, noting that the absolute offsets for these facets are not significantly larger than what we find for other facets, we do not apply astrometric corrections. To conclude, our small astrometric offsets confirm the high accuracy of our astrometry as a result of calibrating the in-field calibrator against an accurate sky model. Despite the presence of hints of minor systematic offsets in a few facets, the accuracy is largely maintained during DD calibration.

Additionally, we conducted a similar astrometric analysis with our $0.6''$ and $1.2''$ data and found within the uncertainties the same accurate results. This consistency is expected because the same calibration solutions were applied for all

resolutions, with the exception of some extra facets at 1.2″ that received additional calibration for the Dutch core and remote stations (see Section 4.3.3.5). Nevertheless, we do not observe any impact from this additional Dutch calibration on the positional offsets because the calibration solutions for the international stations, which primarily determine the smallest angular scales, have remained unchanged.

4.6.4. Flux scale

Similar to the astrometry evaluation, we verified our flux density scale by using the catalogue from Kondapally et al. (2021) (see Section 4.6.3). For this purpose, we utilised only the radio source information from their catalogue. We selected in both catalogues sources that were fitted by `PyBDSF` with a single Gaussian, have a maximum position offset of 6″, and which exhibited a peak intensity at least 25 times greater than the local RMS noise. This brightness threshold ensures we select sources without any loss of flux density, considering the resolution difference. With the remaining 368 sources, we find a median flux density ratio of $\frac{S_6}{S_{0.3}} = 0.97 \pm 0.14$ between the cross-matches of the 6″ catalogue and the 0.3″ catalogue, demonstrating the consistency of our flux scale. Similar to astrometry, the accurate flux density scale stems from the sky model that we used for calibrating the primary in-field calibrator (see Section 4.3.2.2).

We also compared the flux scales across our three resolutions. After selecting again sources identified by `PyBDSF` by a single Gaussian and with peak intensities 25 times above the local RMS noise, we cross-matched the three catalogues using a maximum position offset of 1″. This yielded 215 sources. For these sources, we find flux density ratios of $\frac{S_{0.3}}{S_{0.6}} = 1.00 \pm 0.04$, $\frac{S_{0.6}}{S_{1.2}} = 0.99 \pm 0.08$, $\frac{S_{0.3}}{S_{1.2}} = 0.98 \pm 0.09$. The consistency of these ratios supports the robustness of our flux density scale across different resolutions. When we remove the brightness constraint but keep the 1″ position offset for sources fitted by single Gaussians, we find with the 3607 remaining sources the following flux density ratios $\frac{S_{0.3}}{S_{0.6}} = 0.97 \pm 0.17$, $\frac{S_{0.6}}{S_{1.2}} = 0.92 \pm 0.24$, $\frac{S_{0.3}}{S_{1.2}} = 0.89 \pm 0.28$. Despite the large uncertainties, these results align with our expectations that lower resolutions tend to capture more diffuse emission compared to higher resolutions. The resolution differences are discussed further in the next section.

4.6.5. Sensitivity versus resolution

Higher resolutions allow for accurate astrometry and detailed characterisation of compact and extended structures whilst also allowing for more precise optical and near-infrared identification of host galaxies. The enhanced ability to discriminate between sources is demonstrated in the lower panel of Figure 4.21, where at 6″ we

initially identify 3 Gaussians that appear to belong to the same source, while at 0.3″ resolution, we find this to actually be multiple sources with complex components. However, the detection of objects with low surface brightness is more challenging in high-resolution images and requires further image processing (e.g. smoothing), causing some extended sources, that are visible at lower resolutions, to not be trivially detectable at higher resolutions. For example, we observe that 22% of the sources in our 1.2″ catalogue do not have a counterpart at 0.3″, even though the 0.3″ map is deeper. This demonstrates the importance of considering a trade-off between resolution, sensitivity, and scientific objective.

To illustrate the above, we plot in the left panel of Figure 4.29 the number of sources as a function of integrated flux density across resolution. This figure demonstrates the effect of enhanced sensitivity at higher resolutions on the number of sources detected at lower flux densities. As most of our sources are below 1 mJy, we expect a large fraction of our detections to be part of the star-forming galaxy population (Best et al., 2023). Above ~0.25 mJy, we find the number of detections at 0.6″ resolution to be higher than the 0.3″ resolution detections. This is likely due to the fact that at these flux densities the S/N at 0.6″ is large enough to detect many sources and because the source population here contains many distant star-forming galaxies (with redshifts of approximately $z \sim 0.5 - 1.0$) that have typical sizes of a few tenths of an arcsecond. These sources are therefore more likely to be resolved (out) at 0.3″ compared to the 0.6″. Similarly, above ~0.5 mJy the 1.2″ resolution detects the most sources, which is because fewer sources are resolved at this resolution.

The 1.2″ resolution has the best surface brightness sensitivity of our three resolutions and is therefore the best to identify extended sources. This includes the population of LERGs, which are the most dominant radio source population in the LOFAR Deep Fields DR1 above ~1.5 mJy (Best et al., 2023). However, at higher redshifts, LERGs will be fainter and more compact and more often hosted by star-forming galaxies (Kondapally et al., 2022), leading to selection biases when being unable to separate radio jets from star formation (Mingo et al., 2022; de Jong et al., 2024). This underscores again the value of making wide-field images at multiple resolutions with the same calibrated data.

We assess in the right panel of Figure 4.29 the fraction of sources detected as a function of resolution by cross-matching our catalogues with the compact sources (fitted by a single Gaussian) from a deeper 6″ catalogue (Shimwell et al., in prep.). This figure shows that the fractions detected at 0.3″ and 0.6″ resolutions yield similar results, despite the lower sensitivity of the 0.6″ resolution map. The 1.2″ resolution has a low fraction of sources detected below 0.4 mJy, while above this flux density, all resolution images have about the same detection fraction and approach

completeness between 1 and 10 mJy. The decline in the detected fraction towards lower flux densities is explained by a combination of being less sensitive to detect low surface brightness sources at higher resolutions and the fact that the 6″ resolution map based on 500 hrs of LOFAR data is slightly deeper than our maps.

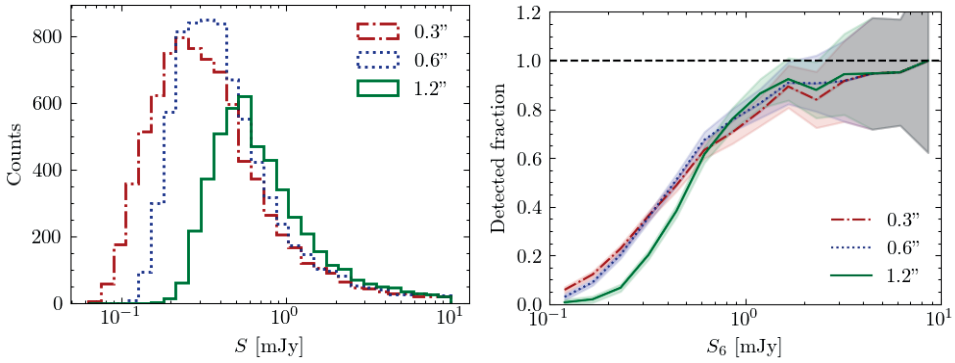


Figure 4.29: Source detection across resolution. *Left panel:* Source counts as a function of the integrated flux density for our 3 resolutions below 10 mJy. *Right panel:* Detected fraction between sources from our radio maps and the compact sources detected by Shimwell et al. (in prep.) at 6″. For the errors, we propagate Poisson uncertainties.

4.7. Summary and conclusion

We have presented the currently deepest wide-field image of ELAIS-N1 at about 0.3″ resolution and 140 MHz. This image has a field of view of 2.5°×2.5° with a sensitivity down to 14 μJy beam^{-1} at the pointing centre. This was achieved by implementing an improved DI and DD calibration strategy built upon the existing VLBI calibration and imaging strategies (Morabito et al., 2022a; Sweijen et al., 2022c; Ye et al., 2024) and applying it to four 8-hrs LOFAR HBA observations. As additional products, we produced wide-field images at 0.6″ and 1.2″ resolution with sensitivities of respectively 21 μJy beam^{-1} and 39 μJy beam^{-1} near the pointing centre. In these radio maps, we report the detection of 9203 sources at 0.3″, 8567 sources at 0.6″, and 5872 sources at 1.2″.

For accurate calibration, we generated a sky model for our primary in-field calibrator by fitting the spectral index using different surveys and by imaging ELAIS-N1 at 54 MHz with LOFAR LBA data. This approach, along with refinements in the calibration steps for calibrating the in-field calibrator, improved our DI calibration of the international LOFAR stations. We adopted a quantitative approach to assess the selection of the DD calibrators. This enabled us to quickly and robustly select the best secondary calibrators to correct for the varying ionosphere across our field

of view. Although we improved the DD calibration, we identified existing inaccuracies in the calibration solutions for the Dutch core and remote stations. This likely stems from the fact that in this work, due to computational cost, we initially ignored bright sources in our field beyond the facet image boundary during self-calibration. To rectify this, we introduced an extra calibration round specifically for the Dutch stations after subtracting the sky outside each facet. After imaging each individual facet separately in parallel, we mosaiced everything back together into our final wide-field images for each resolution. We find our complete data processing from calibration to imaging to be about two times faster compared to the previous work by Sweijen et al. (2022c) and Ye et al. (2024). This is due to software and hardware improvements.

We find the smearing in our images to be the most severe at the highest resolution, which is intensified by the fact that half of our observations from the LTA were pre-averaged to 2 sec. As a result of our primary in-field calibration strategy, using an accurate sky model, we achieved precise astrometry with median offsets of dRA=$0.094 \pm 0.093''$ and dDEC=$-0.067 \pm 0.064''$ after comparing with optical counterparts selected from the catalogues by Kondapally et al. (2021). We also found accurate flux density scales for the wide-field images.

By comparing the three resolutions, we find $1.2''$ to be a good intermediate resolution to detect sources with extended low surface brightness emission, while the depth and detail in our $0.3''$ resolution map are expected to be great for separating source components or detecting compact objects at higher redshifts. The $0.6''$ resolution map complements these two resolutions for objects that are resolved out at $0.3''$ and unresolved at $1.2''$, such as star-forming galaxies at low flux densities. We also find the detected fraction across the three resolutions to reach completeness between 1 and 10 mJy.

Our work demonstrates the feasibility of making deep wide-field images at subarcsecond resolutions with LOFAR. Near the pointing centre, we reached RMS noise values close to what has been recently achieved with the Dutch LOFAR stations at $6''$, with about 16 times less observation time. Currently, computational costs are the primary obstacle in processing all 500 hrs of LOFAR observations of ELAIS-N1 stored in the LTA (Shimwell et al. in prep.). Addressing the computational challenges will enable the creation of the deepest LOFAR wide-field image, which can be used to uncover objects at the smallest angular scales in the low-frequency radio sky.

4.A. Towards an automated VLBI pipeline

Given the large data volumes, processing LOFAR data with international stations incurs significant computing costs (see also Section 4.4.3). This makes it essential to carefully optimise each of the steps in the current calibration strategy. In addition, removing the manual data processing and visual inspection from the current strategy could lead to the possibility of conducting more automated processing, allowing for a larger high-resolution survey of the northern hemisphere survey. We explore in this appendix section parts of the pipeline that could be replaced by automated approaches with strategies to test and implement these.

The initial steps for obtaining the DI calibration for Dutch stations with `Prefactor` pipeline[16] and running the `DDF` pipeline are already automated. However, human intervention is still required to inspect the output from the `Prefactor` and `DDF` pipelines, as it is important to validate the quality of the observations before proceeding with the data reduction of the long-baselines. A preliminary examination of the data flagging percentages catches the severeness of RFI and can help eliminate corrupted observations. Nonetheless, the flagging fraction does not catch all instances of corrupted data. One therefore typically manually inspects calibration solution plots from the calibrator and target solutions generated by `LoSoTo`. The wide-field image at $6''$, after running the `DDF` pipeline, also allowed us to assess the image quality with only Dutch stations of the observation. This provides information about the effects of the ionosphere on the final wide-field image quality and is therefore a direct indicator of the calibratability of our data. The manual interactions with the data in these steps could be replaced by adopting an automated inspection process such that the future pipeline could determine which observations require modifications to the input data, adjust calibration parameters in the pipeline, or perhaps decide to reject a particular observation.

The next important manual input starts when selecting suitable calibrators in both the in-field DI calibrator and DD calibrators. These steps are highlighted in red in Figures 4.4 and 4.8. For the primary in-field calibrator selection (see Section 4.3.2.1), one typically selects first the calibrator candidates from the LBCS catalogue (Jackson et al., 2016, 2022). If no sources are available or the quality of the solutions and calibration proves after visual inspection to be inadequate, one tries other bright unresolved sources from the LoTSS catalogue (Shimwell et al., 2017, 2019, 2022). The standard method for selecting DD calibrators is similar, but begins by directly identifying candidates from the LoTSS catalogue above a specified flux density threshold. Currently, users inspect visually the calibration solutions and

[16]Note that this step has been recently replaced by the LOFAR Initial Calibration (LINC) pipeline: `https://linc.readthedocs.io`.

image qualities and perhaps change the parameters or decide to exclude the candidate calibrator entirely. In both steps, we suggest applying the selection procedure described in Section 4.3.3.1, wherein the selection we evaluate the circular standard deviation on the phase RR and LL phase difference. This is computationally cheap, as it requires just about 1 CPU hr for each source, and eliminates candidates with an insufficient S/N at the longest baselines. Following this up by an additional selection after self-calibration, will eliminate falsely selected sources (as discussed in 4.3.3.3).

It is important to stress that we can currently not guarantee that our empirically selected circular standard deviation score of 2.3 rad from the `scalarphasediff` calibration will be generally good enough for selecting DD calibrators of other fields. For instance, we found the scores to differ up to ~0.5 rad between the individual nights, which is a significant difference on a score that has a maximum around π. One could for example expect that sky areas at lower declinations would have on average much higher scores. Hence, for implementing these steps for automation, additional tests on different observations and fields are required. Moreover, the process of manually adjusting calibration parameters for self-calibration could be automated by linking it to specific circular standard deviation scores from the `scalarphasediff` calibration and by incorporating additional metrics. We have not yet explored this in detail, but as the circular standard deviation links to the brightness of a source, we expect this to link to the solution interval as well.

After obtaining the image output products, it requires – despite having tools such as `PyBDSF` – a vast amount of work to catalogue source detections. The primary challenge involves source association. Work has been done to automate this through machine learning (e.g. Mostert et al., 2022). However, work needs to be done to re-train the models and improve them for our high-resolution data.

To summarise, we suggest replacing the following manual steps in the calibration strategy with automated approaches:

- ☐ **Solution inspection:** Inspection of for instance `Prefactor`/`LINC` calibration solution output to validate data quality before moving to the calibration for the international LOFAR stations. Solution inspection can be similarly done after every calibration step, such as the DI and DD calibration (see Sections 4.3.2.3 and 4.3.3.2).

- ☐ **DDF-image inspection:** Inspection of the DD-corrected wide-field image quality at $6''$, which is produced by the `DDF`-pipeline. This indicates the severeness of the ionosphere and therefore the calibratability of our data.

- ☐ **Quantity inspection:** An inspection of the output data after every main step in the pipeline can ensure that no data gets 'lost', which involves monitoring

for excessive flagging or tracking job failures on portions of the data.

- **Calibrator selection:** Select the best in-field calibrator source and the best DD calibrator sources by using computational cheap metrics, such as the phase noise metric discussed in Section 4.3.3.1.

- **Calibration parameters:** The metrics for the calibrator selection could be linked to optimising the calibrator parameters, such as the solution interval or smoothness constraints.

- **Source association:** To prepare our output images for scientific analysis, it would be advantageous to automatically and accurately perform source association for our high-resolution data.

4.B. Potential self-calibration issues

In Section 4.3.3.2, we discuss the self-calibration of our DD calibrators. While our source selection performed well (see Section 4.3.3.3), we ran during testing also self-calibration on a few sources that were not selected by our selection metric. We highlight below two examples of sources that were above our selection threshold from Section 4.3.3.1 (so were not selected) but diverged strongly due to various issues.

In Figure 4.30 we find in the upper panel a source that has a bright calibrator nearby, which introduces during phase (`scalarphase`) calibration strong artefacts. This results in bad calibration solutions and no improvements of their images, as is evident from the graphs in the left panel of Figure 4.31, where neither the RMS nor the dynamic range shows any improvement. The phase solutions remain also unstable in the right panel. In the lower panel of Figure 4.30 we display self-calibration cycles from a source that was partly subtracted on the edge of our 2.5°×2.5° field of view. This introduced strong artefacts after amplitude calibration, starting from cycle 3. In Figure 4.32, we find in the left panel the RMS goes up after cycle 4, which corresponds in the right panel to the instability of the phase solutions after this same cycle. Although these sources were not selected by our algorithm, they do demonstrate the effectiveness of performing a pre-selection for bright compact sources and for self-calibration inspection in case of similar or other issues that are not guaranteed to be captured by the phase noise metric selection in Section 4.3.3.1.

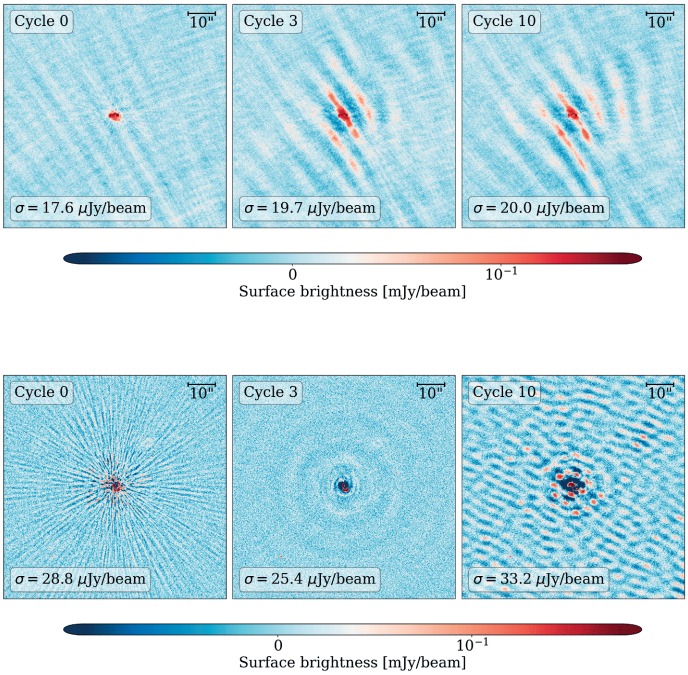


Figure 4.30: Two examples of self-calibration issues. Cycle 0 is the first image with only DI solutions applied from the in-field calibrator (see Section 4.3.2). Cycle 3 corresponds to the self-calibration image after 3 rounds of `scalarphase` calibration. After this cycle, `scalarcomplexgain` calibration is added. This also calibrates for amplitude errors. Cycle 10 shows the result after the 10th self-calibration round.

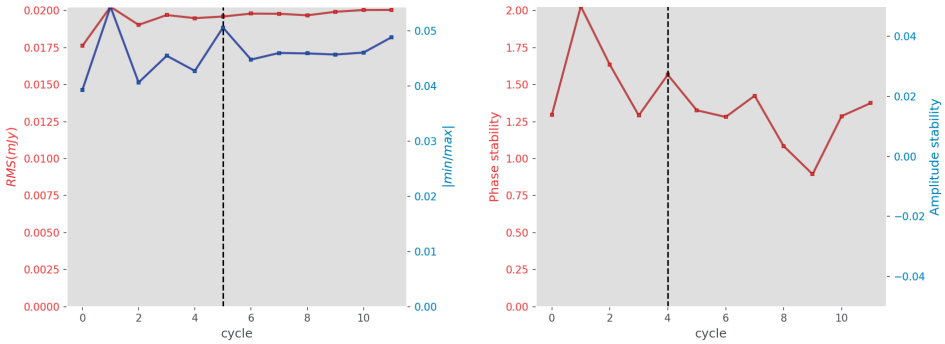


Figure 4.31: Self-calibration and solution stabilities for the source in the upper panel of Figure 4.30. The interpretation of these figures is discussed in the captions of Figures 4.12 and 4.13. It is important to note that this source did not have amplitude corrections, as according to the `auto` settings from `facetselfcal`, this source was not sufficiently bright enough to trigger `scalarcomplexgain` corrections.

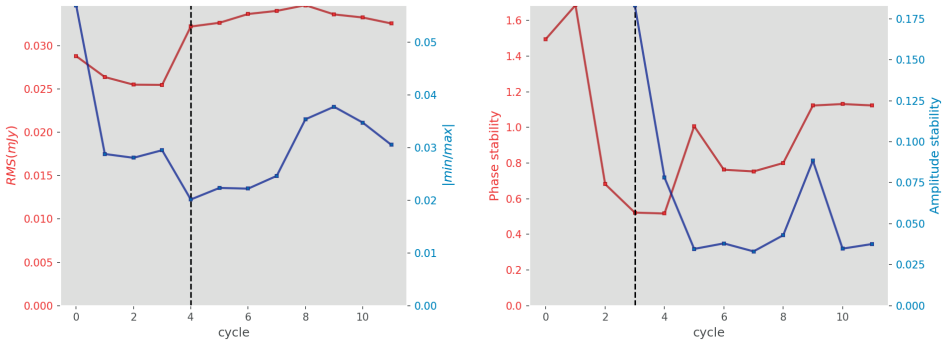


Figure 4.32: Self-calibration and solution stabilities for the source in the lower panel of Figure 4.30. The interpretation of these figures is discussed in the captions Figures 4.12 and 4.13.

Acknowledgements

This publication is part of the project CORTEX (NWA.1160.18.316) of the research programme NWA-ORC which is (partly) financed by the Dutch Research Council (NWO). This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-6218. This work is co-funded by the EGI-ACE project (Horizon 2020) under Grant number 101017567. This work was sponsored by NWO Domain Science for the use of the national computer facilities. LKM and EE are grateful for support from a UKRI Future Leaders Fellowship [MR/T042842/1]. RJvW acknowledges support from the ERC Starting Grant ClusterWeb 804208. MB acknowledges support from INAF under the Large Grant 2022 funding scheme (project "MeerKAT and LOFAR Team up: a Unique Radio Window on Galaxy/AGN co-Evolution". RK and PNB are grateful for support from the UK STFC via grant ST/V000594/1. JP acknowledges support for their PhD studentship from grants ST/T506047/1 and ST/V506643/1. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation. LOFAR data products were provided by the LOFAR Surveys Key Science project (LSKSP; `https://lofar-surveys.org/`) and were derived from observations with the International LOFAR Telescope (ILT). LOFAR (van Haarlem et al., 2013) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and which are collectively operated by the ILT foundation under a joint scientific policy. The efforts of the LSKSP have benefited from funding from the European Research Council, NOVA, NWO, CNRS-INSU, the SURF Co-operative, the UK Science and Technology Funding Council and the Jülich Supercomputing Centre.

5

Unlocking ultra-deep wide-field imaging with sidereal visibility averaging

J.M.G.H.J. de Jong, R.J. van Weeren, T.J. Dijkema, J.B.R. Oonk, H.J.A. Röttgering, F. Sweijen

Astronomy & Astrophysics, Volume 694, id.A98, 9 pp.

Abstract

Producing ultra-deep high-angular-resolution images with current and next-generation radio interferometers introduces significant computational challenges. In particular, the imaging is so demanding that processing large datasets, accumulated over hundreds of hrs on the same pointing, is likely infeasible in the current data reduction schemes. In this paper, we revisit a solution to this problem that was considered in the past but is not being used in modern software: sidereal visibility averaging (SVA). This technique combines individual observations taken at different sidereal days into one much smaller dataset by averaging visibilities at similar baseline coordinates. We present our method and validate it using four separate 8-hrs observations of the ELAIS-N1 deep field, taken with the International LOw Frequency ARray (LOFAR) Telescope (ILT) at 140 MHz. Additionally, we assessed the accuracy constraints imposed by Earth's orbital motion relative to the observed pointing when combining multiple datasets. We find, with four observations, data volume reductions of a factor of 1.8 and computational time improvements of a factor of 1.6 compared to standard imaging. These factors will increase when more observations are combined with SVA. For instance, with 3000 hrs of LOFAR data aimed at achieving sensitivities of the order of μJy beam^{-1} at sub-arcsecond resolutions, we estimate data volume reductions of up to a factor of 169 and a 14-fold decrease in computing time using our current algorithm. This advancement for imaging large deep interferometric datasets will benefit current generation instruments, such as LOFAR, and upcoming instruments such as the Square Kilometre Array (SKA), provided the calibrated visibility data of the individual observations are retained.

5.1. Introduction

Modern and upcoming radio interferometers such as the International LOw Frequency ARray (LOFAR) Telescope (ILT van Haarlem et al., 2013) and the Square Kilometre Array (SKA Dewdney et al., 2009) and its pathfinders (Schinckel et al., 2012; Tingay et al., 2013; Jonas & MeerKAT Team, 2016) advance our ability to study the Universe in unprecedented detail. However, the substantial volumes of data generated by these instruments present significant challenges for data processing and storage (costs). This issue is especially pronounced when combining multiple observations of the same sky area for deep high-resolution imaging, as this involves processing much larger volumes of data compared to imaging with a single observation or at lower resolutions. To ensure that the high costs associated with data processing do not outweigh the scientific benefits, it is crucial to employ more efficient data processing and handling techniques.

Recent deep surveys dedicated to specific areas in the sky, each spanning hundreds of hrs with LOFAR, have produced wide-field images with sensitivities reaching 20 μJy beam^{-1} at 144MHz and 6″ resolution (Kondapally et al., 2021; Duncan et al., 2021; Tasse et al., 2021; Sabater et al., 2021; Best et al., 2023; Bondi et al., 2024). The calibration and imaging were conducted using only the Dutch core and remote stations, excluding all international stations. Including also the international LOFAR stations allows for the creation of ultra-deep, degree-scale images with sub-arcsecond resolutions, due to the additional baselines extending up to 2000 km. These images contain, at sub-arcsecond resolutions, up to 10 billion pixels over 2.5×2.5 degrees and are generated from data amounting to tens to hundreds of terabytes when combining all available observations of the same field. The feasibility of generating these degree-scale images after calibrating all LOFAR's international stations has been proven by producing the first 0.3″ wide-field image with a sensitivity of 32 μJy beam^{-1} (Sweijen et al., 2022c). Recent advancements have led to the production of the deepest wide-field image currently available at 0.3″ resolution and 140 MHz, with a sensitivity of 14 μJy beam^{-1}, by combining four observations (de Jong et al., 2024). While these pioneering studies have progressively developed strategies that can be scaled up to process hundreds of observing hrs of the same pointing, they have also highlighted that the final imaging of the calibrated data is a computational bottleneck. This step consumes about 80% of the total computational time of the entire data reduction pipeline, due to the large data volume left after calibration. To address this, a reduction of the data volumes before imaging without significantly compromising image quality is necessary.

An effective way to compress the data volume of interferometric data is by decreasing the number of visibilities that are needed to create an image. These visibilities are measurements of the correlated signals between pairs of antennas

(baselines), capturing the Fourier components of the sky brightness distribution essential for image reconstruction. One commonly used method to do this is by using baseline-dependent averaging (BDA; e.g. Cotton, 1986, 2009; Skipper, 2014; Wijnholds et al., 2018; Atemkeng et al., 2022). This leverages the fact that different baseline lengths may have different time and frequency resolutions to recover the information for imaging without introducing time and bandwidth smearing. For short baselines, which measure large-scale structures in the sky, the visibility function changes slowly with time and frequency. In contrast, long baselines are sensitive to small spatial resolutions, where the visibility function may vary more rapidly with time and frequency. Therefore, visibilities for short baselines can be averaged over longer time periods and broader frequency channels compared to longer baselines. BDA can, for example, reduce data volumes for the SKA-Low and SKA-Mid up to about ~85% (Wijnholds et al., 2018; Deng et al., 2022). Alternatively, it is also possible to perform data volume compression by using Dysco compression Offringa (2016). This method does not reduce the number of visibilities but uses lossy compression to reduce the storage space and, thus, the reading speed of visibility data. Dysco compression works well on lower signal-to-noise ratio (S/N) data by taking advantage of the fact that variations in the data are primarily due to Gaussian noise rather than the actual signal. This allows for efficient data compression without losing important visibility information to reconstruct images. On average, this technique reduces the data volume for a typical LOFAR observation by a factor of 4 to 6, partly also due to more efficient storage of weights.

Despite the successes in terms of compressing datasets of individual observations, data volume and computational time still scale linearly when performing deep imaging with multiple observations, due to the increased number of visibilities that in imaging software all need to be processed. One way to reduce computational wall time is to create separate images from individual observations and then average them in image space. While this approach enables parallel processing for each observation, it still necessitates gridding all visibilities from the different observations during imaging, resulting in about the same total computational time as imaging all visibilities together (which we refer to as 'standard imaging' throughout this paper). Additionally, image quality is worse when averaging images compared to standard imaging because deconvolution of faint sources can only be done when performed with all observations together. An alternative approach to combine observations and reduce data volume and computational time is to utilise the repeating baseline tracks from deep observations taken over multiple sidereal days. In this way, it is possible to reduce the number of visibilities that need to be imaged. This was, for instance, conducted by Owen & Morrison (2008), who averaged visibilities at the same hour angles using data from the Very Large Array (VLA) at 1.4 GHz with a

total integration time of 140 hrs. They implemented this method in the Astronomical Image Processing System (AIPS van Diepen & Farris, 1994; Glendenning, 1996; Greisen, 2003) as the task STUFFR[1], which combines several AIPS tasks to perform visibility averaging over different sidereal days. This was later also used for deep imaging of the GOODS-N field with the VLA (Owen, 2018).

AIPS was originally designed to handle smaller data volumes compared to the large datasets produced by modern interferometers, such as LOFAR and the SKA. While some processing improvements have been developed for AIPS (e.g. Kettenis et al., 2006; Cotton, 2008; Bourke et al., 2014), it is not optimised to deal with direction-dependent effects (DDEs). These pose an added challenge to achieving wide-field images with high dynamic range and S/N below a few hundred MHz. DDEs are primarily introduced by the ionosphere and beam model errors, distorting the 'real' visibilities differently across the field of view. For LOFAR and SKA data processing have been developed, capable of efficiently processing datasets of the order of tera-to petabytes on large powerful multi-CPU machines, while DDEs may be corrected with software packages such as DP3[2] (van Diepen et al., 2018a; Dijkema et al., 2023), SPAM (Intema et al., 2009), Sagecal[3] (Kazemi et al., 2011), KillMS[4] (Tasse, 2014a,b; Smirnov & Tasse, 2015), and facetselfcal[5] (van Weeren et al., 2021). These are typically integrated with a facet-based approach, where the field is divided into multiple facets (van Weeren et al., 2016b; Williams et al., 2016), with each facet receiving its own calibration solutions. Since calibration for systematic and ionospheric effects is best performed on a per-observation basis, it is essential to average visibilities from different observations corresponding to different sidereal days only after calibrating for the DDEs in each facet. This requires, before averaging visibilities from different observations, to split off datasets for each facet from the full dataset and treating each facet separately, as was demonstrated by Sweijen et al. (2022c) and de Jong et al. (2024) without averaging visibilities for similar hour angles or baseline coordinates.

In this paper, we are exploring a revised method to average visibilities over sidereal days on already calibrated facet data. We term this 'sidereal visibility averaging' (SVA). To demonstrate our method, we average visibilities from datasets from four different LOFAR observations calibrated by de Jong et al. (2024). By testing various averaging settings, we compare and optimise the balance between image quality and computing costs. We also consider effects on the binning of similar baseline coordinates and frequency offsets due to Earth's celestial motion.

[1] http://www.aips.nrao.edu/cgi-bin/ZXHLP2.PL?STUFFR
[2] https://dp3.readthedocs.io
[3] https://github.com/nlesc-dirac/sagecal
[4] https://github.com/saopicc/killMS
[5] https://github.com/rvweeren/lofar_facet_selfcal

Given that imaging accounts for approximately 75-80% in the current LOFAR sub-arcsecond wide-field data reduction pipeline (Sweijen et al., 2022c; de Jong et al., 2024), we address a significant part of current computational challenges for ultra-deep imaging of a single pointing on the sky. While we focus on data from LOFAR in this paper, we advocate this method as a viable solution to address computing and storage challenges for deep multi-epoch imaging with other instruments as well.

In Section 5.2, we first discuss the SVA algorithm. This is followed by an overview of the data used in this paper in Section 6.5. We then present our results in Section 6.6, followed by a discussion in Section 6.7. Finally, we conclude with a summary and conclusions in Section 6.8.

5.2. Sidereal visibility averaging

Interferometric datasets consist of several components, including a time axis, a frequency axis, baseline coordinates (uvw), visibilities, and their corresponding weights. The initial time and frequency resolution of these axes are determined by the settings of the interferometer or correlator. Visibilities are the measurements of the correlated signals for baselines at specific moments in time and for different frequencies. As the Earth rotates, these baseline tracks move over time, with each timestamp corresponding to a point in the uvw plane (e.g. Brouw, 1975). This plane is a coordinate system used to describe the relative positions of antennas. When observations with the same pointing centre are conducted over more than one sidereal day, parts of the baseline tracks are repeated. This repetition allows us to average the visibilities in similar baseline coordinate bins. We discuss in this section how the SVA algorithm makes efficient use of the possibility to average observations over multiple sidereal days. In this paper, we use LOFAR as an example because we have recently reduced high-resolution LOFAR data at hand (de Jong et al., 2024).

5.2.1. Frequency and time axis

The detected radio signals are split into specific frequency bands using a polyphase filterbank, which processes and organises these signals into channels. For LOFAR, the central frequency of these channels is pre-defined which ensures that LOFAR data is stored using the same channel centres in the LOFAR Long Term Archive (LTA)[6] (van Haarlem et al., 2013). As frequencies may be averaged during processing, it is essential that these channels are averaged by common denominators, such that when applying SVA, we only have to match the corresponding frequency channels of the visibilities. Matching visibilities with frequency offsets can otherwise

[6] `lta.lofar.eu`

result in image distortions. This may include inaccurate source positioning when channels with different centres are combined or bandwidth smearing if observations with different or too large frequency channel widths are combined. However, in the context of SVA – where adjustments to the *uvw*-plane are only influenced by the time axis – frequency-related effects on image accuracies do not arise, as long as SVA is applied to data with frequency channels that share the same centres and widths. It is important to note that frequency-related issues can still be introduced by Doppler shifts, as we subsequently discuss in Section 5.5.3.

The time axis from observations taken at different moments must be converted to a common time axis, which can be done using the local sidereal time (LST). The LST relates to the hour angle relative to the vernal equinox, allowing us to align the time axes from different observations to a single sidereal day. We evaluate the LST at the canonical centre of LOFAR. The time tracks of different observations in LST do not need to overlap exactly, as observations at the same pointing are not taken simultaneously on a sidereal day. For instance, if the integration time for an observation is 2 seconds, another observation of the same pointing could be offset by 1 second in sidereal time. Since the *uvw*-plane and the binning in this plane between different input datasets depends on the time resolution, careful consideration of the time axis is essential in reconstructing the correct image. A critical factor is time smearing. To achieve high-quality images while minimizing data volume, it is crucial to balance time resolution with data volume and image quality. This balance is further discussed in Sections 5.4.1, 5.4.2, and 5.5.1. Note that mapping different observations to a common LST axis leads to the loss of intrinsic time information for certain astronomical objects. This is most notable for the varying flux of transient sources.

While one day in LST consists of about 23.93 hrs, we use the fact that the Fourier transform of a real-valued function (the sky brightness distribution) is Hermitian, meaning that the transform exhibits complex conjugate symmetry. Hence, a visibility at coordinates (u, v, w) has a corresponding complex conjugate at $(-u, -v, -w)$. Therefore, our output dataset does not cover more than 11.97 hrs.

5.2.2. Algorithm

The sidereal visibility averaging algorithm involves the following steps:

1. First, using the LST, we construct a time axis for our output dataset that encompasses all LST points from our input datasets. Though the measurement set allows specifying a time axis in LST, we convert the times to a 'representative' UTC time around the median time of all observations. The time resolution (Δt) can be specified as input or can be calculated using the angular resolution (θ_{res})

and maximum distance from the phase centre (θ) with the following formula:

$$\Delta t = 2.9 \times 10^4 \left(\frac{\theta_{\rm res}\sqrt{1-\tau}}{\theta} \right), \tag{5.1}$$

where τ is the time smearing, which is equivalent to the peak intensity loss of a source. This formula is based on the time-smearing equations from Bridle & Schwab (1999) for the average smearing effect on an image. We set τ by default equal to 0.95, but it can be changed to a more conservative value closer to 1. We also create a frequency axis that includes all frequency channels from our input datasets.

2. Next, we add all unique LOFAR stations from our input datasets to our output dataset and make mappings that map the LOFAR station IDs from the output to those in the input datasets. This allows us to quickly identify which data entries correspond to which baseline across all our input and output datasets, as not all LOFAR observations are observed using the same set of stations.

3. Using the baseline coordinates from our input datasets and the LST from each observation, we obtain a preliminary estimate of the baseline coordinates for our output dataset by applying nearest neighbour interpolation for each baseline, using `scipy`'s `interpolate` library (Virtanen et al., 2020). Nearest neighbour interpolation yields similar results to other interpolation methods, as we further refine the accuracy of the output baseline coordinates by averaging all nearest *uvw* values from the input datasets. Note that each baseline coordinate from an input dataset can only correspond to one baseline coordinate in the output dataset. In Figure 5.1 we demonstrate this procedure by showing the baseline coordinates from the input dataset compared to the output dataset for one baseline and different time resolutions.

4. We create index mappings between the baseline coordinates of our input and the nearest baseline coordinates from our output datasets. These mappings allow us to quickly determine for each baseline coordinate in our input datasets to which baseline coordinate from the output dataset these correspond during averaging of our visibilities and summing the weights as explained below.

5. Visibility weights are factors that account for the reliability of visibilities, improving overall accuracy and the S/N during imaging. To obtain the visibility weights of the output dataset (W), we sum the visibility weights from the input datasets (W_i):

$$\bar{W}(u,v,w) = \sum_i W_i(u,v,w)$$

This ensures that the most reliable visibilities in the output dataset have the largest weights. Also, *uvw* for which a subset of the input observations contribute points, will have correspondingly lower weights.

6. We utilise the visibility weights to compute a weighted average of the visibility data, yielding the output visibility values as follows:

$$\bar{V}(u,v,w) = \frac{\sum_i V_i(u,v,w) W_i(u,v,w)}{\bar{W}(u,v,w)}$$

7. Finally, we add a flagging column, flagging all values with output visibility weights equal to 0, as these correspond to *uvw* coordinates in the output dataset that did not have neighbouring visibilities from the input data or were already flagged in the input data.

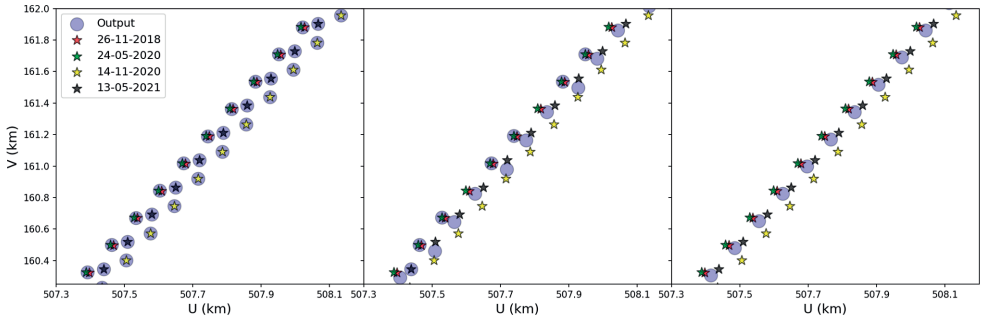


Figure 5.1: Different *uv* samplings for one baseline with length 594 km between a Dutch core station (CS001HBA) and the Swedish LOFAR station at Onsala Space Observatory (SE607HBA). The input datasets are represented by red, green, yellow, and black stars and labelled by their observing dates. The output dataset is represented by blue dots. The input data has a time resolution of 8 seconds. *Left panel:* Output time sampling at $\Delta t = 1$ second. *Centre panel:* Output time sampling according to Equation 5.1 at $\Delta t = 3.6$ seconds. *Right panel:* Output time sampling at $\Delta t = 8$ seconds.

Since our visibility algorithm considers mappings between baselines of the input and output data, it can handle datasets with BDA. We also ensured that our data could be compressed using Dysco compression by testing with the default settings of 10 bits per float for visibilities and 12 bits per float for visibility weights in `DP3`. We measured the image RMS noise with and without Dysco compression and found a difference of less than 0.1%. However, it is important to note that we only tested this with four datasets (see Section 6.5). Using more datasets or the presence of very bright sources could increase the S/N, potentially leading to image quality losses

with Dysco compression. In such scenarios, it may be necessary to consider using more conservative bit-rates or forgoing Dysco compression entirely.

Our sidereal visibility averaging code is currently written with Python and may be implemented as a separate step in data reduction pipelines.[7] The code builds mainly on the functionalities from `casacore`[8] (Casacore Team, 2019; CASA Team et al., 2022) to work with measurement sets, which are the standard data format used for radio interferometric data[9]. We also use `astropy`[10] (Astropy Collaboration et al., 2013, 2018, 2022) for unit conversions and the *uvw* coordinate system and `scipy`[11] (Virtanen et al., 2020) for nearest neighbouring interpolation and retrieving the binary-tree quick nearest neighbour lookup from Maneewongvatana & Mount (2002). To increase the processing speed we also utilise `joblib` (Joblib Development Team, 2020)[12] for parallel processing. To increase the efficiency of the algorithm, it may be investigated in the future to port the code to a more efficient programming language or implement it in an already existing efficient radio astronomical software package.

5.3. Data

To demonstrate the SVA algorithm, we utilise calibrated datasets corresponding to four observations taken by LOFAR of the ELAIS-N1 deep field. These datasets were processed and imaged at sub-arcsecond resolution by de Jong et al. (2024). The corresponding observations are part of two different observing projects (`LT10_012` and `LT14_003`, PI: P.N. Best) and were downloaded from the LTA.[13] A brief summary of the metadata for these observations is provided in Table 5.1. The maximum extent between the observation times is 2.5 years.

SVA should be applied after all calibration have been performed per observation. Otherwise, various systematic or ionospheric effects cannot be properly corrected, as the original time axis information for each observation is lost and only a sidereal time axis remains. The complete calibration process for the datasets used in this work is detailed in de Jong et al. (2024). The most essential step in the data reduction process that we need to highlight in this work, is the use of the facet-based approach to correct for DDEs (e.g. van Weeren et al., 2016b). Initially, direction-independent

[7] The code is currently available at `https://github.com/jurjen93/sidereal_visibility_avg`

[8] `https://casacore.github.io/python-casacore/`

[9] `https://casacore.github.io/casacore-notes/229.pdf`

[10] `https://www.astropy.org`

[11] `https://scipy.org`

[12] `https://joblib.readthedocs.io`

[13] `lta.lofar.eu`

Observation ID	L686962	L769393	L798074	L816272
Observation date	26-11-2018	24-05-2020	14-11-2020	13-05-2021
Start time	07:13:43	19:20:26	08:11:00	19:41:00
Integration time	8 hrs	8 hrs	8 hrs	8 hrs
Frequency range	120-166 MHz	120-166 MHz	115-164 MHz	115-164 MHz
Stations (International)	51 (13)	51 (13)	50 (12)	52 (14)

Table 5.1: Metadata from the four used ELAIS-N1 observations. The pointing centres were aligned during the data reduction process.

effects for the longest baselines are calibrated using one bright calibrator. Then, several other bright calibrators with sufficient S/N at the longest baselines are selected and calibrated. These calibrators define a Voronoi tessellation, where within each facet the calibration solutions are assumed to be constant (Schwab, 1984; van Weeren et al., 2016b). The visibilities from the datasets corresponding to these facets are individually imaged by subtracting sources outside these facets and then phase shifting to their centre. This process allows for additional averaging of the calibrated visibilities for each facet without introducing smearing effects, thereby speeding up the imaging process.

In the following sections, we use datasets with calibrated visibilities from two facets to test the SVA algorithm: facet 12 and facet 25 from Figure 14 in de Jong et al. (2024). These datasets are stored using Dysco compression (Offringa, 2016), which reduces data volume but is incompatible with BDA. However, as the SVA algorithm operates on a per-baseline basis, the analysis presented in the following sections is applicable to data with BDA applied as well. Both facets were, during pre-processing, averaged to a time resolution of 8 seconds and frequency resolution of 97.66 kHz, which reduces the data volume while avoiding time and/or bandwidth smearing. Further averaging would cause smearing and lead to an irreversible loss of information that cannot be restored during SVA. The sky areas covered are 0.20 deg^2 for facet 12 and 0.22 deg^2 for facet 25. These correspond to the time and frequency resolution required for creating images at a resolution of $0.6''$. We conduct the analysis in this paper at a resolution of $0.6''$, as this offers four times better imaging speed, thereby reducing the computational resources needed to generate images. All discussions in this paper are directly applicable to a resolution of $0.3''$ with the same calibrated data.

All images made in this paper are produced with `wgridder` (Arras et al., 2021; Ye et al., 2022) from `WSClean` (Offringa et al., 2014), using the same settings as de Jong et al. (2024). This includes a Briggs weighting of -1.5 (Briggs, 1995), a minimum uv-value of 80λ (corresponding to a largest angular scale of $\sim 43'$), and a pixel size of $0.2''$. For efficient deep cleaning and to better recover extended diffuse

emission, we apply 'auto' masking, multi-scale deconvolution, and an RMS box equal to 50 times the synthesised beam size (Cornwell, 2008; Offringa & Smirnov, 2017). Afterwards, we restore all images to a common resolution of $0.6''$ to allow for direct comparisons.

5.4. Results

A key functionality of the SVA algorithm is to accurately average visibilities and weights from various observations for similar baseline coordinates. The binning is directly controlled by the given time resolution in the output dataset, as demonstrated in Figure 5.1. Frequency offsets between observations are not a concern, as the datasets have aligned frequency channels (see Section 6.5), which is standard for LOFAR data (see Section 5.2.1). The time resolution of the output dataset after applying SVA affects both the image quality, in terms of RMS noise and smearing, and the data volume. It is essential to find an optimised balance between both image quality and data volume for reducing computational time and data volume, while obtaining high-quality science-ready images after imaging. In this section, we evaluate the output of the SVA algorithm using data from the two selected facets, considering both image quality and data volume.

5.4.1. Image quality

The datasets from the facets used in this work have both an initial time resolution of $\Delta t_0 = 8$ seconds, which was identified as a balance between data volume and time smearing by de Jong et al. (2024), when performing standard imaging without SVA. To compare the image qualities for different time resolutions of the output datasets after applying SVA, we plot in Figure 5.2 the comparison between different image properties of the original non-averaged image and the images after SVA for different time resolutions.

We find in the left panel of Figure 5.2 that the background RMS noise improves for smaller Δt. Facet 25 exhibits marginally lower relative RMS noise offsets because, following its DD calibration, the global amplitudes are for this facet slightly elevated compared to those in facet 12, as shown in Figure 25 of de Jong et al. (2024). In the centre panel of Figure 5.2, we find no trend for the different time resolutions and offsets between the original integrated flux density (S_0) and the integrated flux density after SVA of less than 1%. The right panel displays the peak intensity ratios for the calibrator sources of each facet, where we find the peak intensity to reduce up to about 5% towards lower time resolutions. This indicates an increase in time smearing, as smearing impacts only the peak intensity while conserving the flux densities to the first order. This effect is more pronounced for

the calibrator source of facet 25, as this source is located closer to the edge of its facet, where smearing effects are strongest.

In Figure 5.3, we compare cutouts of three extended sources located in the facets when performing standard imaging, with imaging after SVA at a $\Delta t = 3.6$ seconds (which follows from Equation 5.1 at $\tau = 0.95$), and the subtraction between both images. We observe a slight increase in the RMS background noise (as shown in the left panel of Figure 5.2), and we find as expected the sources to disappear in the subtracted images.

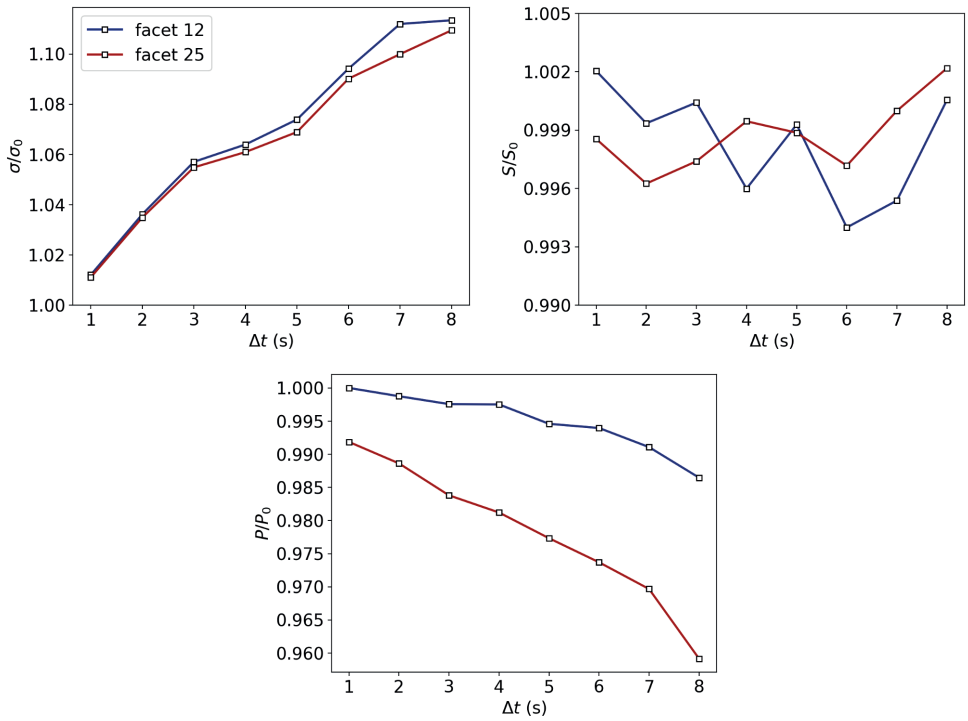


Figure 5.2: Comparing different image properties for different time resolutions (Δt) with and without sidereal visibility averaging for facets 12 and 25 from de Jong et al. (2024) considered in this paper. *Left top panel:* The measured RMS noise background ratios for each facet without (σ_0) and with (σ) sidereal visibility averaging. *Right top panel:* The measured integrated flux density ratios between the calibrator sources for each facet without (S_0) and with (S) sidereal visibility averaging. *Lower panel:* The measured peak intensity ratios between the calibrator sources for each facet without (P_0) and with (P) sidereal visibility averaging.

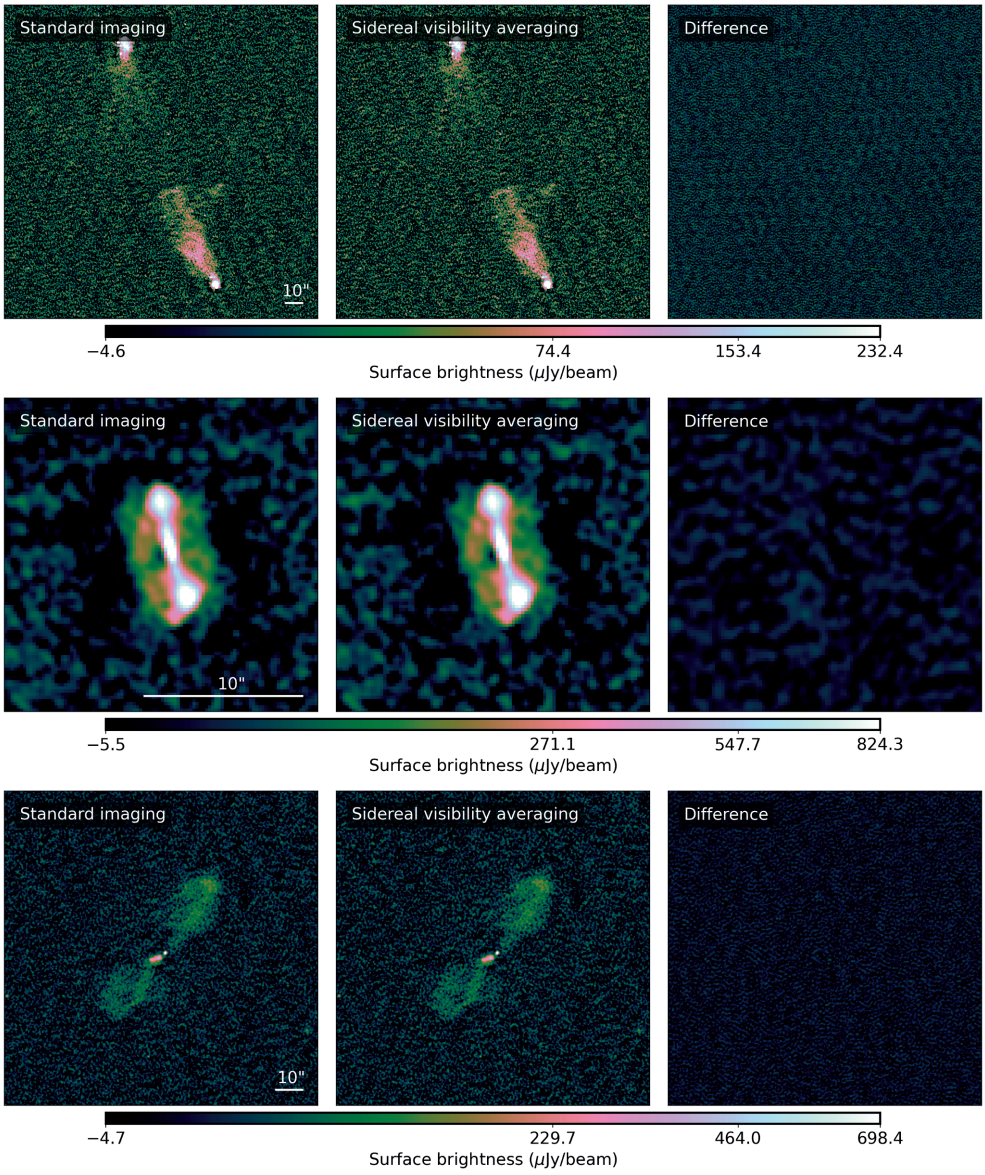


Figure 5.3: Comparison between the 0.6″ resolution images produced with and without sidereal visibility averaging for three extended radio galaxies (*rows*). *First column:* Original image from de Jong et al. (2024) produced after standard imaging without sidereal visibility averaging. *Second column:* Result when imaging after applying sidereal visibility averaging at 4 second resolution. *Third column:* The image after subtracting the second from the first image. Details about the data and imaging settings are provided in Section 6.5.

5.4.2. Computing resources

The main goal of applying SVA is to reduce the required data volume and computational cost for imaging. We will therefore consider the benefits in terms of data storage and computational time.

The data volume after applying SVA ($\mathcal{V}$) depends on the sampling rate of the new time axis. If the time resolution of the new dataset matches that of the N input datasets, the resulting data volume is $\mathcal{V} = \frac{\sum_{i=1}^{N} V_i}{N}$, where V_i are the data volumes of the input datasets. In the more general case, where the new time resolution (Δt) differs from the input time resolution (Δt_0), the input and output data volumes are related by

$$\mathcal{V} \sim \frac{\Delta t_0}{N \cdot \Delta t} \sum_{i=1}^{N} V_i. \tag{5.2}$$

This implies that to achieve any data volume reduction, it is essential that $\Delta t > \frac{\Delta t_0}{N}$, which gives with Equation 5.1

$$N > 3.4 \times 10^{-5} \left(\frac{\Delta t_0 \cdot \theta}{\theta_{\text{res}} \sqrt{1-\tau}} \right), \tag{5.3}$$

an expression for the lower bound for the number of observations that are necessary to obtain a data volume reduction, given the characteristics of the desired image size, smearing allowance, and output time resolution. In our test case with $N = 4$ observations and an input time resolution of $\Delta t_0 = 8$ seconds, we find the data volume

$$\mathcal{V} \sim \frac{2}{\Delta t} \sum_{i=1}^{4} V_i.$$

This shows that to obtain a data volume reduction (less than $\sum V_i$), $\Delta t > 2$ seconds is required. This is satisfied for $\tau = 0.95$, as this corresponds to $\Delta t = 3.6$ seconds (following from Equation 5.1).

In terms of the computational costs, we know that the computational time required for imaging large datasets scales almost linearly with the number of visibilities (e.g. de Jong et al., 2024). This is because the amount spent on deconvolution is almost negligible with respect to the time spent on repeatedly gridding and degridding visibilities, and reading/writing them to disk. Since performing the SVA algorithm also requires additional computational time for applying the algorithm for each dataset (T_{sva}), we need to include this in the total computational time. This results in the following total computational time for imaging and applying the

SVA algorithm:

$$\mathcal{T} \sim T_0 \frac{\Delta t_0}{N \cdot \Delta t} + N \cdot T_{\rm sva},$$

where $\mathcal{T}$ is the total computational time with SVA and T_0 is the total computational time when performing standard imaging without SVA. For the SVA algorithm to be efficient, it must satisfy the condition

$$T_{\rm sva} < \frac{T_0}{N}\left(1 - \frac{\Delta t_0}{N \cdot \Delta t}\right),$$

which again strictly requires $\Delta t > \frac{\Delta t_0}{N}$. Both our facets were originally imaged within about 42 hrs without SVA. So, in this case, using 4 datasets, we find

$$T_{\rm sva} < 10.5\left(\frac{\Delta t - 2}{\Delta t}\right) \text{ hrs.}$$

The SVA algorithm currently takes $T_{\rm sva} = 0.7$ hrs for an output time resolution of $\Delta t = 3.6$ seconds with our 4 datasets on an Intel® Xeon® Gold 5220R Processor with 96 cores. This clearly satisfies the above condition and reduces the computational time by a factor of 1.6. However, it is important to emphasise that for different datasets, more optimised code, or a different processor node, the computing cost in terms of CPU hrs could be more favourable.

5.5. Discussion

We have introduced the SVA method and applied it on deep-calibrated LOFAR data at $0.6''$ resolution. Having examined image qualities at various time resolutions, we delve in this section deeper into the image quality and computing resource balance. This enables us to estimate the advantages of scaling up to combine a larger number of observations. We also investigate the effects of Earth's celestial motion, introducing *uvw* and frequency offsets.

5.5.1. Image quality vs. compute resources

In Section 5.4.1 and Figure 5.2, we found that image quality reduced due to increased RMS noise and increased smearing of the order of a few per cent after interpolating the input baseline coordinates to a new *uvw*-plane. The trend in the left panel of Figure 5.2 also clearly shows that the RMS noise goes up for larger time resolutions. Although Briggs weighting may in our tests have an effect on the noise difference, since the *uvw* coordinates are different between the imaging with and

without SVA, we still observed an increase in the RMS noise in images produced with data from a few international baselines and with uniform weighting when SVA was applied. This suggests that the primary factor behind the noise increase is related to the interpolation process, followed by the imaging where the *uvw* coordinates are interpolated again onto a regular grid to facilitate efficient fast Fourier transforms (FFTs). Consequently, the gridded visibilities represent interpolations of already interpolated *uvw* coordinates, compounding inaccuracies introduced during gridding. The small loss in peak intensity when comparing data imaged with and without SVA, is attributed to additional time smearing effects. This is likely caused by the shifts in baseline coordinates introduced during SVA. This adds to smearing that may already be present in the data.

We found that setting the time resolution to $\Delta t = 3.6$ seconds, allowing a smearing factor of $\tau = 0.95$, results in a computing time improvement of 1.6 times and a data volume reduction of 1.8 times when combining 4 observations of ELAIS-N1 with the current SVA code. While this is a significant gain, these resource savings increase further when combining more datasets with SVA. For example, imaging the 64 available observations of the ELAIS-N1 deep field in the LTA without SVA would be rather costly, as this would currently take about 1.8 million CPU hrs for the final imaging of the data. However, using SVA this wall-time reduces by about a factor of 10, saving almost 1.6 million CPU hrs. Looking even further ahead, with plans to observe a single LOFAR pointing for 3,000 hrs, we anticipate reductions up to a factor of ~169 in data volume and a 14-fold decrease in computing time, while achieving point source sensitivities in the μJy beam^{-1} range.

The image quality reduction in terms of time smearing and RMS noise of around a few per cent compared to standard imaging, becomes in general certainly acceptable given the improved depth of the output image by $\sqrt{N}$ with approximately $\frac{\Delta t_0}{\Delta t} N$ lower data volumes and faster imaging times. For very large numbers of observations, one could also consider to optimise image quality by selecting a finer time resolution and more conservative smearing factors (τ). In the case of 500 or 3000 hrs observational time, one would for instance with $\tau = 0.99$ and our current code still achieve substantial reductions in computational time by factors of approximately 7 and 12, respectively, while reducing data volumes by about 13-fold and 75-fold, respectively. Alternatively, if it is known in advance that SVA will be used to combine observations for deep imaging, these observations could be strategically scheduled to align the start and end time in LST and record the observations with the same time resolutions in LST. This ensures minimal baseline coordinate offsets between observations, disregarding, for now, the effects of Earth's celestial motion discussed in Section 5.5.2.

5.5.2. Precession, nutation, and aberration

One of the main challenges of combining observations taken over different sidereal days, is the fact that the coordinates of baselines are not fixed and alter over time due to Earth's celestial motion. Precession is the conical motion of the Earth's rotation axis, while nutation refers to the smaller oscillations superimposed on the longer-term precession motion (e.g. Rekier et al., 2022). Both are for the most part due to the gravitational forces exerted by the Sun and the Moon. Precession has the most dominant effect on the baseline tracks with a rate of about $50.2''$ per year, whereas nutation has a smaller effect with an amplitude of $9.2''$ over a period of 18.6 years (Mathews et al., 2002; Dehant et al., 2017). In addition, annual aberration, which results from the Earth's orbital motion around the Sun, causes another apparent shift in the observed positions of astronomical objects. This effect introduces a maximum shift in baseline coordinates of approximately $20.5''$ over the course of a year (e.g. Gubanov, 1973; Kovalevsky, 2003).

In Figure 5.1, we observe the above-mentioned effects on the baseline coordinates using different time resolutions. The distances between baseline coordinates within the same dataset represent the maximum allowable separation between *uvw* points, which, as determined by de Jong et al. (2024), stay within acceptable smearing limits. The different *uvw* samplings, represented by the blue dots in Figure 5.1, demonstrate that we remain within these limits, indicating that both precession and aberration have minimal influence on the resulting dataset after SVA. However, when combining observations taken with many years in between or when for instance observing objects at higher declinations, precession becomes increasingly significant and requires adjustments to the *uvw* sampling to avoid substantial time-smearing effects.

The SVA algorithm addresses the challenges introduced by celestial motions in part by using nearest-neighbour interpolation of the baseline coordinates. This can be further refined by employing a finer time resolution, as shown in the three panels of Figure 5.1. In some cases, it may also be more accurate to generate the output *uvw*-plane through interpolation across the entire *uvw* space, rather than on a per-baseline basis, since *uvw* points from different baselines may overlap due to precession. When combining a large number of observations with large time intervals between them, an alternative approach is to group datasets within specified observing time ranges and apply SVA only on these subsets. This approach still reduces data volume without combining all datasets into a single set, thereby minimizing the loss of image quality caused by combining too distant *uvw* coordinates from different observations. The new set of sidereal averaged datasets can then be imaged together. Additionally, to minimise aberration effects, it is advisable to schedule observations for SVA close to each other.

5.5.3. Doppler shifts

When we combine observations taken at different moments in time, it is also important to consider Doppler shifts which occur due to the relative motion between our instrument on earth and the sky direction, causing observed frequency changes. Doppler shift differences between observations result in frequency offsets that may spectrally distort our images. This is in particular relevant to spectral line science.

The radial velocity is given by

$$v_r = \mathbf{v}_\mathrm{e} \cdot \hat{\mathbf{r}},$$

where $\mathbf{v}_\mathrm{e}$ is the velocity vector of the Earth and $\hat{\mathbf{r}}$ is the unit vector pointing from the observer to the sky direction, using the pointing centre of the observation and the antenna locations relative to the centre of the Earth. The Doppler shift for an observation is

$$\Delta\nu = \nu_\mathrm{obs} \left(\frac{v_r}{c}\right),$$

where ν_obs is the observing frequency at 140 MHz. We find Doppler shifts ranging from 2.6 kHz for the ELAIS-N1 observation with ID `L686962` to -2.3 kHz for the observation with ID `L769393` (see Table 5.1). Given the frequency resolutions of 97.66 kHz, these Doppler shifts are too small to significantly contribute to frequency offsets during visibility averaging. This is because ELAIS-N1 is relatively favourably positioned on the sky. When observing at lower ecliptic latitudes, Doppler shifts could reach values of up to about ±15 kHz for LOFAR observations. This might have severe effects when the observations are taken half a year apart, introducing ~30 kHz frequency offsets.

To avoid the need to apply Doppler shift corrections when applying SVA, it may be beneficial to create for instance smaller facets. These smaller facets can be averaged more in frequency before SVA, making the Doppler shifts relatively smaller compared to the frequency channel width. Alternatively, if it is known in advance that SVA is used for imaging deep surveys, it is best to schedule the observations strategically. This way, observations can be scheduled to minimise the introduction of large Doppler shifts between the individual observations. A potentially better solution is to apply default Doppler corrections during the measurement pre-processing phase, before any visibility data is calibrated. This approach would eliminate the need to account for Doppler corrections during data processing. However, this may also need to involve additional corrections for *uvw* coordinates.

5.6. Summary and conclusion

We have in this paper revisited a method called 'sidereal visibility averaging' (SVA) to enable ultra-deep imaging when combining multiple observations of a single pointing on the sky. This method takes advantage of the repetitive nature of baselines each sidereal day, allowing us to average calibrated visibilities from different observations at similar baseline coordinates. While this approach eliminates information about the time-varying flux of transient sources, it significantly reduces the number of visibilities to process during imaging, alleviating the computational bottleneck for deep imaging with multiple observations and lowering the long-term data archiving costs of calibrated visibilities. It can be used in addition to other data volume reduction methods, such as BDA and Dysco compression.

By testing the SVA algorithm with four previously calibrated datasets from de Jong et al. (2024), corresponding to images of two facets at a $0.6''$ resolution, we found that we could reduce the data volume by a factor of 1.8 and speed up imaging by a factor of 1.6 compared to standard imaging when we allow a 5% additional smearing increase towards the edge of the imaged facet. The improvements in data volume and computational time become larger when more observations are combined. For example, applying this method to the approximately 500 hrs of LOFAR data available for the ELAIS-N1 deep field, we estimate reductions in data volume of up to a factor of 28, while computing times may be decreased by a factor of around 10. For even larger projects, with over 3000 hrs of combined integration time, the improvements may reach up to a factor of 169 in data volume and a factor of 14 in computing time, while achieving imaging sensitivity of the order of a μJy beam^{-1} at 150 MHz. The computational time reductions are likely to further improve as the software becomes more optimised and more advanced hardware becomes available.

We also examined the effects of Earth's celestial motion on the baseline coordinates of the combined dataset. Although the baseline coordinates between the four observations are offset due to precession and aberration, these have a small effect on SVA. However, we anticipate that these effects could become problematic when combining observations taken many more years apart. Depending on the resolution, this could potentially introduce additional smearing effects. This issue can be addressed by adding more *uvw* points to the output dataset using finer time resolutions, which reduces the final computational speedup factors but maintains image quality. We also assessed the effects of Doppler shifts and found that for the utilised ELAIS-N1 data, they are too small to have a significant impact on the dataset after applying SVA. However, observations at different positions on the sky may necessitate additional frequency corrections, depending on the times of the year during which the observations were conducted. Ideally, to optimise baseline

coordinate binning accuracy and minimise Doppler shift effects, observations should be scheduled during the same period of the year and start at the same sidereal time.

We have demonstrated that SVA is an effective method for reducing data volumes and processing time for imaging calibrated visibilities. We believe this approach serves as an important building block for producing the deepest single-pointing images using current and upcoming interferometric radio instruments, as long the calibrated visibility data of the observations being combined are retained before applying SVA. This enables us to explore the Universe at unprecedented depths and spatial resolution.

Acknowledgements

This publication is part of the project CORTEX (NWA.1160.18.316) of the research programme NWA-ORC which is (partly) financed by the Dutch Research Council (NWO). This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-1287. This work is co-funded by the EGI-ACE project (Horizon 2020) under Grant number 101017567. RJvW acknowledges support from the ERC Starting Grant ClusterWeb 804208. FS appreciates the support of STFC [ST/Y004159/1]. LOFAR data products were provided by the LOFAR Surveys Key Science project (LSKSP; `https://lofar-surveys.org/`*) and were derived from observations with the International LOFAR Telescope (ILT). LOFAR (van Haarlem et al., 2013) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and which are collectively operated by the ILT foundation under a joint scientific policy. The efforts of the LSKSP have benefited from funding from the European Research Council, NOVA, NWO, CNRS-INSU, the SURF Co-operative, the UK Science and Technology Funding Council and the Jülich Supercomputing Centre. The use of the national computer facilities in this research was subsidized by NWO Domain Science.*

6

Advanced strategy for deep sub-arcsecond wide-field facet calibration with LOFAR

J.M.G.H.J. de Jong, L. Veefkind, R.J. van Weeren, R.J. Schlimbach, J.B.R. Oonk, D.N.G. Kampert, H.J.A. Röttgering, M. van der Wild

Abstract

We address key bottlenecks that limit the processing of a large number of observations of the same sky area for sub-arcsecond wide-field imaging with the International Low-Frequency Array (LOFAR). Our improvements are focused on three critical aspects: First, we establish a framework to streamline the data reduction of LOFAR datasets on HPC systems. Second, we refine and further automate the direction-dependent calibration by employing performance metrics, linking S/N to solution intervals, and integrating a neural network to identify image artefacts. Third, by revisiting the imaging strategy, we reduce the total required computational resources up to a factor of 3 to 4, depending on whether processing a single 8-hour observation or combining hundreds of hours of data. We demonstrate the robustness of our new data reduction strategy through sub-arcsecond imaging on 1 already successfully reduced dataset and 1 new challenging dataset from the ELAIS-N1 deep field. This work enables the processing of hundreds of hours of LOFAR data of the same pointing, paving the way for the creation of the deepest high-quality sub-arcsecond wide-field images with sensitivities on the order of a few μJy beam^{-1} at 144 MHz. We also outline additional development steps needed to establish an automated survey pipeline for sub-arcsecond wide-field imaging of other sky regions using LOFAR.

6.1. Introduction

Deep high-resolution wide-field surveys at low radio frequencies provide us with essential information about the evolution of our universe. These surveys can be done using the International LOw Frequency ARray (LOFAR) Telescope (ILT van Haarlem et al., 2013). This radio telescope is sensitive to detecting low-frequency radio waves between 10-80 MHz with the low-band antennas (LBAs) and between 110-240 MHz with the high-band antennas (HBAs). Recent efforts to automatically calibrate and image LOFAR observations with only the Dutch LOFAR HBAs have led to the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al., 2017, 2019, 2022; Williams et al., 2019) and the LoTSS-Deep Fields (Kondapally et al., 2021; Duncan et al., 2021; Tasse et al., 2021; Sabater et al., 2021; Best et al., 2023; Bondi et al., 2024; Shimwell et al., 2025), providing us with wide-field images of the northern sky at 144 MHz and 6″ resolution. The deepest of these maps is created with over ~500 hrs of observations of the ELAIS-N1 deep field, reaching a central sensitivity of about 11 μJy beam^{-1} (Shimwell et al., 2025). However, 40% of the image noise in this map is due to confusion noise. This can be mitigated by producing higher-resolution images through the inclusion of visibility data from all LOFAR stations across Europe. This extends baselines to ~2,000 km, enabling 20 times better resolutions, and increasing the telescope's collecting area as well (Varenius et al., 2015; Morabito et al., 2022a). The resulting sub-arcsecond resolution resolves much smaller angular structures, which unlocks a wealth of scientific opportunities to for instance study supernovae (e.g. Venkattu et al., 2023), AGN (e.g. Mahatma et al., 2023; Jurlin et al., 2024), galaxy clusters (e.g. van Weeren et al., 2024), or separate AGN activity from star-formation (Morabito et al., 2022b, 2025b). Nonetheless, the scalability of calibrating and imaging many observations with all LOFAR stations for ultra-deep wide-field imaging is currently constrained by the lack of a framework for automatic processing, the high computational costs, and challenges in achieving high-quality calibration solutions for both short and long baselines simultaneously.

Building on earlier successful work to process LOFAR data for sub-arcsecond resolution imaging (e.g. Moldón et al., 2015; Varenius et al., 2015, 2016; Jackson et al., 2016; Harris et al., 2019), the next step to advance the development of a sub-arcsecond imaging strategy for LOFAR was introduced by Morabito et al. (2022a), who focused on the general direction-independent calibration of the international stations followed by postage stamp imaging of targets. Sweijen et al. (2022c) extended this work to wide-field imaging, producing the first image that captured thousands of sources within a 2.5×2.5 deg field of view at a resolution of 0.3″. This was followed by de Jong et al. (2024), who conducted imaging with four observations of the same pointing of the ELAIS-N1 deep field, resulting in the detection of four times more sources within the same sky area compared to imaging only one

observation. In addition to increasing depth, they introduced automated direction-dependent (DD) calibrator selection, aiming to reduce the need for manual intervention. They also demonstrated that the current DD calibration strategy required additional attention since the calibration solutions, corresponding to the Dutch LOFAR stations and hence the shorter baselines, were of insufficient quality to enable high-resolution imaging. This issue was mitigated by introducing an additional ad hoc calibration step specifically for the Dutch LOFAR stations after completing all other calibration steps. While this solution reduced the problem, image artefacts across different facets of the images persist and therefore a refinement of the calibration strategy is required. Moreover, Sweijen et al. (2022c), Ye et al. (2024), and de Jong et al. (2024) demonstrated that the final imaging steps consume the majority of data processing costs, accounting for approximately 80% of the total expenses. Revising key aspects of the imaging strategy is therefore essential before scaling up sub-arcsecond wide-field imaging for deeper imaging. One promising approach to address these high computational costs when combining multiple observations of the same sky area is sidereal visibility averaging (SVA), which was recently revisited by de Jong et al. (2025). This approach takes advantage of the ability to average visibilities over repeating baseline tracks for each sidereal day, offering a solution to significantly compress data volume and thereby reduce computational demands.

Unlike for wide-field imaging at 6″ resolution with LoTSS (e.g. Shimwell et al., 2017, 2019; Mechev et al., 2017, 2018; Tasse et al., 2021), a fully developed framework to automatically process LOFAR data for sub-arcsecond wide-field imaging has not yet been realised. The lack of such a comprehensive framework is largely due to the exploratory nature of previous studies on sub-arcsecond wide-field imaging, which focused efforts on creating and refining individual processing steps and software. However, with most of the essential main building blocks for data processing for sub-arcsecond wide-field imaging in place, it is now possible to take the next steps in the development. Processing the large data volumes – approximately 4 TB of compressed data per observation using Dynamical Statistical Compression (Dysco; Offringa, 2016) – along with the intensive computational demands of hundreds of thousands of CPU core hrs (Sweijen et al., 2022c; de Jong et al., 2024), requires the implementation of automated data processing pipelines and workflows. Here, a pipeline is defined as a structured sequence of processing steps where the output of one step directly feeds into the next, whereas a workflow refers to a flexible, organised arrangement of interconnected tasks that manage the coordination and orchestration of those tasks. To achieve feasible wall times, these processes are best executed on a high-performance computing (HPC) cluster, capable of distributing the workload efficiently across multiple interconnected computing nodes, each equipped with numerous CPUs.

In this work, we establish the next step towards a framework to streamline the data reduction of LOFAR datasets on HPC clusters, revisit and automate the DD calibration, and develop strategies to reduce the required computational resources. We demonstrate the new calibration and imaging strategy on 2 LOFAR datasets from the ELAIS-N1 field, of which one has already been reduced by de Jong et al. (2024) and one new dataset. With hundreds of hrs of data available for this field with international LOFAR stations, we pave the way for ultra-deep imaging of a single pointing, reaching a sensitivity on the order of a few μJy beam^{-1} at 144 MHz.

In Section 6.2, we introduce a framework designed to manage large-scale data processing on high-performance computing infrastructures. This is followed by Section 6.3 where we discuss steps to further automate the self-calibration of LOFAR data with long-baselines, setting the stage for Section 6.4 where we present the optimised DD calibration and imaging strategy. In Section 6.5 we outline the data used to implement the new data reduction strategy. Section 6.6 presents our results, which are discussed along with future prospects in Section 6.7. Finally, we end our work with a summary and conclusions in Section 6.8.

6.2. Data processing framework

Efficient management of processing large data volumes in the order of a few to hundreds of TB with many individual steps on high-performance computing (HPC) systems requires a well-designed framework with workflows that orchestrate task execution, handle dependencies, and utilise computational resources across multiple nodes and CPUs and their cores efficiently. For this purpose, we use the combination of the Common Workflow Language (CWL; Amstutz et al., 2016; Crusoe et al., 2022), Toil (Vivian et al., 2017), and the Simple Linux Utility for Resource Management (SLURM; Yoo et al., 2003). This approach is effective because CWL, Toil, and SLURM seamlessly integrate to facilitate efficient workflow execution:

1. CWL defines how the tasks in a workflow are structured, including their inputs, outputs, computing resources, and dependencies.

2. Toil interprets the CWL workflow, orchestrates the execution of tasks, and interacts with the computing infrastructure.

3. SLURM handles the actual job scheduling and resource allocation on an HPC cluster.

For the data reduction in this work, we utilise Spider[1], a high-throughput data

[1] `https://doc.spider.surfsara.nl`

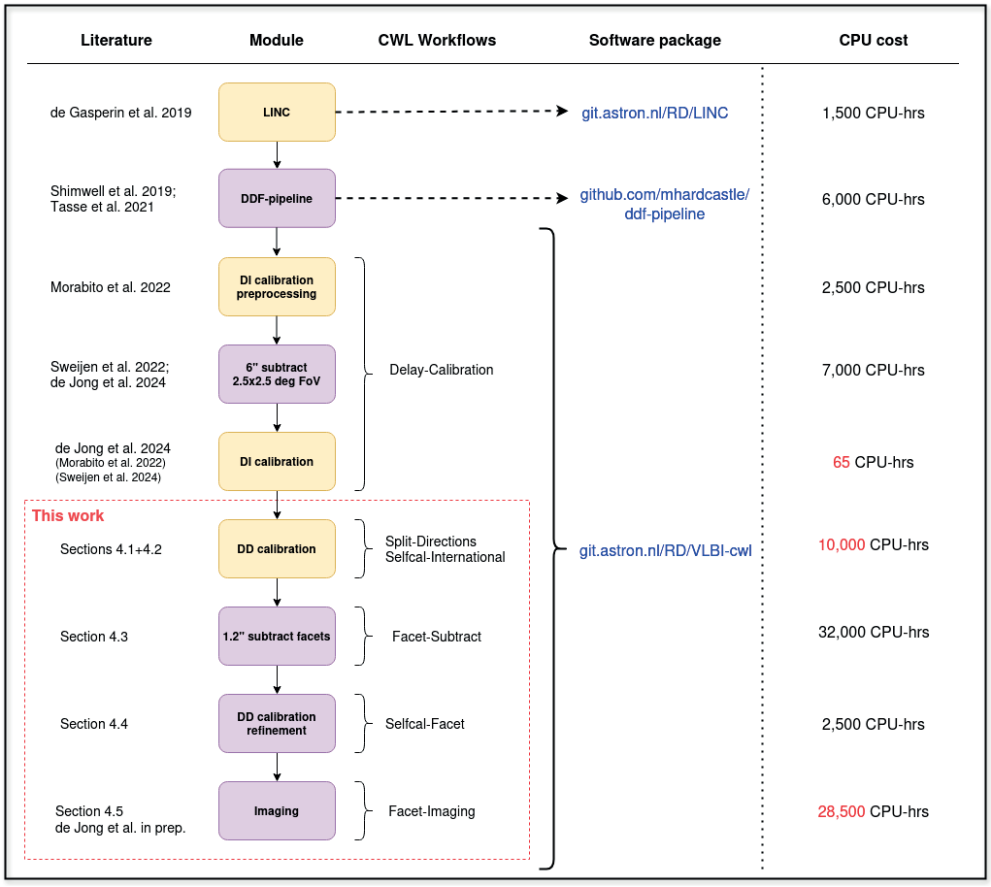


Figure 6.1: Diagram outlining the framework consisting of different modules with corresponding CWL workflows, software `git` directories, and computing requirements. The yellow modules are essential for postage stamp imaging, while the purple modules are additional modules necessary for wide-field imaging. The first 2 modules, the LOFAR Initial Calibration (`LINC`) and `DDF`-pipeline, are independent pipelines but are required to process LOFAR data for calibration of the international stations and for sub-arcsecond wide-field imaging. For the DI calibration step, we have put Morabito et al. (2022a) and Sweijen et al. (2022c) between brackets, since we refer here to the optimised step by de Jong et al. (2024) for ELAIS-N1. The CPU core hrs given in red are variable, as these will be smaller when more observations are combined for (ultra-)deep imaging (see Section 6.6.2.2) or when processing a different dataset with a different DI calibration strategy. The modules enclosed by the dashed red box represent the parts that we optimise in this work, while other steps are referenced in related literature. A more in-depth explanation of the different CWL workflows for postage stamp imaging will be discussed in van der Wild (in prep.).

processing platform provided by SURF[2]. We use AMD EPYC 7551 and AMD EPYC

[2] `https://www.surf.nl/`

7702P processor nodes, each with 60 cores, 0.5 TB and 1 TB RAM respectively, and about 4 TB of scratch storage on a local solid-state drive (SSD). A shared Ceph file system (CephFS) serves as central storage, from which data is copied to the local SSDs to increase the processing speed of I/O-heavy jobs. Once jobs are complete, the output is transferred back to CephFS to proceed to the next step, as Toil and CWL require a central shared point for communication between jobs with input and output dependencies.

The framework's overarching data processing workflow, reducing data from raw to image, is structured into distinct, self-contained modules. Each module can operate autonomously, enabling straightforward development, testing, and maintenance. This modular design enhances the workflow's scalability by allowing modules to be added, modified, and replaced independently, without having to directly affect the input of the next module. These modules are chained together as outlined in the second column in Figure 6.1, and can correspond to an entire pipeline (e.g. the `DDF`-pipeline[3]) or one or multiple CWL workflows. This modular approach allows users to customise their entire data reduction workflow by selecting the specific steps needed to achieve their desired output. The yellow modules in this figure correspond to the modules typically required to create postage stamp images (Morabito et al., 2022a), while for wide-field imaging, we use the additional purple modules (Sweijen et al., 2022c; Ye et al., 2024; de Jong et al., 2024). In Figure 6.1, we have provided an overview of the corresponding CWL workflows with the currently advised order for wide-field imaging, together with software packages and an estimate of the CPU requirements. We developed multiple CWL workflows because CWL has the flexibility to implement these as sub-workflows within an automated larger overarching workflow. This design also allows users to execute workflows step by step, providing the flexibility to manually inspect or modify the data output as needed. The modules and workflows enclosed in the red dashed box are the parts that are further improved in this work. In the discussion in Section 6.7.1 we further discuss opportunities to reorder some of the modules.

It is important to note that we measure CPU core hrs in this work as the number of allocated CPUs requested for a job per hour. This measure does not account for idle CPUs, such as when a job temporarily uses only one CPU for data transfer between nodes during a job. This way of measuring may therefore be less accurate than tracking actual CPU usage.

The required software for data processing is distributed using Singularity containers (Kurtzer et al., 2017).[4] These containers include standardised tools such

[3] `https://github.com/mhardcastle/ddf-pipeline`
[4] `https://github.com/tikk3r/flocs`

as `casacore`[5] (Casacore Team, 2019; CASA Team et al., 2022) for working with measurement sets – the standard data format for radio interferometric datasets; `astropy`[6] for general-purpose astronomy utilities (Astropy Collaboration et al., 2013, 2018); and `DP3`[7] for preprocessing LOFAR datasets (van Diepen et al., 2018a; Dijkema et al., 2023), including tasks such as flagging, averaging, and calibration. These Singularity software containers can be compiled for different AMD and Intel architectures, optimizing performance and compatibility across a wide range of HPC systems. This ensures efficient utilisation of hardware resources.

6.3. Automated long-baseline self-calibration

The ionosphere induces DD effects (DDEs) that lead to artefacts on scales from arcseconds to arcminutes (e.g. Intema et al., 2009; Smirnov, 2011b). To address these issues, we apply self-calibration on various calibrators distributed across the field of view. These sources correspond to facets in a Voronoi tessellation, where the calibration solutions are assumed to be constant (Schwab, 1984; van Weeren et al., 2016b). We aim to achieve sub-arcsecond resolution images, which necessitates finding calibrators with high enough S/N ratios at the longest baselines. This is more challenging towards higher resolutions because fewer sources have sufficient flux densities at $0.3''$ scales.

In this section, we revisit the calibrator selection from de Jong et al. (2024) and connect it to a metric for determining solution intervals, aimed at enhancing automatic self-calibration. Additionally, we introduce a neural network to assess self-calibration convergence. This helps to refine the automation of calibration strategies by identifying the optimal self-calibration cycle and set of automatically determined parameters.

6.3.1. Calibrator selection

de Jong et al. (2024) implemented a source selection method that exploits the fact that circular polarisation is a rare phenomenon at low radio frequencies, as demonstrated by Callingham et al. (2023). Therefore, the difference between right- and left-handed polarisation should be minimal for compact high S/N calibrators.[8] Consequently, the circular standard deviation, accounting for phase wrapping around π (e.g. Mardia, 1972; Fisher et al., 1993), is applied to the solutions that correct for the

[5] `https://casacore.github.io/python-casacore/`

[6] `https://www.astropy.org`

[7] `https://dp3.readthedocs.io`

[8] This is a simplified assumption ignoring the effects of polarisation leakage variations and the minor impact of differential Faraday rotation for different directions.

phase difference between both polarisation directions. The resulting values for the most distant international stations (excluding Dutch and German stations) provide a measure of the S/N on the smallest angular scales, as these stations correspond to the longest baselines. Having enough S/N is essential to correct for DDEs. de Jong et al. (2024) selected a solution interval of $\hat{\delta}_t = 10$ min for these corrections, since this value worked well for all their sources. Having a fixed solution interval helps to compare the S/N across different sources within the same field of view. In this work, we refer to this metric with the phasediff-score (denoted by variable $\hat{\sigma}_c$). While de Jong et al. (2024) used a phasediff-score threshold of $\hat{\sigma}_c < 2.3$ rad, we opt to split our calibrators in two groups:

- ☐ *Main facet calibrators*: $\hat{\sigma}_c < 2.0$ rad.
- ☐ *Weak facet calibrators*: $2.0 \leq \hat{\sigma}_c < 2.6$ rad.

The main facet calibrators define our facet layout (see Section 6.4.2 and Figure 6.2), while the second group consists of weaker 'secondary' DD-calibrators, which can be used to refine DD calibration (see Section 6.4.4).

The threshold of $\hat{\sigma}_c = 2.0$, which separates strong and weak facet calibrators, is based on a re-evaluation of the calibration quality in de Jong et al. (2024). This work showed that the selected DD-calibrators with scores above this threshold tend to exhibit less stable calibration behaviour. Therefore, a more conservative approach ensures the selection of generally more stable main DD-calibrators, with the option to incorporate weaker secondary sources to enhance image quality by cleaning more local DDEs after facet subtraction (see Section 6.4.3). An additional minimal distance criteria between calibrator candidates prevents having neighbouring facets with calibrator sources near the edge of the other facet. This potentially leads to DD artefacts crossing the facet boundary (this will be later demonstrated in an example in Section 6.6.1.1) or to having sources obtaining higher $\hat{\sigma}_c$ values due to picking up flux from bright neighbours. DDEs do not vary as much on small angular scales, which made us decide to take 0.15 degrees as a minimal distance value. If two or more sources are neighbouring within these distance thresholds, the source with the lowest phasediff-score is kept and the other one(s) is removed from the selection. The calibration of weak facet calibrators is the final step of our data processing strategy (as will be outlined in Section 6.4.4). This is not required in order to perform the final wide-field imaging, when for instance these weak calibrator are near an already properly calibrated main calibrator, or when the ionospheric effects on the data are mild. Therefore, users can decide to skip this step or manually enhance it by, for example, adding additional sources.

For the datasets used from the ELAIS-N1 field, we removed 2 sources from the main facet calibrators because they were located within 0.15 deg of another

calibrator. From the weak facet calibrators, we removed 7 calibrators, since those were within 0.2 deg from another calibrator. This yields 23 strong calibrators next to the already calibrated in-field calibrator, next to 8 weak facet calibrators. This is different to the 30 calibrators selected by de Jong et al. (2024). The facet layout corresponding to the main DD-calibrators, along with the positions of the main and weak facet calibrators, is shown in Figure 6.2.

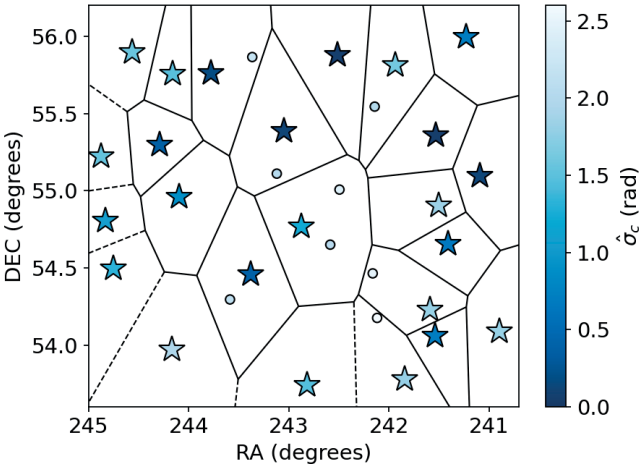


Figure 6.2: Main DD calibrator sources (stars) and weak facet calibrators (circles) for ELAIS-N1 with their respective phasediff-scores (σ_c) and corresponding facet boundaries in black.

6.3.2. Solution interval metric

Since the phasediff-scores are linked to the S/N of the longest international baselines, they also link to the optimal width of the solution interval to calibrate international stations. This enables us to derive automatically an appropriate value for the solution intervals for each calibrator source.

By varying the solution intervals between 1 min and 20 min for different calibrator sources when performing phase calibrations to correct for the differences between right-handed and left-handed polarisations, we empirically derived the following relation between the phasediff-scores (σ_c) and the corresponding solution intervals (δ_t):

$$\sigma_c \approx \pi \sqrt{1 - \exp\left(-\frac{\varsigma}{\delta_t}\right)}, \tag{6.1}$$

where ς is a unique constant value for each calibrator source in the data, which must be derived individually for each source and each observation. Increasing the solution intervals leads to a reduction in the variance of the phase solution data, while shorter intervals result in higher variance. The value for ς is a measure corresponding to the S/N of the source and therefore indicates the extent to which the solution interval can be narrowed to achieve a desired phase solution variance on the correction solutions.[9] Since the phasediff-scores are determined using calibration corrections on the data, we are effectively assessing the calibratability of the data and thus accounting for systematic effects, such as ionospheric variations unique to each observation. Consequently, Equation 6.1 provides a more precise method for determining solution intervals compared to theoretical approaches (e.g. Sob et al., 2021), which do not include direct information about the calibratability of a dataset.

For each source, we derive ς by evaluating $\sigma_c = \hat{\sigma}_c$ at a fixed solution interval of δ_t =10 min, such that we can derive a general expression for a metric for the solution intervals for different σ_c. By experimenting with self-calibrations of calibrator sources with different solution intervals, we found $\sigma_c = 1.75$ rad to correspond to δ_t values indicative of stable calibration. Using $\hat{\delta}_t = 10$ min and the fact that ς is constant, we derive

$$\begin{aligned} \delta_t &= \hat{\delta}_t \frac{\ln\left(1 - \left(\frac{\hat{\sigma}_c}{\pi}\right)^2\right)}{\ln\left(1 - \left(\frac{\sigma_c}{\pi}\right)^2\right)} \\ &\approx -26.92 \ln\left(1 - \left(\frac{\hat{\sigma}_c}{\pi}\right)^2\right). \end{aligned} \tag{6.2}$$

This expression functions as the metric for deriving solution intervals for different types of calibration (see Section 6.4.2). In Figure 6.3, we illustrate for two calibrator sources how the curve from Equation 6.1 fits well to simulated normally distributed data. The red star corresponds to the solution interval determined with Equation 6.2. The left source has a higher S/N than the right source, leading to a smaller δ_t for the left source compared to the right. The difference in S/N also accounts for the increased noise in the simulated data for the right source compared to the left source.

[9] This relation becomes particularly evident for $\delta_t \gg \varsigma$, where a Taylor expansion allows us to approximate the inverse square-root relationship $\sigma_c \propto \sqrt{\frac{\varsigma}{\delta_t}}$.

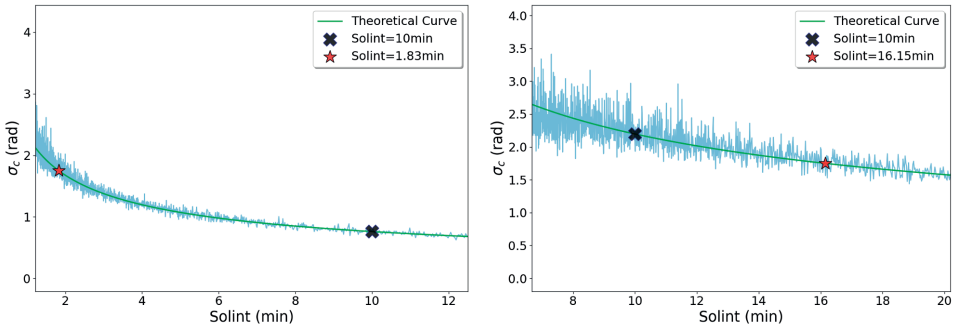


Figure 6.3: Phasediff-scores (σ_c) as a function of solution interval (δ_t) for two different calibrators. The black cross is the phasediff-score as calculated and described in de Jong et al. (Section 3.3.1; 2024) for $\delta_t = 10$ min, the green curve corresponds to the fit from Equation 6.1, and the red star is the best solution interval corresponding to $\sigma_c = 1.75$, according to Equation 6.2. The blue scattered line corresponds to determining circular standard deviations on randomly drawn normally distributed data with different solution intervals.

6.3.3. Automatic self-calibration assessment

By utilizing metrics that combine commonly used image metrics, such as the RMS background noise, the dynamic range, and by visually assessing images and solution inspection plots, Ye et al. (2024) and de Jong et al. (2024) found that many of the calibrator sources required fewer or in a few cases also more self-calibration cycles than the often selected 10 or 12 cycles. In certain instances, self-calibration even began to diverge after just 5 or 6 cycles. While the RMS and dynamic range provide useful information, they are not sufficient to fully determine if self-calibration has converged. For some calibrators, the RMS may for instance slightly increase during self-calibration when solving for amplitudes. In some cases, this is due to unstable diverging behaviour when a weaker calibrator absorbs signal from a strong nearby calibrator, while in other cases these are valid corrections to correct for small amplitude offsets. Similarly, the dynamic range may converge to favourable values during self-calibration, yet subtle image artefacts can persist and remain visible to the human eye. de Jong et al. (2024) also examined calibration solution stability, which indicates whether the calibration solutions converge across different cycles. While this confirms that self-calibration has indeed stabilised, it does not necessarily reflect whether the result is of sufficient quality, particularly if overly conservative self-calibration parameters are used.

To address this, we automate the decision-making process to remove the need for visual assessment. For this purpose, we adopt a pre-trained DINOv2 model (Oquab et al., 2023), utilizing register-based methods (Darcet et al., 2023) and with a vision transformer (ViT; Dosovitskiy et al., 2020) as the backbone. Specifically, we use

DINOv2 with the ViT-L backbone, comprising 300 million parameters, which has been trained using knowledge distillation on a larger DINOv2 model with a ViT-g backbone containing 1.1 billion parameters. The DINOv2 pre-training dataset (LVD-142) is a combination of existing curated datasets, including different ImageNet datasets (Russakovsky et al., 2015), Google Landmarks (Weyand et al., 2020), and images scraped from the internet by examining their similarity to the already existing images (without being duplicated).[10] The total combined dataset consists of 142 million images. The original DINOv2 model was trained in a self-supervised manner on this comprehensive dataset. We leverage this pre-trained model by substituting only the classification sub-model with a custom classifier. This classifier is a two-layer multi-layer perceptron (Rumelhart et al., 1986). In this classifier, a dropout is first applied to the feature extractor's output for regularisation. Subsequently, two linear layers are applied: the first layer, activated by a rectified linear unit (ReLU; Nair & Hinton, 2010), maps the feature extractor's output to the same dimensionality. The second layer reduces this output to a single value, followed by a sigmoid function to produce a pseudo-probabilistic output.

While the original DINOv2 model has mostly been trained on generic natural images curated from the internet, the self-supervised training scheme enables more generic and widely applicable features compared to feature extractors trained with a supervised training scheme (e.g. Huang et al., 2021). The primary advantage of transfer learning from the DINOv2 model is that, as long as the data modalities remain relatively similar, the extracted features are likely to be transferable (e.g. Gerace et al., 2022; Tahir et al., 2024). This allows us to achieve high performance by utilizing the well-trained feature extractor of the DINOv2 model, which is pre-trained on a large dataset, rather than training the feature extractor from scratch on a smaller dataset. By training a classifier that acts on these features (and potentially fine-tuning the feature extractor) on our domain-specific self-calibration images, we obtain a model that is robust for detecting artefacts in these images. This can be assessed by evaluating the model on test images that are unseen during the training stage. The effectiveness of utilizing DINOv2 as a base model on 'new' images has been demonstrated across various domains, including applications in medical imaging (e.g. Kundu et al., 2024; Song et al., 2024) and geological imaging (e.g. Brondolo & Beaussant, 2024).

We trained the model on 2,360 binary-labelled self-calibration images from different observations, including all calibrator sources considered by de Jong et al. (2024), Bondi et al. (in prep.), and Escott et al. (in prep.). These images include both high-S/N calibrator sources, weaker sources that were not selected for final DD

[10] This similarity has been determined using another pre-trained ViT, which embeds these images.

calibration, and sources that are resolved out and thus correspond to images dominated by noise. We further extend the dataset by applying data augmentations, specifically using random mirroring and rotations in multiples of 90 degrees. This enhances the model's capability to train on images associated with more unique and complex extended calibrator sources. The label *continue* ($P = 1$) is assigned to an image when significant artefacts and continued calibration is required. In all other cases, the images are labelled as *stop* ($P = 0$), which covers the following scenarios:

- ☐ Self-calibration has converged.
- ☐ The source is resolved out at high resolution, resulting in insufficient S/N for calibration.
- ☐ Artefacts are present but are not caused by the calibrator source in the image.

Figure 6.4 presents examples of images with labels from our training dataset. To improve regularisation we apply label smoothing as well (Szegedy et al., 2015). This is a technique that slightly reduces the *continue* class $P = 1$ to for instance $P = 0.9$ and adds the remaining probability (e.g., 0.1) to the *stop* class. This reduces overconfidence in the model's predictions, helping it generalise better by not focusing too strictly on exact class labels.

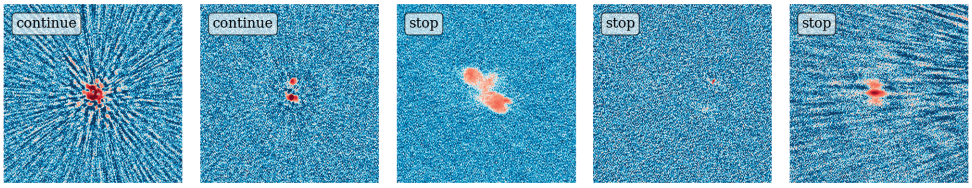


Figure 6.4: Images of sources at various stages of self-calibration. *Left:* Labelled *continue*, as no self-calibration has been applied and significant artefacts are present. *Middle left:* Labelled *continue*, after several cycles of self-calibration, though noticeable artefacts remain. *Middle:* Labelled *stop* since self-calibration has successfully converged. *Middle right:* Labelled *stop* because the source is resolved, and further self-calibration is unnecessary. *Right:* Labelled *stop* as artefacts from a nearby bright calibrator leak into the calibrator in the centre.

After training the model, we achieved an accuracy of 0.94, which corresponds to the confusion matrix given in Figure 6.5. This indicates that the model maintains high precision in correctly labelling the sources in our validation dataset and according to our visual inspections. Note, however, that the labelling of when to stop self-calibration remains subjective, as calibration is often never entirely perfect and some artefacts may persist in the image. Although expert labelling is involved, a certain level of bias is therefore inevitable.

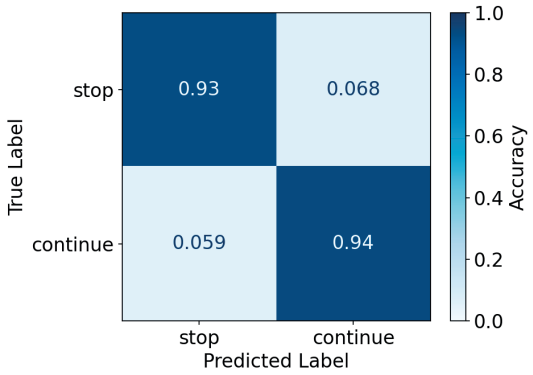


Figure 6.5: Confusion matrix with the fraction of true positives, true negatives, false positives, and false negatives.

Figure 6.6 displays examples of label predictions for calibrators across various self-calibration cycles. The sources in the top three rows all achieve convergence by the 5th cycle, with P values below 0.5. In each of these cases, we find that the model recognises subtle residual amplitude artefacts (4th column) that could still be corrected. In the fourth row, convergence is reached by the 3rd cycle, after which the label briefly increases, before reaching its lowest value in the 5th cycle. While the image shows notable improvement compared to the uncalibrated version, this example highlights the subjectivity of the labelling, as other astronomers might still find minor amplitude artefacts around the hotspots of this FRII source (Fanaroff & Riley, 1974) unsatisfactory. In the fifth row, we observe a source gradually approaching convergence by the 5th cycle, though it has not yet met the early-stopping criteria for self-calibration. The final row illustrates a point source that fails to converge to an artefact-free image, although the model detects slight improvements compared to the initial uncalibrated image.

Since the model is trained purely to classify individual images, it does not account for amplitude drifts or information from prior self-calibration cycles. Therefore, we set our self-calibration stopping criteria on the requirements that the image is labelled with $P < 0.5$, an RMS increase compared to the initial uncalibrated image of no more than 5%, and an improvement in dynamic range compared to the uncalibrated image. We also require the solutions to be stabilised between cycles as well. This is determined by subtracting the current phase solutions from the previous one and calculating the circular standard deviation of the difference, which must be less than 0.1 rad. This threshold is based on the results from the self-calibration inspections from de Jong et al. (2024).

Our stopping criteria not only reduce the computational cost of self-calibration but also aid in identifying optimal calibration settings for our sources. This parameter optimisation allows us to fine-tune the automatic settings, as described in Section 6.4.2. A potential future application is to integrate these automatic self-calibration stopping criteria to dynamically adjust parameters during self-calibration, using the different stopping criteria to decide to refine settings after each cycle. As this approach is still in its early stages and requires validation of data from other observations, further potential applications are discussed in Section 6.7.2.

6.4. Improved DD calibration and imaging strategy

We follow the same data reduction strategy up to the primary in-field calibration as described in de Jong et al. (2024, Section 3.1 and 3.2). We only replaced the software from the first calibration step using `prefactor` (van Weeren et al., 2016b; de Gasperin et al., 2019b) by its successor the LOFAR Initial Calibration (`LINC`[11]). `LINC` uses CWL and therefore integrates well into our data processing framework. In this section, we detail DD calibration and imaging improvements, aimed at enhancing image quality and reducing computing costs to enable scalable data reduction for ultra-deep imaging. The full new data processing strategy discussed in this Section is presented in Figure 6.7 and will be referred to throughout the text.

The different calibration steps utilise `facetselfcal`[12] (van Weeren et al., 2021), which is an advanced flexible low-frequency self-calibration software package. This software integrates the Default Preprocessing Pipeline (`DP3`; van Diepen et al. 2018a; Dijkema et al. 2023) and `WSClean` (Offringa et al., 2014) for (self-)calibration on a source. `facetselfcal` executes multiple self-calibration cycles, adjusting the source model iteratively before proceeding to the next cycle.

6.4.1. Dutch station calibration

For the final imaging, de Jong et al. (2024) created datasets for each individual facet, as this allows to perform parallel imaging across facets and reduces overall computing wall-time. Before imaging, sources outside each facet were subtracted using $1.2''$ resolution model images with corresponding calibration solutions. However, during imaging, they found that the DD calibration solutions for the Dutch core and remote stations were of poor quality at lower resolutions. To address this, an additional calibration refinement step was introduced on the datasets corresponding to each facet, selecting only the Dutch core and remote stations. While

[11] `https://linc.readthedocs.io`
[12] `https://github.com/rvweeren/lofar_facet_selfcal`

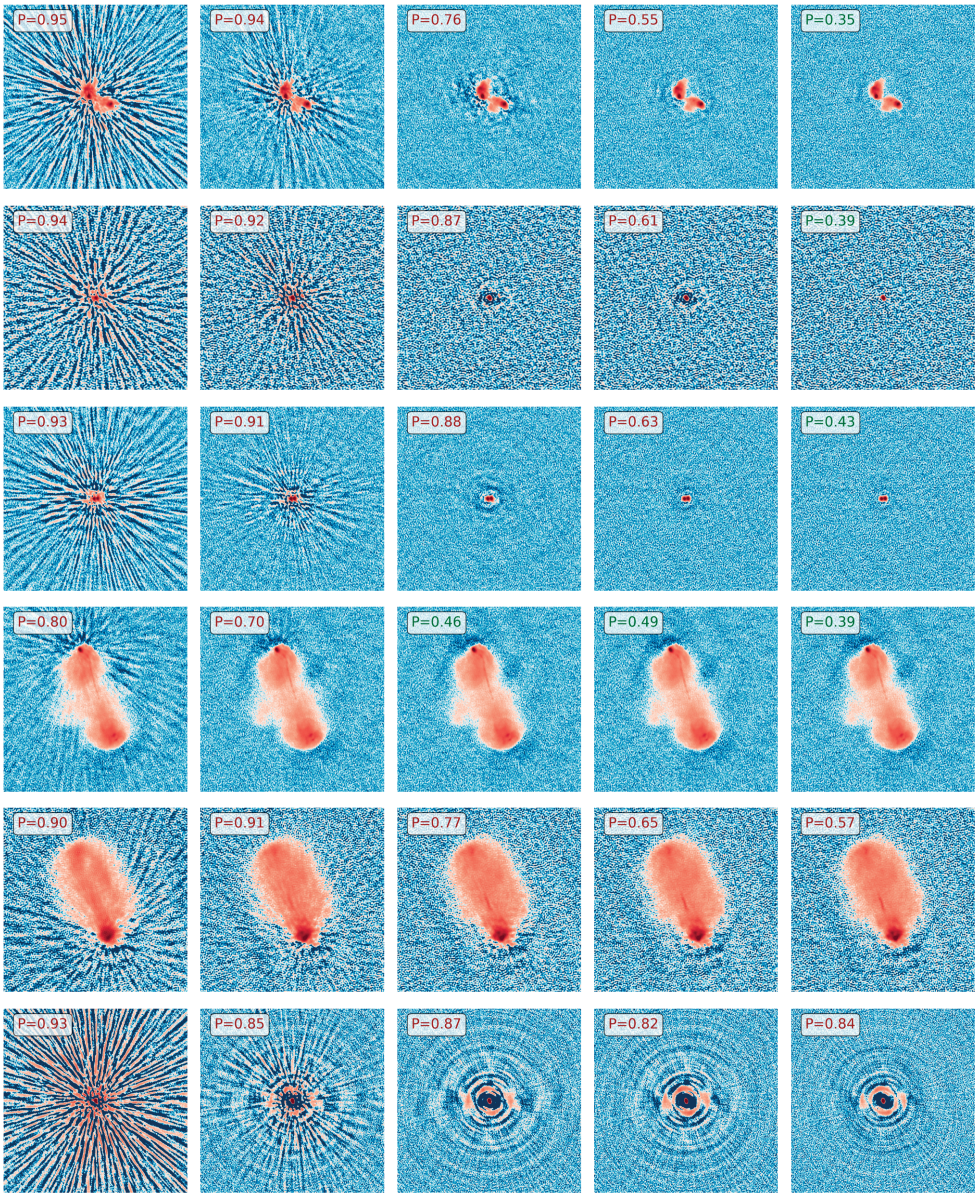


Figure 6.6: Examples of self-calibration images with early-stopping scores from our neural network. The *rows* represent different sources, while the *columns* correspond to self-calibration cycles. $P < 0.5$ values in green indicate successful convergence and will make self-calibration stop, while $P \geq 0.5$ in red suggests poor image quality, indicating that further self-calibration is required.

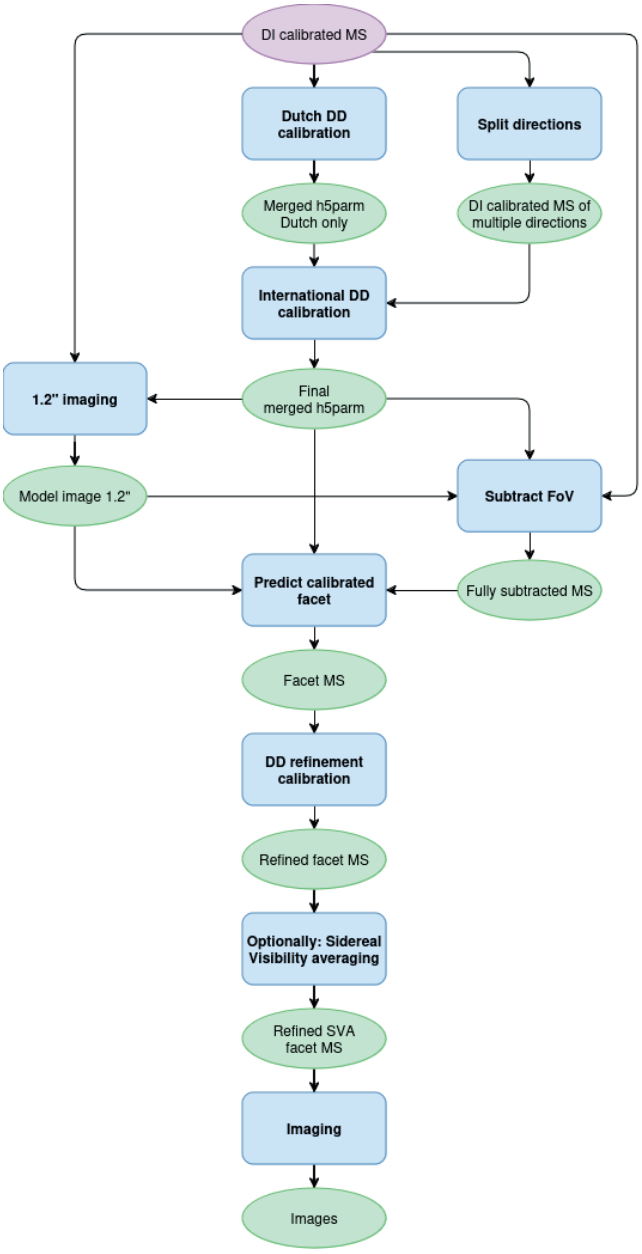


Figure 6.7: Flowchart corresponding to the full DD calibration discussed in Section 6.4. The workflow starts with DI-corrected uv-data and ends with images. Purple ovals are input data, blue boxes are operations on the data, and green ovals are output data.

this improved image quality around the calibrator source of the facet, it did not eliminate DDEs originating from neighbouring facets. This was because the refinement was applied after splitting the facets and subtracting external sources using the 1.2″ resolution model images, which were created with poor Dutch calibration solutions. Hence, it is essential that poor Dutch calibration solutions are corrected before any subtraction is performed.

To enhance the calibration solutions for short baselines, we introduce an additional DD calibration step for the Dutch core and remote stations before performing subtraction of sources outside the facets. We already obtained DD solutions from the `DDF`-pipeline[13], which utilises `KillMS`[14] (Tasse, 2014a,b; Smirnov & Tasse, 2015) to obtain phase and amplitude corrections. However, we have already corrupted our data with the DI in-field calibration solutions for both the Dutch and international stations, which makes pre-applying Dutch DD solutions from the `DDF`-pipeline to the closest calibrator selected in Section 6.3.1 no longer valid. An alternative approach for incorporating the `DDF`-pipeline DD solutions is discussed in Section 6.7.1.

We opt instead for using a new joint-solve feature from `facetselfcal`, which performs self-calibration in multiple directions simultaneously using `DP3`. For this, we average our datasets to 16 sec and 195.36 kHz after removing the international stations. This reduces the data volume by a factor ~300 and therefore reduces the computing resources required for a multi-directional joint-solve while having enough time and frequency resolution to calibrate for the changing ionospheric effects. Recent updates in `DP3` and `WSClean` have also enabled Stokes I-only data processing, which implies that 4 times less data has to be processed, saving a significant amount of RAM and computational time. Since our selection (see Section 6.3.1) is based on the S/N for the international stations, we select calibrators that have a peak intensity above 85 mJy beam^{-1} in the 6″ ELAIS-N1 catalogue from Sabater et al. (2021), as we find this threshold to correspond to stable calibrators. This leaves us with 15 calibrators which we calibrate with the following strategy:

1. We first perform phase calibration focused on correcting the Dutch remote stations by taking solution intervals of 32 seconds for sources with peak intensities below 300 mJy beam^{-1}, and 16 seconds for sources with peak intensities above this threshold. We also use a frequency smoothness kernel of 20 MHz, which from experience has shown to result in stable solutions at 6″. After this step, we reset the Dutch core station solutions to amplitudes equal to 1 and phases to 0, such that only the Dutch remote stations get calibration corrections.

2. We then carry out phase calibration for slower phase variations from the Dutch

[13] `https://github.com/mhardcastle/ddf-pipeline`
[14] `https://github.com/saopicc/killMS`

core stations, using longer solution intervals of 64 seconds and a broader frequency smoothness kernel of 40 MHz. This approach works well for the Dutch core stations because their shorter baselines are less sensitive to rapid phase variations and primarily capture larger-scale structures.

3. After doing 3 cycles, we have calibrated the phases well enough to also include calibration for the combination of phases and amplitudes as well. This applies longer solution intervals, compared to the previous phase calibration steps, since phases tend to vary on much shorter time scales compared to amplitudes. The solution intervals are 40 min for sources with peak intensities below 300 mJy beam^{-1}, while 20 min for sources with peak intensities above this threshold. We use a frequency smoothness kernel of 10 MHz.

All calibration steps are polarisation-independent, as polarisation corrections have already been applied through the `DDF`-pipeline full-Jones DI solutions and during a full-Jones in-field calibration step (see Section 3.2.3; de Jong et al., 2024). We also use a *uv*-cut of 10λ which corresponds to largest angular scales of about ~5.7 deg. The used settings are optimised on an empirical basis for ELAIS-N1. We expect these to work well in the general case, but for further optimisation, these settings can be adjusted, using a configuration file. This step corresponds to the 'Dutch DD calibration' box and its merged output solutions in Figure 6.7.

6.4.2. International station calibration

With the improved Dutch DD calibration solutions, we proceed to calibrating the international stations. We first create datasets of each of the 24 calibrator sources, selected by the metrics in Section 6.3.1, by phase-shifting to the centre of the source. We then average the datasets to 32 sec and 390.72 kHz. These averaging settings were shown by de Jong et al. (2024) to be effective in averaging out signal from nearby sources, while providing sufficient time and frequency resolution to correct fast phase variations for the brightest calibrator sources, without introducing bandwidth or time smearing in the self-calibration images. Subsequently, we apply on each dataset the Dutch calibration solutions from the nearest of the Dutch calibrators from Section 6.4.1. To suppress the signal of nearby sources at short baselines and reduce the computational cost, we adopt, similar to previous works (e.g. Moldón et al., 2015; Morabito et al., 2022a), a phase-up of the Dutch core stations into a superstation. Since the calibration solutions for the Dutch core and remote stations have already been applied, the goal is now fully focused on calibrating the international stations. For self-calibration, we use again `facetselfcal`.

During our analysis and testing, we observed that when performing self-calibration with all stations, including the Dutch phased-up stations, the Dutch core and re-

mote stations drifted, despite having their initial optimised calibration solutions applied, to the same poor solutions previously identified in de Jong et al. (2024). This drift is likely caused by flux from other high S/N sources in the field leaking into the signal from the target, thereby contaminating and corrupting the self-calibration process. This effect is particularly pronounced on the short baselines from the Dutch stations, despite the use of the phase-up of the core to partially mitigate this effect and aid convergence of the calibration for the international stations (Morabito et al., 2022a). While alternative solutions exist, by for instance drawing boxes around calibrators and predicting and extracting all sources outside these boxes (e.g. van Weeren et al., 2021; de Jong et al., 2022), these methods are computationally far too expensive for our large data volumes. Instead, we need to address this by ensuring that the already obtained solutions for the Dutch core stations and some of the Dutch remote stations change as little as possible after each self-calibration cycle and by applying large *uv*-cuts of $\geq 20,000\lambda$, corresponding to angular scales smaller than $10.3''$. These baseline cuts correspond to smaller angular scales, more compact source models during self-calibration, and focus on corrections for the longest baselines.

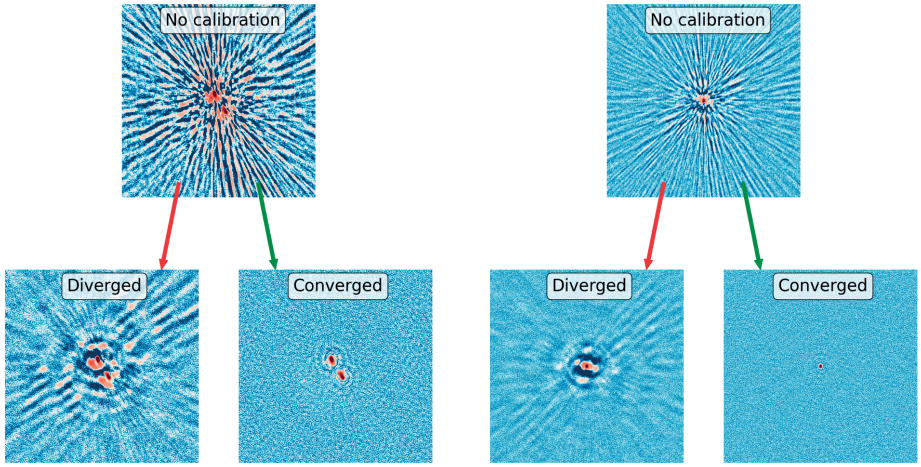


Figure 6.8: Self-calibration stability for various calibration parameters. The top row displays two images of uncalibrated sources. Below, a left and right image illustrate the corresponding images after performing self-calibration on the uncalibrated sources, where the calibration either diverged or successfully converged based on different parameter settings. The difference in the settings between the diverging and converging images was dependent on changing the *uv*-cut settings and by incorporating an additional round of polarisation-independent phase calibration for stations and resetting the Dutch core and a few additional remote stations, as we further explain in Section 6.4.2.

Starting with the parameter settings from de Jong et al. (2024) and utilizing

the solution interval metric δ_t (Equation 6.2) along with the image quality-based stopping criteria from the neural network discussed in Section 6.3.3, we optimised our calibration settings for different calibrator sources. To preserve the good calibration solutions from the DD Dutch calibration, we first simply reset all Dutch calibration solutions to the original pre-applied solutions obtained in Section 6.4.1, after self-calibration. However, instabilities during calibration and the emergence of spots with negative signals around compact high S/N sources highlighted the need to allow more flexibility for Dutch stations in adjusting their solutions. This is likely due to the fact that the Dutch stations have only a quarter of the collecting area compared to the international stations. Hence, for the most compact sources, the self-calibration process may run into the situation that S/N gets too low for the short baselines from the Dutch stations, while still being sufficient for calibrating the long baselines with the international stations. Two examples of different diverging and converging behaviour during self-calibration for different parameter settings are presented in Figure 6.8 and demonstrate the importance of careful parameter tuning.

The self-calibration divergence issues led us to explore an approach that involves calibrating with multiple cycles for different station groups linked to the brightness of the source. The options from `facetselfcal` provide the possibility for calibrating specific stations by resetting the solutions from station groups after each self-calibration cycle. The reset ensures that solutions for these stations are unchanged compared to the starting solutions. This provides a workaround for the limitation of the `DP3` solvers, which cannot apply different time intervals for different stations. This enables us to obtain accurate calibration solutions while minimizing the degrees of freedom by reducing the number of solutions needed to correct the data effectively. After experimenting with various parameter settings in a grid search and selecting the best-performing configuration based on the automatic image and solution quality-based stopping criteria, we identified the following strategy as the most effective:

1. We perform the first calibration for fast wrapping phases of the most distant international stations by calibrating these using solution intervals with size $\sqrt{\delta_t}$ (using Equation 6.2) and frequency smoothness kernels of 8 MHz. The solution intervals have a minimal size equal to the time resolution of the datasets at 32 sec and a maximum size of 3 min. After this calibration step, we reset the solutions of all Dutch core and remote stations and German stations closest to the Dutch stations (DE601HBA and DE605HBA).

2. With the first phase solutions for most of the international stations, we add another round of phase calibration but now with larger solution intervals with

size $\sqrt{2\delta_t}$ and frequency smoothness kernels of 10 MHz. The solution intervals have a minimal size equal to the time resolution of the datasets at 32 sec and a maximum size of 5 min. After calibration, we reset the solutions for all Dutch core and remote stations.

3. The following step only continues for sources with solution intervals $\delta_t < 3$ min, where we add another round of phase calibration for a specific (sub-)group of Dutch remote stations. Since this step involves solving for some of the Dutch stations, which correspond to shorter baselines, we use larger solution intervals and frequency smoothness kernels compared to the previous steps. The solution intervals are set to $2\sqrt{\delta_t}$ and the smoothness kernel to 15 MHz. For the brightest sources, with $\delta_t < 0.3$ min, the best results were obtained by resetting only the solutions of the Dutch core stations, which implies that all remote stations can be freely adjusted during self-calibration. For sources with $0.3 \leq \delta_t < 1$ min, we limit the adjustments for remote stations by resetting the solutions of all Dutch core stations and the five remote stations closest to the Dutch core (RS106HBA, RS205HBA, RS305HBA, RS306HBA, RS503HBA). Lastly, for sources with $1 \leq \delta_t < 3$ min, we reset the same remote station, including a few more distant remote stations (RS307HBA, RS406HBA, RS407HBA). This approach helps balance image quality and solution stability by including only specific groups of Dutch remote stations, selected based on their solution interval, which directly relates to their S/N.

4. Finally, after performing 3 cycles with only phase calibration, we also calibrate for both phases and amplitudes together but with larger solution intervals of $20\delta_t$ with a minimum of 18 min, since phases vary on much shorter time scales than amplitudes. This step is entirely skipped when $20\delta_t > 4$ hrs. The smoothness kernels depends on the δ_t, as we set this to 8 MHz if $\delta_t < 1$ min, 10 MHz if $1 \leq \delta_t < 3$ min, and 12 MHz otherwise. We reset in this step the solutions from all Dutch core stations and the same specific sub-group of remote stations as for the previous step if $\delta_t \geq 3$.

All these calibration steps are polarisation-independent, as we have already performed polarisation corrections during in-field calibration (see Section 3.2.3. in de Jong et al. (2024)). Since calibration for the Dutch stations has already been performed, the short Dutch-only baselines have already been corrected, which allows us to use large *uv*-cuts to focus on calibrating against a more compact sky model corresponding to long baselines (e.g. the international stations). We found by experimenting, using the automatic self-calibration quality stopping criteria, that varying the *uv*-cut based on their S/N led to better results as well. This is likely because when a source has a high S/N at long baselines, it is sufficient to construct

the sky model during self-calibration with strong signal at smaller angular scales. By doing so, we mitigate the effects of signal from other high S/N sources in the field leaking into the target source, which we already identified to corrupt the calibration solutions in particular at shorter Dutch baselines. While for less compact high S/N calibrators, we need more baselines with enough signal to reach convergence, which requires a smaller *uv*-cut. Linked to the freedom we give to the Dutch remote stations to vary in the 3rd step of our calibration strategy, we found the following relation between the *uv*-cut and δ_t to work well:

$$uv\text{-cut} = \begin{cases} 40,000\lambda & \text{if } \delta_t < 0.3 \\ 35,000\lambda & \text{if } 0.3 \leq \delta_t < 1 \\ 25,000\lambda & \text{if } 1 \leq \delta_t < 3 \\ 20,000\lambda & \text{if } \delta_t \geq 3 \end{cases}$$

A *uv*-cut of 20,000λ corresponds to a largest angular scale of approximately 10.3″, while a *uv*-cut of 40,000λ excludes most remote-remote baselines and corresponds to a largest angular scale of around 5.2″. Thus, even though we do not reset the remote stations for the brightest sources in step 3, we constrain their source model by applying a larger *uv*-cut to focus on correcting smaller angular scales. In this way, we balance, based on the S/N at the longest baselines, the freedom of the Dutch remote stations to be adjusted with a constraint on the baseline length.

For deciding on early-stopping during self-calibration, we utilise the neural network in combination with the other quality metric assessments discussed in Section 6.3.3. We use between 3 and 20 self-calibration cycles, providing a suitable range to ensure self-calibration convergence before proceeding to the next step in the data reduction process. In Section 6.4.4, we further refine calibration and provide a larger maximum number of self-calibration cycles to reach complete convergence. Unlike de Jong et al. (2024), we do not need to perform self-calibration with all observations combined since this will be carried out in the final self-calibration step. This allows in this intermediate calibration step for embarrassingly parallel processing over each calibrator source and each observation, reducing the wall time when enough CPU cores are available. The step discussed in this Section corresponds to the 'International DD calibration' box and its merged output solutions in Figure 6.7.

6.4.3. 1.2″ facet subtraction

Producing a typical 0.3″ wide-field image requires a wall-time of about 1.5 months per 8-hrs observation when using the 'standard' facet-mode from `WSClean` on a

single node. To reduce wall-time, we divide, similar to Sweijen et al. (2022c) and de Jong et al. (2024), the full dataset into smaller subsets, each corresponding to a single facet, enabling parallel processing of all facets across an HPC cluster up to the final single-facet imaging stage. This approach enables more data averaging and therefore accelerates imaging through parallelisation, while also allowing for SVA as this can only be performed with fully calibrated facets across multiple observations (see Section 6.4.5). To avoid calibration errors introduced by sources outside of the facet, it is crucial to subtract all sources that do not correspond to the current facet. de Jong et al. (2024) associated this step as part of the final imaging, as it utilises imaging software for prediction of sources within the field of view and is the final part of the pre-processing stage before the actual imaging. This step accounted for 76% of the total imaging costs, which represents 62% of the overall computational costs, including all calibration steps. The remaining 24% of the imaging costs were attributed to imaging the facets using all observations combined, as discussed in Section 6.4.5. This implies that reducing the computational cost at this stage can significantly lower the overall computational cost of the total data processing.

To enhance the computational efficiency of the source subtraction step before imaging, we introduce an improved method to make datasets of each facet. Instead of copying the full datasets for each facet and masking a facet and subtracting all sources outside this facet, we first create a dataset for each observation where we subtract all sources in the entire field of view. For this subtraction step, we use the improved $1.2''$ resolution model images of each observation, which result from imaging with our improved Dutch calibration solutions. This 'empty' dataset is then copied for each facet, whereafter we add back the sources corresponding to that facet, as illustrated in Figure 6.9. This approach eliminates the need to repeatedly predict the same sky, as was done with the original method from de Jong et al. (2024). All these steps can be done in parallel for each facet, each frequency subband, and each observation. To minimise I/O overhead, the data volume may be further reduced using stricter Dysco compressions (Offringa, 2016), as outlined in Section 6.6.2.1. In Section 6.6.2.2, we also discuss the significant computational cost improvements from our new approach. The subtraction steps correspond to the 'Subtract FoV' and 'Predict calibrated facet' boxes and their corresponding output in Figure 6.7.

6.4.4. Final calibration refinement

Since the sources with low phasediff-scores ($\sigma_c < 2.0$ rad), indicating high S/N, are now contained in separate datasets corresponding to individual facets, we have mitigated the interfering effect that high S/N calibrators have on each other when they are not corrected for DDEs and are contained in the same dataset (see Section

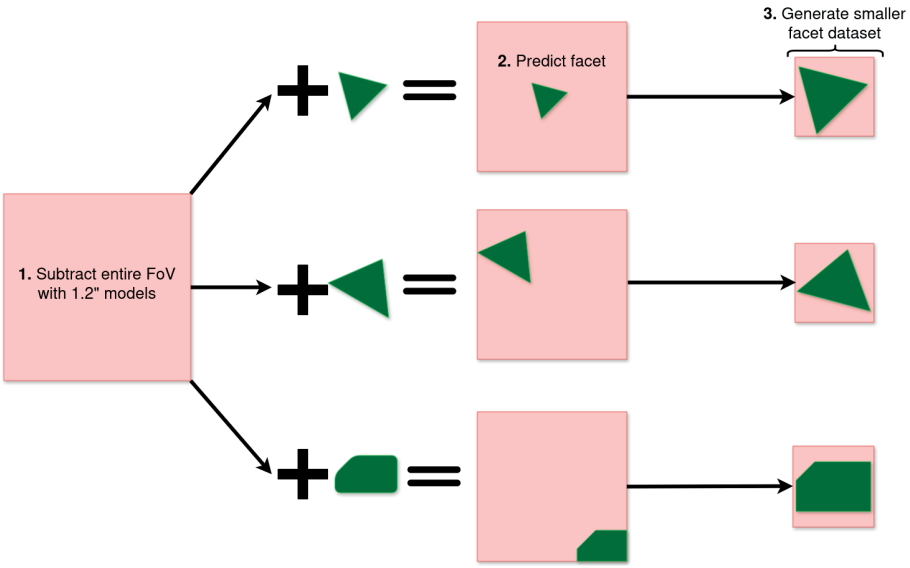


Figure 6.9: Updated subtraction and prediction strategy. The first large box represents a fully subtracted dataset, which has been created after predicting and subtracting sources with a DD-corrected 1.2″ model image. Next, sources within a polygon-shaped region are predicted and added back, creating a dataset with sources inside the facet while sources outside are subtracted. Finally, the data is phase-shifted to the centre of the polygon, and after applying beam corrections, solutions, and averaging, the smaller facet data is imaged.

6.4.2). This enables us to apply a final calibration step to improve calibration accuracy for all baselines and all observations together.

Similar to de Jong et al. (2024), we calibrate the Dutch stations by performing self-calibration at 6″ on the entire facet. Doing this at 0.3″ would in theory also be possible but is currently computationally too expensive, while at 6″the data can be averaged by a factor of 8 in both time and frequency and the international stations can be ignored, which reduces the computational cost significantly. Given that for facets with calibrators with solution interval metrics of $\delta_t > 3$ min the corrections on the Dutch and remote were left unchanged (see Section 6.4.2), we do not have to perform this step for those facets. Adopting the same recipe as de Jong et al. (Section 3.3.5; 2024), we proceed as follows, using a *uv*-cut of 750λ.

1. The first calibration step focuses on solving for 'fast' phase changes for the Dutch remote stations, using a solution interval of 1 min and a frequency smoothness kernel of 10 MHz, after which we reset the Dutch core stations.

2. Next, we apply another solve for phases, using a solution interval of 5 min and a larger frequency smoothness kernel of 20 MHz, but without resetting solutions.

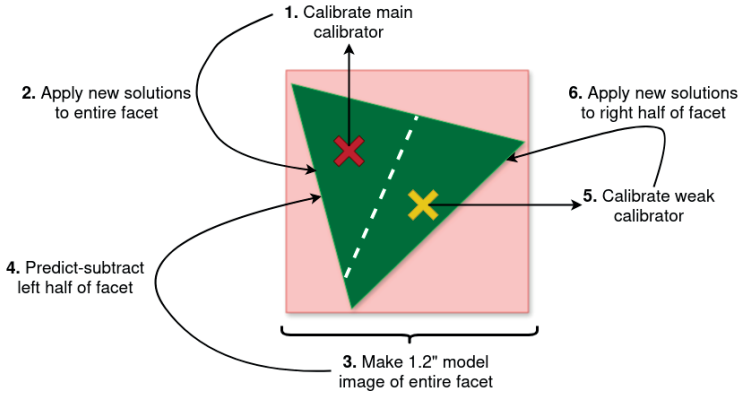


Figure 6.10: Additional facet calibration refinement for facets with weaker calibrators with low phasediff-scores ($2.0 \leq \sigma_c < 2.6$ rad). This calibration involves several steps. First, the main facet calibrator is calibrated using all observations combined. The derived solutions are then applied to the facet, and a $1.2''$ model image of the entire facet is created. With this model, the left half of the facet is predicted and subtracted, where both halves are defined with a Voronoi tessellation. This allows for the calibration of a secondary, weaker calibrator, after which the new solutions are applied to the remaining half of the facet. Finally, both halves of the facet are imaged.

3. After having done 2 rounds of self-calibration with only phase calibration, we perform a phase and amplitude correction with a solution interval of 30 min and a frequency smoothness kernel of 15 MHz as well.

After pre-applying the resulting solutions to the data, we employ a final round of self-calibration, where we first perform phaseshifts back to the centre of the calibrator source corresponding to each facet before averaging the datasets back to 32 sec and 390.72 kHz. Our calibration strategy follows a similar approach to what we described in Section 6.4.2, but with slightly modified parameters and different station resets.

1. We begin by calibrating the fast-wrapping phases of the international stations using solution intervals of size $\sqrt{\delta_t}$ and frequency smoothness kernels of 5 MHz. The solution intervals range from a minimal size equal to the time resolution of the datasets at 32 sec to a maximum size of 3 min. After this calibration step, we reset the calibration solutions for all Dutch core and remote stations.

2. We now perform phase calibration with all stations together, using a solution interval of $2\sqrt{\delta_t}$, and a frequency smoothness kernel of 20 MHz. The solution intervals range from a minimum of 64 sec to a maximum of 10 min.

3. Finally, after performing two cycles of phase-only calibration, we proceed with calibrating both phases and amplitudes using solution intervals of size $20\delta_t$, with a minimum interval of 18 min. This step is skipped if $20\delta_t > 4$ hrs. The smoothness kernel depends on δ_t, set to 5 MHz if $\delta_t < 1$ min, 10 MHz if $1 \leq \delta_t < 3$ min, and 15 MHz otherwise.

All steps are performed using polarisation-independent calibration, with the *uv*-cut constrained to 20,000λ. This helps to avoid potential issues from an incomplete high-resolution sky model, as remaining sources within the full facet but outside the smaller imaging box are not imaged or included in the model. We assess self-calibration convergence using our neural network model in combination with the convergence criteria described in Section 6.3.3, setting a minimum of 5 cycles and a maximum of 20. This is the first step where for deep imaging, the data from all observations are combined together, ensuring proper alignment before final imaging. This alignment is essential for correcting amplitude scale differences and preventing image blurring from small astrometric offsets between observations.

Finally, there is an optional step to include calibration of the weaker 'secondary' DD-calibrators, which were defined in Section 6.3.1 as sources with phasediff-scores of $2.0 \leq \sigma_c < 2.6$ rad. These sources may not have high enough S/N to have their own facet during the previous stages of the calibration, but they may still suffer from DDEs that can be corrected through an additional refinement step. This depends largely on the severity of the ionospheric conditions, which varies per observation. However, because the facet datasets are smaller at this stage, these refinement steps are significantly less computationally expensive than earlier calibration and prediction steps. In this refinement, we first create a $1.2''$ mode image of the entire facet and apply a Voronoi tessellation, using the positions of the main facet calibrator (which has already been corrected and solutions have been applied to the data) and the weaker calibrators. The area corresponding to the main facet calibrator is predicted and subtracted, followed by a phase-shift to the weaker calibrator before further calibration, using the same method as for the main facet calibrators, skipping the $6''$ full facet calibration. This process splits the facet into two (or more) sub-facets, where the weaker calibrator benefits from the pre-applied solutions of the primary calibrator, allowing for further refinement of local DDEs. Figure 6.10 illustrates this process, beginning with the calibration of the primary calibrator and showing the subsequent steps leading to the final calibrated data. In the case of multiple weaker calibrators, this can be easily extended to multiple facet splits, where all weaker calibrators have their own local solutions. The entire calibration refinement step, as described in this subsection, corresponds to the 'DD refinement calibration' box and its dataset output in Figure 6.7.

6.4.5. Final imaging

Imaging with multiple observations of the same field is typically done by concatenating the time axis from all observations, leading to an increase in data volume proportional to the number of observations. Consequently, the computational cost of imaging increases approximately linearly with the number of observations. However, we can exploit the fact that baseline tracks repeat every sidereal day. Since each observation is conducted on a different sidereal day, we can apply SVA, as described by de Jong et al. (2025) after having fully calibrated the facet datasets. They highlight that when combining observations separated over several years and conducted at different times of the year, offsets in the baseline coordinates must be accounted for by selecting a higher time resolution than the input datasets. These offsets are influenced by celestial motions, including aberration, precession, and nutation. The algorithm includes functionality that determines the optimal time resolution by considering image size, resolution, and baseline coordinate offsets. This minimises time smearing while significantly reducing the data volume compared to the total input. Although frequency effects like Doppler shifts are present, de Jong et al. (2025) found these to be negligible for ELAIS-N1 and therefore not a concern. The imaging step corresponds to the optional 'sidereal visibility averaging' and 'imaging' boxes with the images as output product in Figure 6.7.

For the final imaging, we follow the same approach as de Jong et al. (2024), using `WSClean`'s `wgridder` module (Offringa et al., 2014; Arras et al., 2021; Ye et al., 2022), with Briggs weighting set to -1.5 (Briggs, 1995). We adopt a minimum uv-value of 80λ and use pixel sizes of $0.1''$, $0.2''$, and $0.4''$, with corresponding Gaussian tapers of $0.3''$, $0.6''$, and $1.2''$. For efficient cleaning, we employ 'auto' masking, multi-scale deconvolution, and set an RMS box size 50 times the synthesised beam (Cornwell, 2008; Offringa & Smirnov, 2017). We apply a correction to account for the primary beam attenuation as well. We also use baseline-dependent averaging (BDA; e.g. Cotton, 1986, 2009; Wijnholds et al., 2018) during imaging.

6.5. Data

Observ. ID	Project	Calibrator ID	Observ. date	Pointing centres	Stations (int)	Frequencies
L686962	LT10_012	L686958	26-11-2018	16:11:00, +54.57.00	49 (11)	120-166 MHz
L833466	LT14_003	L833474	09-10-2021	16:11:00, +55.00.00	50 (13)	118-166 MHz

Table 6.1: Metadata from the 2 ELAIS-N1 observations used in this work. Observations used calibrator 3C 295 for calibration with `LINC`. The number of stations and frequencies are recorded after flagging. International stations are given between brackets.

ELAIS-N1 is a famous deep field, which has been explored across various wave-

lengths. This includes X-rays (e.g. Manners et al., 2003), ultraviolet (e.g. Martin et al., 2005), optical (e.g. McMahon et al., 2001; Aihara et al., 2018), infrared (e.g. Lawrence et al., 2007; Mauduit et al., 2012), and radio (e.g. Sirothia et al., 2009; Ocran et al., 2020; Best et al., 2023). This extensive multi-wavelength coverage has established ELAIS-N1 as an important field for extragalactic studies. Over 500 hrs of observing time are available for this field (Shimwell et al., 2025). About 200 hrs of these observations include only a few international stations or are heavily averaged, making them unsuitable for sub-arcsecond imaging. However, with hundreds of hrs of data still remaining, this field presents a great opportunity to produce the deepest LOFAR image to date. With the work done by Ye et al. (2024) and de Jong et al. (2024), we also have extensive prior knowledge of this field for calibrating and imaging data at arcsecond and sub-arcsecond resolutions. This gives us an advantage in experimenting with different calibrator sources and automated imaging settings to enhance the already existing images.

For the purpose of this work, we have selected 2 datasets from the ELAIS-N1 deep field to serve as test cases for our upgraded data reduction strategy. These datasets are detailed in Table 6.1. The first dataset, `L686962`, is selected from project `LT10_012`, representing their best observation and for us a benchmark for validating the improved calibration and imaging strategies. We used the already processed dataset up to the in-field calibration. The second dataset, `L833466`, comes from project `LT14_003` and features data recorded with a 2-sec integration time, compared to the default 1-sec, which leads to additional smearing at the edges (de Jong et al., 2024). Additionally, this observation was taken closer to solar maximum and shows stronger ionospheric effects compared to `L686962`, which is reflected in its `LINC` solution plots. This makes it an ideal candidate to test the robustness of our upgraded strategy in processing challenging observations. By combining these two datasets, we also evaluate the effectiveness of joint-calibration across multiple observations (see Section 6.4.4) and assess the impact of SVA (see Section 6.4.5).

6.6. Results

We have in this work, implemented enhancements in calibration quality and computational efficiency, enabling ultra-deep imaging at sub-arcsecond resolutions with LOFAR. To demonstrate these improvements, we have applied the improvements on two datasets. In this section, we highlight the resulting improvements in terms of calibration and computing costs.

6.6.1. Image quality improvements

The calibration strategy has been improved by addressing challenges highlighted in previous work (de Jong et al., 2024), and incorporating the solution interval metric and neural network described in Section 6.3. This has led to significant improvements in image quality, as outlined in this subsection.

6.6.1.1. Facet boundary leakage

In Section 6.4.1, we introduced an additional step to perform DD calibration for the Dutch core and remote stations before proceeding with DD calibration for the international stations. This improves the Dutch calibration solutions, which is in particular important for producing the $1.2''$ model images required for subtracting sources outside each facet before imaging (see Section 6.4.3). To illustrate the improvements in the $1.2''$ resolution image, Figure 6.11 shows a cutout of a challenging region. In the left panel of this figure, DD artefacts from two neighbouring calibrators were previously leaking into each other's facets. It is clear that in the new image in the right panel, this effect has been significantly reduced, which enables performing an improved subtraction of source signal at short (Dutch) baselines. This improvement also facilitates the final calibration refinement step (see Section 6.4.4) by minimizing interference from residual artefacts leaking from neighbouring facets.

6.6.1.2. Facet refinements

With better source subtraction before splitting out datasets for individual facets of our full mosaic, we mitigate the negative effect of high S/N sources affecting the self-calibration of the calibrators corresponding to each facet. This allows us, as outlined in Section 6.4.4, to perform a final self-calibration refinement step.

We tested the final calibration step for both the Dutch and international stations, with and without phasing up the Dutch core stations. For some facets, we found that the phase-up was not required, as most of the bright calibrator sources had been removed from the data, minimizing their impact on the calibration of shorter baselines. However, in a few cases, sources within the facet, other than the calibrator source, contained enough signal to still disrupt the calibration of short baselines. Therefore, in particular for automated approaches, it may be advisable to include a phase-up of the Dutch core stations to ensure robust calibration in the final refinement step.

Figure 6.12 illustrates the image quality enhancements that we achieve with the final refinement step, for calibrating one of the most challenging calibrator sources in the ELAIS-N1 field. This source introduced the most artefacts across the

full wide-field image from de Jong et al. (2024). At 1.2″ resolution we find most of the spike-like structure originating from the calibrator source to have reduced significantly from the left to right panel. At the 0.3″ resolution, all circular-type artefacts appear to be completely mitigated.

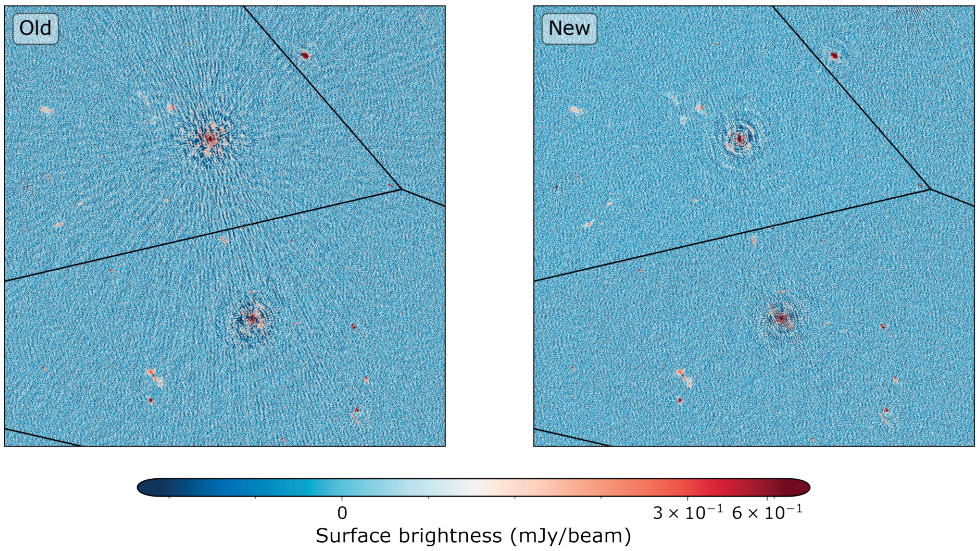


Figure 6.11: Improvements in direction-dependent (DD) calibration for 1.2″ imaging. The *left* image displays the image quality using the calibration strategy from de Jong et al. (2024), while the *right* image shows the results employing the new calibration strategy where we first apply Dutch-only DD calibration before calibrating the international stations for DDEs. This example corresponds to a challenging observation, which was previously difficult to calibrate. The black lines represent the facet boundaries.

6.6.2. Computing costs

Reducing LOFAR data for sub-arcsecond wide-field imaging is computationally intensive, due to its substantial storage demands, leading to significant CPU core hr requirements. In this subsection, we discuss the computational improvements we have made with the new data processing strategy.

6.6.2.1. Data volume

The large data volumes required for processing LOFAR data pose a significant bottleneck, especially for sub-arcsecond wide-field imaging, as the visibility data cannot be further averaged due to time and bandwidth smearing constraints. The initial unaveraged and uncompressed input data is approximately 16 TB. However, Dysco

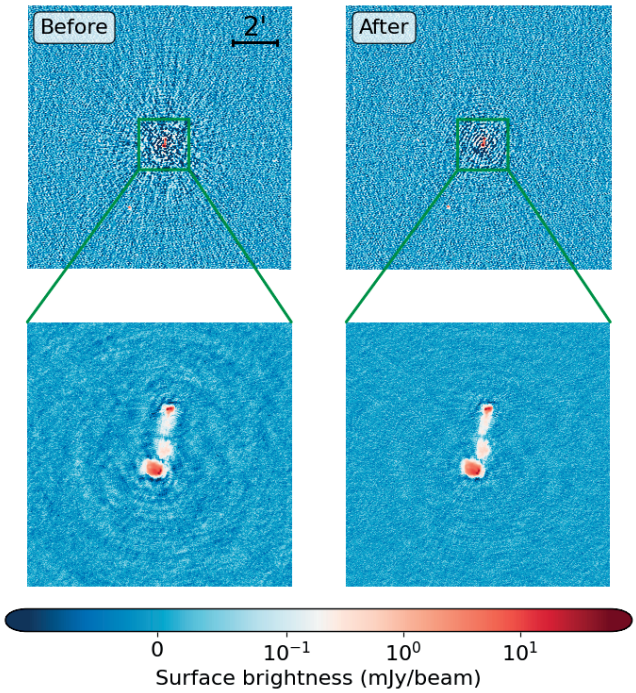


Figure 6.12: Final image quality improvements at 1.2″ resolution (above) and 0.3″ resolution (below) for one of the most challenging calibrators in the field. The *left* panels display the image quality before refinement (up to the calibration from Section 6.4.2), while the *right* panel presents the image after refinement (including calibration discussed in Section 6.4.4). We included a 2′ scale bar to illustrate the size of the image.

compression (Offringa, 2016) offers a solution to bring this number down with lossy compression. In the LOFAR Long-Term Archive (LTA), data are compressed by default at a factor of 4, with visibility data stored at 10 bits and their associated weights at 12 bits. These settings do generally not lead to a significant loss of information in LOFAR data. As noted by Offringa (2016), at high time and frequency resolutions, the data are expected to be noise-dominated. Therefore, compressing with higher bit rates may be possible for datasets for sub-arcsecond wide-field imaging.

To determine the extent to which we can further compress our data, we selected two facets from the ELAIS-N1 data after the 1.2″ subtraction, where one of the facets corresponds to a high S/N calibrator and the other to a low S/N calibrator. We applied various levels of Dysco compression by adjusting the bit rate used to store the visibilities and their corresponding weights and compared the results after imaging with the standard settings described in Section 6.4.5. Visibility weights

are already compressed by a factor of 10 compared to uncompressed data (Offringa, 2016), by storing only one polarisation and using a 12-bit compression, leaving limited potential for further compression. Changing the bit-rate reduces the total data volume only in the order of a few per cent. In contrast, much larger data volume reductions can be achieved beyond the default Dysco settings by adjusting the number of bits used for storing visibilities. This is because visibilities are stored as complex values and are more noise-dominated compared to weights. For varying compression levels expressed in the number of bits, we show in Figure 6.13, the RMS increase in the images as a function of compression level, given the data volume compressed size compared to the default 10-bit visibility and 12-bit weight compression with Dysco. This figure indicates that with 6-bit compression, the RMS background noise remains unchanged, and only a an increase in the order of a few per cent is observed for 4-bit compression, with more pronounced noise increases at higher compression levels, reaching about 100% at 2-bit compression. We also find that the image quality at $1.2''$ imaging is more affected at higher bit rates to RMS increases compared to the $0.3''$ resolution imaging. This is because the data is more averaged over time and frequency, making the data less noise-dominated. We verified that the residual images at these bit rates – obtained after subtracting the image corresponding to the original 10-bit stored data – remain purely noise-dominated and show no unusual artefacts. We also performed $1.2''$ wide-field imaging with datasets stored at different bit rates and found similar results to the $0.3''$ low S/N case, which is probably because these datasets are less averaged compared to the facet-subtracted data. Additionally, we confirmed that for bit rates below 6, the peak fluxes remain unaffected. Our results demonstrate that for the ELAIS-N1 dataset, and likely for many other LOFAR pointings, the visibility data can be compressed to 6-bits, resulting in a 40% data volume reduction. In cases where data volume presents a significant bottleneck, such as large-scale data processing with limited storage, 4-bit compression may also be viable.

The above applies to the data compression of one observation. For deep imaging using multiple observations, data compression is also achieved by reducing the number of visibilities through SVA, as discussed in Section 6.4.5 and by de Jong et al. (2025). For example, when combining 500 hrs of data without significant precession or aberration effects due to large observation time offsets, the number of visibilities can be reduced by a factor of 29. However, applying Dysco compression on top of SVA must be done with caution, as the combined data has lower noise levels depending on the number of observations merged.

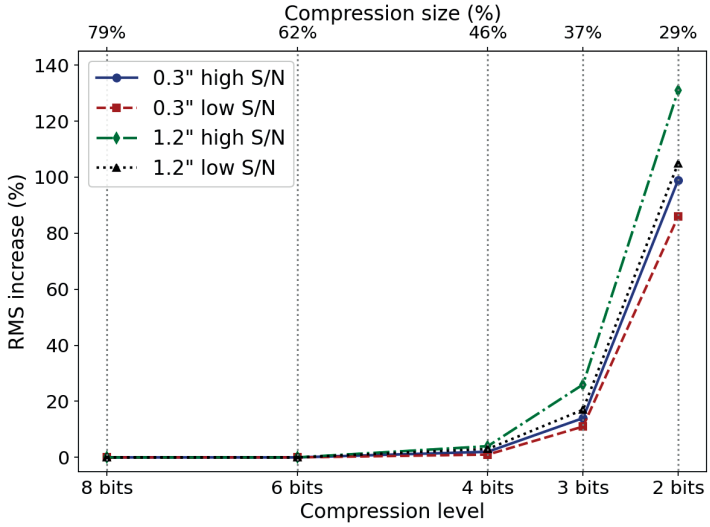


Figure 6.13: Comparison of RMS noise increase in image space as a function of the number of bits the visibilities are stored and compression size for two different facets, imaged at two different resolutions. The 1.2″ images are 4 times more in time and frequency averaged compared to the 0.3″ images. The compressed size is compared with Dysco's default settings (10-bit visibility storage and 12-bit weight storage).

6.6.2.2. CPU core hours

The computing costs reported in de Jong et al. (2024) summed to 680,000 CPU core hrs for calibrating and imaging 32 hrs of data, which corresponded to 170,000 CPU core hrs for 8 hrs of data. Since half of the data was before data reduction already averaged to 2 sec, this corresponds to about 250,000 CPU core hrs for one 8 hrs at 1 sec. With the strategy improvements presented in this work, we have achieved a significant reduction in these high costs, as outlined in this subsection. The costs for the different modules from our upgraded data processing strategy are given in Figure 6.1. The cost reductions enable us to scale data processing of LOFAR data for ultra-deep wide-field imaging and achieve sensitivities in the order of μJy beam^{-1}.

A new step, as part of the 'DD calibration' module in Figure 6.1, is the pre-self-calibration steps for the Dutch core and remote stations (see Section 6.4.1). This increases the total computational costs for every observation by about 8,000 CPU core hrs. However, for ultra-deep imaging, this cost can be partly mitigated by applying the Dutch calibration to for instance 6 observations to generate a sufficiently deep model, which can then be used as input for calibrating the remaining observations. This approach reduces the number of self-calibration cycles required for subsequent observations, lowering the overall cost by a factor of ~3 compared to

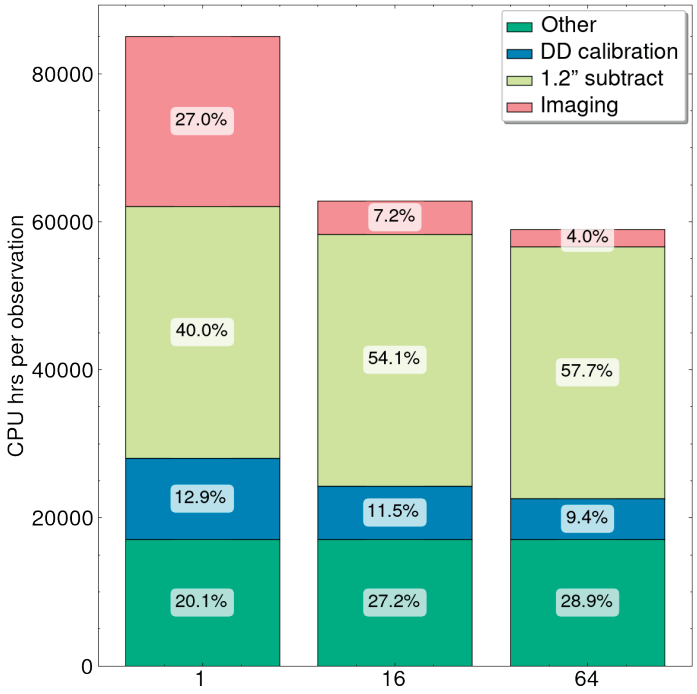


Figure 6.14: Breakdown of the computational cost across the major data processing steps when combining 1, 16, or 64 LOFAR observations of 8 hrs. We have separated the full cost for the DD calibration (outlined in Sections 6.4.1, 6.4.2, and 6.4.4) from the 1.2″ subtraction (outlined in 6.4.3), and imaging (outlined in Section 6.4.5). Percentages represent the fraction of the computational cost for full processing. For comparison, we normalised the total cost to process one 8 hrs observation. The lower computational cost when processing more observations is because using sidereal visibility averaging and model images during calibration reduces the required CPU core hrs. The computational cost for 'other' encompasses the sum of all additional steps to reduce the data in this work, which are associated with the modules not covered within 'This work' in Figure 6.1.

self-calibration without an initial model.

Employing early-stopping with the neural network from Section 6.3.3, allows us to self-calibrate the main facet calibrators to already reach convergence after 5 or 6 self-calibration cycles. Only in a few cases, this exceeds up to cycle 10 (in Section 6.4.2). This nearly halves the original computational cost for this calibration step compared to de Jong et al. (2024), which was about 1,500 CPU core hrs. However, the computational cost increases again with the introduction of an additional refinement step from Section 6.4.4, where we perform another round of self-calibration for all facets (see Section 6.4.4). Nonetheless, due to recent improvements in speed and data volume handling within `facetselfcal`, the total computational costs for

DD calibration of the international stations remain still comparable to that of de Jong et al. (2024). Thus, this step does not present a computational bottleneck.

The most computationally expensive module is the 1.2″ facet subtraction. While de Jong et al. (2024) included the facet subtraction with 1.2″ models as part of imaging, we have in this work separated this step from the imaging (see Section 6.4.3), as it is in the new strategy an essential part of performing the final high-quality self-calibration outlined in Section 6.4.4. We have significantly reduced the cost for this step by modifying the source subtraction strategy (see Section 6.4.3), resulting in a computational cost reduction of a factor of 5, bringing it down to approximately 32,000 CPU core hrs per observation. This is significant given that this step originally accounted for over 158,000 CPU core hrs for a single 8 hrs observation, which was 62% of the total data reduction. As a result, this improvement alone accelerates the entire overarching data processing workflow by about a factor of 2.

The final imaging module, utilizing SVA for deep imaging (see Section 6.4.5), introduces another significant speed improvement, especially when processing a large number of observations. This is because each observation has its own sidereal day. The CPU core hr reduction scales linearly with the data volume reductions mentioned in Section 6.6.2.1, with a slight compute overhead for applying the SVA algorithm itself. For 128 hrs of data, the reduction in CPU core hrs is approximately a factor of ~5 (equivalent to 5,700 CPU core hrs per observation), while for 500 hrs, the speedup increases to roughly ~10 (equivalent to 2,850 CPU core hrs per observation).

Beyond our calibration and imaging strategy changes, we are also benefiting from more efficient usage of computing resources, by combining CWL, Toil, and Slurm. This allows us to more efficiently distribute our jobs and optimise the number of CPU cores for each substep. In Figure 6.14, we compare the current total computational costs for reducing LOFAR data for wide-field 0.3″ imaging for combining 1, 16, and 64 observations. We normalised the cost to processing one 8 hrs observation. To simplify the plot and to focus on the steps optimised in this work, we have combined `LINC`, the `DDF`-pipeline, 6″ subtraction, and DI calibration in 'other'. We find a total computational cost for 1 observation of ~85,000 CPU core hrs, and estimate costs for 16 observations of ~1,000,000 CPU core hrs, and for 64 observations of ~4,000,000 CPU core hrs. This demonstrates that, depending on the number of combined observations, the total data processing cost has been reduced by a factor of 6 to 8 compared to Sweijen et al. (2022c, taking into account they averaged the data by a factor 2), and by a factor of 3 to 4 compared to de Jong et al. (2024, taking into account half of their data was averaged by a factor 2). This factor depends on how many observations are combined together with SVA.

6.7. Discussion

By enhancing our data reduction strategy, we have enabled the production of sub-arcsecond resolution images with sensitivities in the order of μJy beam^{-1} at 3 to 4 lower computational costs and with improved image qualities compared to what was previously possible, as we show in Section 6.6. In this section, we discuss possible remaining improvements and implications of our results for a potential future automated sub-arcsecond wide-field survey with LOFAR.

6.7.1. Modular data processing

The data processing framework presented in Section 6.2 and Figure 6.1 comprises different modules, each responsible for a specific data processing task. While all modules are essential for wide-field imaging, they function as individual blocks (as already highlighted by de Jong et al., 2024), which can be modified independently without directly affecting the input of the next module. This allows for flexible updates to the data processing strategy by improving an isolated single module or having the option of adding new modules.

Reordering of some of these modules may also be considered for performing DD calibration for the Dutch stations. This type of calibration happens currently both during the `DDF`-pipeline and after the DI in-field calibration of the international stations with `facetselfcal` (see Section 6.4.1). However, it may be more efficient to allow the option to pre-apply the DD solutions from the `DDF`-pipeline before performing the DD calibration of the international stations. This implicates that we could place the `DDF`-pipeline module in Figure 6.1 after the DI in-field calibration of the international stations, making the additional DD calibration step for the Dutch stations with `facetselfcal` (from Section 6.4.1) unnecessary. This may lead to computing cost reductions by approximately 10% over the entire data processing.

Given the calibration strategy of our in-field calibrator (as discussed in Section 3.2.3. from de Jong et al., 2024) with full-Jones corrections, we can also perform polarisation studies. However, we need for this to align the different observations by applying an additional correction to the polarisation angles across observations. An automatic approach for this correction may be implemented as a separate module prior to the final imaging step. Similarly, additional calibration refinement steps or automated inspections can be added between already existing modules as well.

A benefit of the new data processing strategy for deep imaging is that we fully process each observation embarrassingly parallel up to the 'DD calibration refinement' module in Figure 6.1. This is possible because only at the DD calibration refinement stage do the different observations need to be calibrated together, as discussed in Section 6.4.4. Parallel processing is in particular beneficial when pro-

cessing a large number of observations of the same field of view for ultra-deep imaging on different nodes or even different HPCs. It also enables full processing up to the calibrated facet datasets and removing all intermediate products before processing the next dataset. This may be essential when storage space is limited.

6.7.2. Advanced decision making

In Section 6.3.3, we introduced the use of neural networks to employ early-stopping during self-calibration and optimise our automatic DD calibration parameter settings in Section 6.4.2. This approach shows that incorporating steps for automatic decision-making in our data reduction strategy enhances performance by removing human intervention to select the optimal self-calibration cycle, while also reducing computational costs. In this subsection, we discuss a few other use-cases to incorporate automated decision-making during data reduction.

6.7.2.1. Calibratability

Not every dataset is calibratable, by which we mean that some datasets are dominated by RFI or bad ionospheric conditions and therefore too challenging to calibrate. It may therefore in some cases not be worth processing a dataset, given the high computing costs involved. To avoid wasting the use of computing resources on calibrating bad data, it is worth to assess the calibratability of data at an early stage. This could for instance be done before or after running `LINC`, which is the first step in processing LOFAR data (see Figure 6.1). This may be best done automatically by including various metrics on the calibration solutions or UV data directly, or perhaps by using external data with information on the ionospheric conditions during observations (e.g. Flisek et al., 2023). A final data quality score would not only aid in assessing our specific use case but also benefit other applications of LOFAR data across various scientific objectives, as `LINC` serves as a general data reduction step for imaging LOFAR datasets.

Similarly, the `DDF`-pipeline returns data products, including wide-field images, that indicate the calibratability of the data as well. An assessment of the remaining DDEs may indicate the severity of the ionosphere for a particular dataset. Providing postage stamps of bright sources in the field to our neural network from Section 6.3.3, can assist in deciding whether to continue or stop data processing. This would only require a re-training of the model including images at $6''$. Since the majority of CPU core hrs are spent after `LINC` and the `DDF`-pipeline (see Figure 6.1), this approach can save a considerable amount of computing time, allowing resources to be allocated to processing other observations with fewer data issues.

6.7.2.2. In-field calibration

Ye et al. (2024) and de Jong et al. (2024) devoted much attention to manually select and calibrate the in-field calibrator for ELAIS-N1 (identified as ICRF J160607.6+552135; Charlot et al., 2020; Sexton et al., 2022). Since for ELAIS-N1, we are aiming to calibrate the same field of view for many observations, we do not have to modify this calibration step. However, for a potential future wide-field imaging survey (e.g. ILoTSS) this process needs to be automated for different sky areas.

The first source of information for selecting potential in-field calibrators comes from the Long-Baseline Calibrator Survey (LBCS; Jackson et al., 2016, 2022). However, if no candidate sources from the LBCS catalogue are available, one could consider using the brightest sources above a specific mJy threshold from the LoTSS catalogues (Shimwell et al., 2017, 2019, 2022), which covers most of the northern sky. In the rare instance where a LoTSS catalogue is unavailable, one could also consider using source finder software such as `PyBDSF`[15] (Mohan & Rafferty, 2015), on the 6″ map that the `DDF`-pipeline produced in the step before the international DI calibration (see Figure 6.1).

Having at least one calibrator source after the initial catalogue-based selection, allows us to assess the S/N for the longest baselines by employing the same calibrator selection method based on the phasediff-scores as described in Section 6.3.1, but with a different score threshold. To find the best calibrator and (self-)calibration parameters, a good starting point is using similar settings as those optimised by de Jong et al. (2024). Next to this, it may be worth applying a grid search with different (starting) parameters on the candidate(s). In this way, the neural network model from Section 6.3.3 and additional image and solution quality measures can be utilised to select the best parameters and/or candidate source. Given that the CPU cost is low for the DI calibration (see Figure 6.1), this step does not substantially increase the overall computing time. Over time, when many in-field calibrators have been optimised and the parameters from the different grid-searches have been collected, a more advanced new neural network can be trained to immediately set the best initial parameters corresponding to the potential in-field calibrator source that has been automatically selected.

6.7.2.3. Self-calibration parameter-tuning

In the DD self-calibration images from Figure 6.6, we observe that while the self-calibration results in the last two rows show minor improvements after the initial cycles, additional cycles do not lead to full convergence. In the fourth row, despite the neural network providing a positive assessment, there remains room for further

[15] `https://pybdsf.readthedocs.io`

refinement. In such cases, it may be beneficial to not only assess convergence and image quality but also adjust the calibration parameters to enhance results. A possible approach could involve beginning with a conservative set of parameters in the initial self-calibration round, then gradually shifting to less restrictive settings, such as shorter solution intervals and reduced frequency smoothness constraints, continuing until convergence is achieved according to our metrics. If self-calibration starts to diverge, parameters can be adjusted back to more conservative values.

6.7.3. Data compression and I/O

With our data reduction strategy enhancements and data volume reductions, we have significantly reduced the computing costs compared to previous works. Nonetheless, more gains in terms of computing cost reductions are still possible.

As outlined in Section 6.6.2.1, we use Dysco compression to reduce the size of our visibility data. For the ELAIS-N1 dataset, we found that stricter compression using fewer bits to store visibilities data, reduced data volumes without loss of image quality. This works well for ELAIS-N1 due to the absence of extremely bright sources in the field. However, for other observations containing brighter sources or datasets with more averaging applied, the optimal bit rates may differ, as the data is less noise-dominated. Additionally, we do not recommend using high compression levels for phased-up data, as the S/N is much higher on baselines involving the Dutch superstation. Since bit rates depend on the S/N across baselines in each dataset, optimizing compression for automated data processing would require a method to dynamically determine the maximum allowable compression level for each specific dataset. With higher bit rates, we can improve the I/O overhead in our data reduction, such as during the $1.2''$ subtraction, where the unaveraged subtracted data has to be copied multiple times to perform the prediction for each facet before averaging down in time and frequency.

A limitation of Dysco is that it cannot compress model data or be applied to BDA data. Recent work with MultiGrid Adaptive Reduction of Data (MGARD; Dodson et al., 2024) compression on high-resolution ELAIS-N1 data from de Jong et al. (2024) discuss the potential for combining lossy compression with BDA, while performing similar levels of compression as Dysco on the visibility data and weights. MGARD may also have the potential to further compress model data, which could help reduce data volume requirements during self-calibration, particularly in the multi-directional self-calibration process discussed in Section 6.4.1.

6.8. Summary and conclusion

In this work, we tackled several key challenges that limit the efficient processing of numerous observations for ultra-deep sub-arcsecond wide-field imaging with LOFAR. Our main improvements include:

- An efficient modular framework for data reduction on HPC systems with a scheduler supporting Toil and CWL, allowing for embarrassingly parallel processing of many observations of the same field for ultra-deep imaging and having the potential to perform ILoTSS surveys as well.

- An improved automated DD calibration strategy, with a focus on improving the calibration solutions for the Dutch stations and removing the need for human interventions, by combining already existing image and solution quality metrics with a neural network that assesses image quality.

- optimised strategy for reducing the computing costs of wide-field imaging by a factor of 3 to 4 compared to the most recent similar work. The exact factor depends on the number of observations that are being combined since more observations of the same sky area allow with our new strategy for more speedup.

Our new data reduction strategy has been validated using two observations from the ELAIS-N1 deep field, where one was used to compare results with de Jong et al. (2024), and the other for testing on a more challenging observation. We find clear improvements for in particular the calibration of the shorter baselines, compared to previous work, which enhances the quality for both the $1.2''$ and $0.3''$ resolutions.

We have also identified and discussed several steps to further improve the automated data reduction strategy:

- In the new data calibration strategy we perform DD calibration on the Dutch stations twice. This is inefficient from a workflow perspective. A possible solution is to perform DI in-field calibration before running the `DDF`-pipeline or `facetselfcal` for joint-calibration of the Dutch stations. With the modular design of our framework, this can be done by reordering the main processing steps, which may eventually save an additional 10% on total computing time.

- We have demonstrated the value of automated decision-making through a neural network for image quality validation. Additional steps earlier in the data reduction process could also be improved with more data quality assessments. This would help reduce the waste of computing resources on observations impacted by poor ionospheric conditions.

- To extend this work to other sky areas through automatic processing, an automated DI in-field calibration step is currently missing. We propose using similar techniques as we have presented in this work for DD calibration. However, this needs careful testing on more observations from sky areas with various conditions.

- Further improvements in data processing speed can be achieved by focusing on reducing data volumes, by incorporating data compression techniques to compress model data or by combining lossy data compression with BDA.

The next major goal is to apply our calibration and imaging strategy to reduce hundreds of hrs of ELAIS-N1 data available in the LTA and other deep surveys, aiming to reach sensitivities on the order of 5μJy beam^{-1}. This work also marks significant progress toward establishing a fully automated survey workflow for (deep) sub-arcsecond imaging with LOFAR.

Acknowledgements

This work is part of the project CORTEX (NWA.1160.18.316) of the research programme NWA-ORC which is (partly) financed by the Dutch Research Council (NWO). This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-1287. This work is co-funded by the EGI-ACE project (Horizon 2020) under Grant number 101017567. RJvW acknowledges support from the ERC Starting Grant ClusterWeb 804208. LOFAR data products were provided by the LOFAR Surveys Key Science project (LSKSP; `https://lofar-surveys.org/`*) and were derived from observations with the International LOFAR Telescope (ILT). LOFAR (van Haarlem et al., 2013) is the Low Frequency Array designed and constructed by ASTRON. It has observing, data processing, and data storage facilities in several countries, which are owned by various parties (each with their own funding sources), and which are collectively operated by the ILT foundation under a joint scientific policy. The efforts of the LSKSP have benefited from funding from the European Research Council, NOVA, NWO, CNRS-INSU, the SURF Co-operative, the UK Science and Technology Funding Council and the Jülich Supercomputing Centre. The use of the national computer facilities in this research was subsidised by NWO Domain Science.*

Bibliography

Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, Publications of the Astronomical Society of Japan, 70, S8

Akamatsu, H., Fujita, Y., Akahori, T., et al. 2017, in The X-ray Universe 2017, ed. J.-U. Ness & S. Migliari, 30

Akritas, M. G. & Bershady, M. A. 1996, The Astrophysical Journal, 470, 706

Alexander, P. & Leahy, J. P. 1987, MNRAS, 225, 1

Amstutz, P., Crusoe, M. R., Tijanić, N., et al., eds. 2016, Common Workflow Language, v1.0. Specification (Common Workflow Language working group)

Arakawa, D. 1936, Report on Radio Research in Japan, 3, 31

Arras, P., Reinecke, M., Westermann, R., & Enßlin, T. A. 2021, A&A, 646, A58

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, The Astrophysical Journal, 935, 167

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, The Astronomical Journal, 156, 123

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Atemkeng, M., Perkins, S., Kenyon, J. S., Hugo, B. V., & Smirnov, O. 2022, in Astronomical Society of the Pacific Conference Series, Vol. 532, Astronomical Society of the Pacific Conference Series, ed. J. E. Ruiz, F. Pierfedereci, & P. Teuben, 71

Baade, W. & Minkowski, R. 1954, The Astrophysical Journal, 119, 206

Baars, J. W. M., van der Brugge, J. F., Casse, J. L., et al. 1973, IEEE Proceedings, 61, 1258

Bacchi, M., Feretti, L., Giovannini, G., & Govoni, F. 2003, A&A, 400, 465

Baldi, R. D., Capetti, A., & Giovannini, G. 2015, A&A, 576, A38

Baldi, R. D., Capetti, A., & Massaro, F. 2018, A&A, 609, A1

Banfield, J. K., Wong, O. I., Willett, K. W., et al. 2015, MNRAS, 453, 2326

Barišić, I., van der Wel, A., van Houdt, J., et al. 2019, The Astrophysical Journal, 872, L12

Barthel, P. D. & Arnaud, K. A. 1996, MNRAS, 283, L45

Becker, R. H., White, R. L., & Helfand, D. J. 1995, The Astrophysical Journal, 450, 559

Berger, E., Ball, S., Becker, K. M., et al. 2001, Nature, 410, 338

Best, P. N. 2009, Astronomische Nachrichten, 330, 184

Best, P. N., Kaiser, C. R., Heckman, T. M., & Kauffmann, G. 2006, MNRAS, 368, L67

Best, P. N., Kondapally, R., Williams, W. L., et al. 2023, MNRAS, 523, 1729

Biava, N., de Gasperin, F., Bonafede, A., et al. 2021, MNRAS, 508, 3995

Bicknell, G. V. 1994, The Astrophysical Journal, 422, 542

Bicknell, G. V. 1995, The Astrophysical Journal, 101, 29

Blandford, R. & Eichler, D. 1987, Physics Reports, 154, 1

Blandford, R., Meier, D., & Readhead, A. 2019, Annual Review of Astronomy and Astrophysics, 57, 467

Blasi, P. & Colafrancesco, S. 1999, Astroparticle Physics, 12, 169

Böhringer, H. & Werner, N. 2010, Astronomy and Astrophysics Review, 18, 127

Bolton, J. G. & Stanley, G. J. 1948a, Australian Journal of Scientific Research A Physical Sciences, 1, 58

Bolton, J. G. & Stanley, G. J. 1948b, Nature, 161, 312

Bonafede, A., Brunetti, G., Rudnick, L., et al. 2022, The Astrophysical Journal, 933, 218

Bonafede, A., Brunetti, G., Vazza, F., et al. 2021, The Astrophysical Journal, 907, 32

Bondi, M., Scaramella, R., Zamorani, G., et al. 2024, A&A, 683, A179

Bonjean, V., Aghanim, N., Salomé, P., Douspis, M., & Beelen, A. 2018, A&A, 609, A49

Bonnassieux, E., Sweijen, F., Brienza, M., et al. 2022, A&A, 658, A10

Botteon, A., Brunetti, G., van Weeren, R. J., et al. 2020a, The Astrophysical Journal, 897, 93

Botteon, A., Cassano, R., Eckert, D., et al. 2019, A&A, 630, A77

Botteon, A., Gastaldello, F., & Brunetti, G. 2018a, MNRAS, 476, 5591

Botteon, A., Giacintucci, S., Gastaldello, F., et al. 2021, A&A, 649, A37

Botteon, A., Shimwell, T. W., Bonafede, A., et al. 2018b, MNRAS, 478, 885

Botteon, A., Shimwell, T. W., Cassano, R., et al. 2022, A&A, 660, A78

Botteon, A., van Weeren, R. J., Brunetti, G., et al. 2020b, MNRAS, 499, L11

Bourke, S., Mooley, K., & Hallinan, G. 2014, in Astronomical Society of the Pacific Conference Series, Vol. 485, Astronomical Data Analysis Software and Systems XXIII, ed. N. Manset & P. Forshay, 367

Boxelaar, J. M., van Weeren, R. J., & Botteon, A. 2021, Astronomy and Computing, 35, 100464

Braude, S. I., Men, A. V., & Sodin, L. G. 1978, Antenny, 26, 3

Bridle, A. H. & Perley, R. A. 1984, Annual Review of Astronomy and Astrophysics, 22, 319

Bridle, A. H. & Schwab, F. R. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 371

Briggs, D. S. 1995, in American Astronomical Society Meeting Abstracts, Vol. 187, American Astronomical Society Meeting Abstracts, 112.02

Brondolo, F. & Beaussant, S. 2024, arXiv e-prints, arXiv:2407.18100

Brouw, W. N. 1975, Methods in Computational Physics, 14, 131

Brunetti, G. 2004, Journal of Korean Astronomical Society, 37, 493

Brunetti, G., Cassano, R., Dolag, K., & Setti, G. 2009, A&A, 507, 661

Brunetti, G., Giacintucci, S., Cassano, R., et al. 2008, Nature, 455, 944

Brunetti, G. & Jones, T. W. 2014, International Journal of Modern Physics D, 23, 1430007

Brunetti, G. & Lazarian, A. 2011, MNRAS, 412, 817

Brunetti, G., Setti, G., Feretti, L., & Giovannini, G. 2001, New Astronomy, 6, 1

Brunetti, G. & Vazza, F. 2020, Physical Review Letters, 124, 051101

Buote, D. A. 2001, The Astrophysical Journal, 553, L15

Burns, J. O., Sulkanen, M. E., Gisler, G. R., & Perley, R. A. 1992, The Astrophysical Journal Letters, 388, L49

Calistro Rivera, G., Williams, W. L., Hardcastle, M. J., et al. 2017, MNRAS, 469, 3468

Callingham, J. R., Pope, B. J. S., Feinstein, A. D., et al. 2021, A&A, 648, A13

Callingham, J. R., Shimwell, T. W., Vedantham, H. K., et al. 2023, A&A, 670, A124

Capetti, A., Brienza, M., Baldi, R. D., et al. 2020, A&A, 642, A107

Capetti, A., Massaro, F., & Baldi, R. D. 2017a, A&A, 598, A49

Capetti, A., Massaro, F., & Baldi, R. D. 2017b, A&A, 601, A81

Carilli, C. L. & Rawlings, S. 2004, New Astronomy Reviews, 48, 979

CASA Team, Bean, B., Bhatnagar, S., et al. 2022, Publications of the Astronomical Society of the Pacific, 134, 114501

Casacore Team. 2019, casacore: Suite of C++ libraries for radio astronomy data processing, Astrophysics Source Code Library, record ascl:1912.002

Cassano, R., Botteon, A., Di Gennaro, G., et al. 2019, The Astrophysical Journal, 881, L18

Cassano, R., Brunetti, G., & Setti, G. 2006, MNRAS, 369, 1577

Cassano, R., Ettori, S., Brunetti, G., et al. 2013, The Astrophysical Journal, 777, 141

Cassano, R., Ettori, S., Giacintucci, S., et al. 2010, The Astrophysical Journal, 721, L82

Cattaneo, A., Faber, S. M., Binney, J., et al. 2009, Nature, 460, 213

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560

Charlot, P., Jacobs, C. S., Gordon, D., et al. 2020, A&A, 644, A159

Ciliegi, P., McMahon, R. G., Miley, G., et al. 1999, MNRAS, 302, 222

Clark, B. G. 1980, A&A, 89, 377

Cochrane, R. K., Kondapally, R., Best, P. N., et al. 2023, arXiv e-prints, arXiv:2305.15510

Condon, J. J., Cotton, W. D., & Broderick, J. J. 2002, The Astronomical Journal, 124, 675

Condon, J. J., Cotton, W. D., Fomalont, E. B., et al. 2012, The Astrophysical Journal, 758, 23

Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, The Astronomical Journal, 115, 1693

Cooley, J. W. & Tukey, J. W. 1965, Mathematics of Computation, 19, 297

Cornwell, T. & Fomalont, E. B. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 187

Cornwell, T. J. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 793

Cotton, W. 2009, Effects of baseline dependent time averaging of uv data, Tech. rep., National Radio Astronomy Observatory

Cotton, W. D. 1986, in Synthesis Imaging, ed. R. A. Perley, F. R. Schwab, & A. H. Bridle, 123–136

Cotton, W. D. 2008, Publications of the Astronomical Society of the Pacific, 120, 439

Croft, S., Bower, G. C., & Whysong, D. 2013, The Astrophysical Journal, 762, 93

Croston, J. H., Hardcastle, M. J., Mingo, B., et al. 2019, A&A, 622, A10

Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11

Crusoe, M. R., Abeln, S., Iosup, A., et al. 2022, Communications of the ACM, 65, 54

Dabhade, P., Röttgering, H. J. A., Bagchi, J., et al. 2020, A&A, 635, A5

Darcet, T., Oquab, M., Mairal, J., & Bojanowski, P. 2023, arXiv e-prints, arXiv:2309.16588

De Breuck, C., van Breugel, W., Röttgering, H. J. A., & Miley, G. 2000, A&AS, 143, 303

de Gasperin, F. 2017, MNRAS, 467, 2234

de Gasperin, F., Brunetti, G., Brüggen, M., et al. 2020, A&A, 642, A85

de Gasperin, F., Dijkema, T. J., Drabent, A., et al. 2019a, A&A, 622, A5

de Gasperin, F., Dijkema, T. J., Drabent, A., et al. 2019b, A&A, 622, A5

de Gasperin, F., Mevius, M., Rafferty, D. A., Intema, H. T., & Fallows, R. A. 2018, A&A, 615, A179

de Jong, J. M. G. H. J., Röttgering, H. J. A., Kondapally, R., et al. 2024, A&A, 683, A23

de Jong, J. M. G. H. J., van Weeren, R. J., Botteon, A., et al. 2022, A&A, 668, A107

de Jong, J. M. G. H. J., van Weeren, R. J., Dijkema, T. J., et al. 2025, A&A, 694, A98

de Ruiter, H. R., Willis, A. G., & Arp, H. C. 1977, A&AS, 28, 211

De Young, D. S. 1992, The Astrophysical Journal, 386, 464

DeForest, C. E. 2004, Solar Physics, 219, 3

Dehant, V., Laguerre, R., Rekier, J., et al. 2017, Geodesy and Geodynamics, 8, 389, geodesy, Astronomy and Geophysics in Earth Rotation

Deng, Q.-W., Wang, F., Deng, H., et al. 2022, Research in Astronomy and Astrophysics, 22, 045014

Dennison, B. 1980, The Astrophysical Journal Letters, 239, L93

Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, IEEE Proceedings, 97, 1482

Dey, A., Schlegel, D. J., Lang, D., et al. 2019, The Astronomical Journal, 157, 168

Dijkema, T. J., Nijhuis, M., van Diepen, G., et al. 2023, DP3: Streaming processing pipeline for radio interferometric data, Astrophysics Source Code Library, record ascl:2305.014

Dodson, R., Williamson, A., Gong, Q., et al. 2024, arXiv e-prints, arXiv:2410.15683

Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2020, arXiv e-prints, arXiv:2010.11929

Dubois, Y., Gavazzi, R., Peirani, S., & Silk, J. 2013, MNRAS, 433, 3297

Duchesne, S. W., Johnston-Hollitt, M., & Wilber, A. G. 2021, Publications of the Astronomical Society of Australia, 38, e031

Duncan, K. J., Kondapally, R., Brown, M. J. I., et al. 2021, A&A, 648, A4

Duncan, K. J., Sabater, J., Röttgering, H. J. A., et al. 2019, A&A, 622, A3

Dunlop, J. S. & Peacock, J. A. 1990, MNRAS, 247, 19

Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E., & Archer, S. 1959, Memoirs of the Royal Astronomical Society, 68, 37

Edler, H. W., de Gasperin, F., & Rafferty, D. 2021, A&A, 652, A37

Elbers, A. 2015, PhD thesis, -

Ellingson, S. W., Clarke, T. E., Cohen, A., et al. 2009, IEEE Proceedings, 97, 1421

Ewen, H. I. & Purcell, E. M. 1951, Nature, 168, 356

Fabian, A. C. 2012, Annual Review of Astronomy and Astrophysics, 50, 455

Fabian, A. C., Peres, C. B., & White, D. A. 1997, MNRAS, 285, L35

Fanaroff, B. L. & Riley, J. M. 1974, MNRAS, 167, 31P

Feretti, L., Fusco-Femiano, R., Giovannini, G., & Govoni, F. 2001, A&A, 373, 106

Fisher, N. I., Lewis, T., & Embleton, B. J. J. 1993, Statistical Analysis of Spherical Data

Flisek, P., Forte, B., Fallows, R., et al. 2023, Journal of Space Weather and Space Climate, 13, 27

Folk, M., Heber, G., Koziol, Q., Pourmal, E., & Robinson, D. 2011, ACM International Conference Proceeding Series, 36

Fujita, Y., Koyama, K., Tsuru, T., & Matsumoto, H. 1996, Publications of the Astronomical Society of Japan, 48, 191

Fujita, Y., Tawa, N., Hayashida, K., et al. 2008, Publications of the Astronomical Society of Japan, 60, S343

Garn, T., Green, D. A., Riley, J. M., & Alexander, P. 2008, MNRAS, 383, 75

Gehrels, N. 1986, The Astrophysical Journal, 303, 336

Gendre, M. A., Best, P. N., & Wall, J. V. 2010, MNRAS, 404, 1719

Gendre, M. A., Best, P. N., Wall, J. V., & Ker, L. M. 2013, MNRAS, 430, 3086

Gendre, M. A. & Wall, J. V. 2008, MNRAS, 390, 819

Gerace, F., Saglietti, L., Sarao Mannelli, S., Saxe, A., & Zdeborová, L. 2022, Machine Learning: Science and Technology, 3, 015030

Ghisellini, G., Celotti, A., Tavecchio, F., Haardt, F., & Sbarrato, T. 2014, MNRAS, 438, 2694

Giacintucci, S., Venturi, T., Brunetti, G., et al. 2005, A&A, 440, 867

Gitti, M., Brunetti, G., & Setti, G. 2002, A&A, 386, 456

Glendenning, B. E. 1996, in Astronomical Society of the Pacific Conference Series, Vol. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes, 271

Gopal-Krishna & Wiita, P. J. 2000, A&A, 363, 507

Govoni, F., Enßlin, T. A., Feretti, L., & Giovannini, G. 2001, A&A, 369, 441

Govoni, F., Orrù, E., Bonafede, A., et al. 2019, Science, 364, 981

Greisen, E. W. 2003, in Astrophysics and Space Science Library, Vol. 285, Information Handling in Astronomy - Historical Vistas, ed. A. Heck, 109

Groeneveld, C., van Weeren, R. J., Miley, G. K., et al. 2022, A&A, 658, A9

Gu, L., Akamatsu, H., Shimwell, T. W., et al. 2019, Nature Astronomy, 3, 838

Gubanov, V. S. 1973, Soviet Astronomy, 16, 907

Ha, J.-H., Ryu, D., Kang, H., & van Marle, A. J. 2018, The Astrophysical Journal, 864, 105

Hales, C. A., Norris, R. P., Gaensler, B. M., et al. 2014, MNRAS, 441, 2555

Ham, R. A. 1975, Journal of the British Astronomical Association, 85, 317

Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&A, 117, 137

Hardcastle, M. J. & Croston, J. H. 2020, New Astronomy Reviews, 88, 101539

Hardcastle, M. J., Croston, J. H., Shimwell, T. W., et al. 2019a, MNRAS, 488, 3416

Hardcastle, M. J., Gürkan, G., van Weeren, R. J., et al. 2016, MNRAS, 462, 1910

Hardcastle, M. J., Williams, W. L., Best, P. N., et al. 2019b, A&A, 622, A12

Harris, D. E., Lari, C., Vallee, J. P., & Wilson, A. S. 1980, A&A, 42, 319

Harris, D. E., Moldón, J., Oonk, J. R. R., et al. 2019, The Astrophysical Journal, 873, 21

Harwood, J. J., Hardcastle, M. J., Croston, J. H., & Goodger, J. L. 2013, MNRAS, 435, 3353

Harwood, J. J., Mooney, S., Morabito, L. K., et al. 2022, A&A, 658, A8

Harwood, J. J., Vernstrom, T., & Stroe, A. 2020, MNRAS, 491, 803

Heightman, D. W. 1936, T&R Bulletin, 1937, 496

Herd, J. S. & Conway, M. D. 2016, Proceedings of the IEEE, 104, 519

Herrera Ruiz, N., O'Sullivan, S. P., Vacca, V., et al. 2021, A&A, 648, A12

Herschel, W. 1800, Philosophical Transactions of the Royal Society of London Series I, 90, 255

Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, Nature, 217, 709

Hey, J. S., Parsons, S. J., & Phillips, J. W. 1946, Nature, 158, 234

Hill, J. M. & Oegerle, W. R. 1993, The Astronomical Journal, 106, 831

Hincks, A. D., Radiconi, F., Romero, C., et al. 2022, MNRAS, 510, 3335

Hoang, D. N., Shimwell, T. W., Osinga, E., et al. 2021, MNRAS, 501, 576

Hoeft, M., Brüggen, M., Yepes, G., Gottlöber, S., & Schwope, A. 2008, MNRAS, 391, 1511

Hoeft, M., Dumba, C., Drabent, A., et al. 2021, A&A, 654, A68

Högbom, J. A. 1974, Astronomy and Astrophysics Supplement Series, 15, 417

Hoyle, F. & Fowler, W. A. 1963, MNRAS, 125, 169

Hu, M.-K. 1962, IRE Transactions on Information Theory, 8, 179

Huang, W., Yi, M., Zhao, X., & Jiang, Z. 2021, arXiv e-prints, arXiv:2111.00743

Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, The Astrophysical Journal, 199, 26

Huynh, M. T., Jackson, C. A., Norris, R. P., & Prandoni, I. 2005, The Astronomical Journal, 130, 1373

Ignesti, A. 2022, New Astronomy, 92, 101732

Ignesti, A., Brunetti, G., Gitti, M., & Giacintucci, S. 2020, A&A, 640, A37

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

Jackson, C. A. & Wall, J. V. 1999, MNRAS, 304, 160

Jackson, N., Badole, S., Morgan, J., et al. 2022, A&A, 658, A2

Jackson, N., Tagore, A., Deller, A., et al. 2016, A&A, 595, A86

Jaffe, W. J. 1977, The Astrophysical Journal, 212, 1

Jaffe, W. J. & Rudnick, L. 1979, The Astrophysical Journal, 233, 453

Jamrozy, M. 2004, A&A, 419, 63

Jamrozy, M., Konar, C., Machalski, J., & Saikia, D. J. 2008, MNRAS, 385, 1286

Jannuzi, B. T. & Dey, A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 193, The Hy-Redshift Universe: Galaxy Formation and Evolution at High Redshift, ed. A. J. Bunker & W. J. M. van Breugel, 258

Jansky, K. G. 1933, Nature, 132, 66

Joblib Development Team. 2020, Joblib: running Python functions as pipeline jobs

Jonas, J. & MeerKAT Team. 2016, in MeerKAT Science: On the Pathway to the SKA, 1

Joye, W. A. & Mandel, E. 2003, in Astronomical data analysis software and systems XII, Vol. 295, 489

Jurlin, N., Morganti, R., Sweijen, F., et al. 2024, A&A, 682, A118

Kaiser, C. R. & Best, P. N. 2007, MNRAS, 381, 1548

Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7733, Ground-based and Airborne Telescopes III, ed. L. M. Stepp, R. Gilmozzi, & H. J. Hall, 77330E

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055

Kazemi, S., Yatawatta, S., Zaroubi, S., et al. 2011, MNRAS, 414, 1656

Kellermann, K. I., Bouton, E. N., & Brandt, S. S. 2020, The Postwar Explosion in Radio Astronomy: The US Falls Behind (Cham: Springer International Publishing), 35–75

Kellermann, K. I. & Moran, J. M. 2001, Annual Review of Astronomy and Astrophysics, 39, 457

Kettenis, M., van Langevelde, H. J., Reynolds, C., & Cotton, B. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 497

Kim, K. T., Kronberg, P. P., Giovannini, G., & Venturi, T. 1989, Nature, 341, 720

Klamer, I. J., Ekers, R. D., Bryant, J. J., et al. 2006, MNRAS, 371, 852

Kondapally, R., Best, P. N., Cochrane, R. K., et al. 2022, MNRAS, 513, 3742

Kondapally, R., Best, P. N., Hardcastle, M. J., et al. 2021, A&A, 648, A3

Kop, M., Aboy, M., De Jong, E., et al. 2023, arXiv e-prints, arXiv:2303.16671

Kormendy, J. & Ho, L. C. 2013, Annual Review of Astronomy and Astrophysics, 51, 511

Kovalevsky, J. 2003, A&A, 404, 743

Kraus, J. D. 1966, Radio Astronomy (New York: McGraw-Hill)

Kravtsov, A. V. & Borgani, S. 2012, Annual Review of Astronomy and Astrophysics, 50, 353

Krolik, J. H. & Chen, W. 1991, The Astronomical Journal, 102, 1659

Kruithof, G., Bassa, C., Bonati, I., et al. 2023, Experimental Astronomy, 56, 687

Kukreti, P., Morganti, R., Shimwell, T. W., et al. 2022, A&A, 658, A6

Kundu, B., Khanal, B., Simon, R., & Linte, C. A. 2024, arXiv e-prints, arXiv:2411.09598

Kurtzer, G. M., Sochat, V., & Bauer, M. W. 2017, PLOS ONE, 12, 1

Lane, W. M., Cotton, W. D., van Velzen, S., et al. 2014, MNRAS, 440, 327

Lara, L., Márquez, I., Cotton, W. D., et al. 2001, A&A, 378, 826

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

Ledlow, M. J. & Owen, F. N. 1996, The Astronomical Journal, 112, 9

Lockman, F. J., Jahoda, K., & McCammon, D. 1986, The Astrophysical Journal, 302, 432

Longair, M. S., Ryle, M., & Scheuer, P. A. G. 1973, MNRAS, 164, 243

Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., et al. 2003, Publications of the Astronomical Society of the Pacific, 115, 897

Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J., & Crawford, F. 2007, Science, 318, 777

Lynden-Bell, D. 1969, Nature, 223, 690

Magliocchetti, M. 2022, Astronomy and Astrophysics Review, 30, 6

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, The Astronomical Journal, 115, 2285

Mahatma, V. H. 2023, Galaxies, 11, 74

Mahatma, V. H., Basu, A., Hardcastle, M. J., Morabito, L. K., & van Weeren, R. J. 2023, MNRAS, 520, 4427

Mahony, E. K., Morganti, R., Prandoni, I., et al. 2016, MNRAS, 463, 2997

Malarecki, J. M., Jones, D. H., Saripalli, L., Staveley-Smith, L., & Subrahmanyan, R. 2015, MNRAS, 449, 955

Mandal, S., Intema, H. T., van Weeren, R. J., et al. 2020, A&A, 634, A4

Mandal, S., Prandoni, I., Hardcastle, M. J., et al. 2021, A&A, 648, A5

Maneewongvatana, S. & Mount, D. 2002, Analysis of approximate nearest neighbor searching with clustered point sets, 105–123

Manners, J. C., Johnson, O., Almaini, O., et al. 2003, MNRAS, 343, 293

Mardia, K. V. 1972, Statistics of Directional Data (Academic Press), 18–24

Markevitch, M., Forman, W. R., Sarazin, C. L., & Vikhlinin, A. 1998, The Astrophysical Journal, 503, 77

Markevitch, M., Govoni, F., Brunetti, G., & Jerius, D. 2005, The Astrophysical Journal, 627, 733

Markevitch, M. & Vikhlinin, A. 2007, Physics Reports, 443, 1

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, The Astrophysical Journal, 619, L1

Martini, P., Miller, E. D., Brodwin, M., et al. 2013, The Astrophysical Journal, 768, 1

Mathews, P. M., Herring, T. A., & Buffett, B. A. 2002, Journal of Geophysical Research (Solid Earth), 107, 2068

Matthews, T. A. & Sandage, A. R. 1963, The Astrophysical Journal, 138, 30

Mauch, T. & Sadler, E. M. 2007, MNRAS, 375, 931

Mauduit, J. C., Lacy, M., Farrah, D., et al. 2012, Publications of the Astronomical Society of the Pacific, 124, 714

McAlpine, K., Smith, D. J. B., Jarvis, M. J., Bonfield, D. G., & Fleuren, S. 2012, MNRAS, 423, 132

McKay-Bukowski, D., Vierinen, J., Virtanen, I. I., et al. 2015, IEEE Transactions on Geoscience and Remote Sensing, 53, 1440

McMahon, R. G., Walton, N. A., Irwin, M. J., et al. 2001, New Astronomy Reviews, 45, 97

McNamara, B. R. & Nulsen, P. E. J. 2007, Annual Review of Astronomy and Astrophysics, 45, 117

McNamara, B. R. & Nulsen, P. E. J. 2012, New Journal of Physics, 14, 055023

Mechev, A., Oonk, J. B. R., Danezi, A., et al. 2017, in Proceedings of the International Symposium on Grids and Clouds (ISGC) 2017, 2

Mechev, A. P., Oonk, J. B. R., Shimwell, T., et al. 2018, arXiv e-prints, arXiv:1808.10735

Mevius, M. 2018, RMextract: Ionospheric Faraday Rotation calculator, Astrophysics Source Code Library, record ascl:1806.024

Miley, G. & De Breuck, C. 2008, Astronomy and Astrophysics Review, 15, 67

Mills, B. Y. 1952a, Australian Journal of Scientific Research A Physical Sciences, 5, 266

Mills, B. Y. 1952b, Australian Journal of Scientific Research A Physical Sciences, 5, 456

Mingo, B., Croston, J. H., Best, P. N., et al. 2022, MNRAS, 511, 3250

Mingo, B., Croston, J. H., Hardcastle, M. J., et al. 2019, MNRAS, 488, 2701

Mohan, N. & Rafferty, D. 2015, PyBDSF: Python Blob Detection and Source Finder, Astrophysics Source Code Library, record ascl:1502.007

Moldón, J., Deller, A. T., Wucknitz, O., et al. 2015, A&A, 574, A73

Morabito, L. K., Deller, A. T., Röttgering, H., et al. 2016, MNRAS, 461, 2676

Morabito, L. K. & Harwood, J. J. 2018, MNRAS, 480, 2726

Morabito, L. K., Jackson, N., de Jong, J., et al. 2025a, arXiv e-prints, arXiv:2502.06946

Morabito, L. K., Jackson, N. J., Mooney, S., et al. 2022a, A&A, 658, A1

Morabito, L. K., Kondapally, R., Best, P. N., et al. 2025b, MNRAS, 536, L32

Morabito, L. K., Sweijen, F., Radcliffe, J. F., et al. 2022b, MNRAS, 515, 5758

Morganti, R. 2017, Frontiers in Astronomy and Space Sciences, 4, 42

Mostert, R. I. J., Duncan, K. J., Alegre, L., et al. 2022, A&A, 668, A28

Mostert, R. I. J., Oei, M. S. S. L., Barkus, B., et al. 2024, arXiv e-prints, arXiv:2405.00232

Murgia, M., Govoni, F., Feretti, L., & Giovannini, G. 2010, A&A, 509, A86

Murgia, M., Govoni, F., Markevitch, M., et al. 2009, A&A, 499, 679

Murphy, E. J., Momjian, E., Condon, J. J., et al. 2017, The Astrophysical Journal, 839, 35

Muzzin, A., Wilson, G., Yee, H. K. C., et al. 2009, The Astrophysical Journal, 698, 1934

Myers, S. T. & Spangler, S. R. 1985, The Astrophysical Journal, 291, 52

Nair, V. & Hinton, G. E. 2010, in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 807–814

Napier, P. J. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 37

Nesvadba, N. P. H., Lehnert, M. D., De Breuck, C., Gilbert, A. M., & van Breugel, W. 2008, A&A, 491, 407

Netzer, H. 2013, The Physics and Evolution of Active Galactic Nuclei

Northover, K. J. E. 1973, MNRAS, 165, 369

Nunhokee, C. D., Bernardi, G., Manti, S., et al. 2021, arXiv e-prints, arXiv:2102.02900

Ocran, E. F., Taylor, A. R., Vaccari, M., Ishwara-Chandra, C. H., & Prandoni, I. 2020, MNRAS, 491, 1127

O'Dea, C. P. & Saikia, D. J. 2021, Astronomy and Astrophysics Review, 29, 3

Oei, M. S. S. L., Hardcastle, M. J., Timmerman, R., et al. 2024a, Nature, 633, 537

Oei, M. S. S. L., van Weeren, R. J., Gast, A. R. D. J. G. I. B., et al. 2023, A&A, 672, A163

Oei, M. S. S. L., van Weeren, R. J., Hardcastle, M. J., et al. 2022, A&A, 660, A2

Oei, M. S. S. L., van Weeren, R. J., Hardcastle, M. J., et al. 2024b, A&A, 686, A137

Offringa, A. R. 2016, A&A, 595, A99

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

Oliver, S., Rowan-Robinson, M., Alexander, D. M., et al. 2000, MNRAS, 316, 749

Oquab, M., Darcet, T., Moutakanni, T., et al. 2023, arXiv e-prints, arXiv:2304.07193

Osinga, E., van Weeren, R. J., Boxelaar, J. M., et al. 2021, A&A, 648, A11

Oswald, A. A. 1930, Transactions of the American Institute of Electrical Engineers, 49, 629

O'Toole, S. & Tocknell, J. 2022, FAIR standards for astronomical data

Owen, F. N. 2018, The Astrophysical Journal Supplement Series, 235, 34

Owen, F. N. & Morrison, G. E. 2008, The Astronomical Journal, 136, 1889

Owen, F. N. & Rudnick, L. 1976, The Astrophysical Journal, 205, L1

Padovani, P. 2016, Astronomy and Astrophysics Review, 24, 13

Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, Astronomy and Astrophysics Review, 25, 2

Parma, P., de Ruiter, H. R., & Fanti, R. 1996, in Extragalactic Radio Sources, ed. R. D. Ekers, C. Fanti, & L. Padrielli, Vol. 175, 137

Peacock, J. A., Miller, L., & Longair, M. S. 1986, MNRAS, 218, 265

Penzias, A. A. & Wilson, R. W. 1965, The Astrophysical Journal, 142, 419

Perley, R. A. 1999, in Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley

Petrosian, V. 2001, The Astrophysical Journal, 557, 560

Pignataro, G. V., Bonafede, A., Bernardi, G., et al. 2024, A&A, 685, L10

Pilkington, J. D. H. & Scott, J. F. 1965, Memoirs of the Royal Astronomical Society, 69, 183

Pinzke, A., Oh, S. P., & Pfrommer, C. 2017, MNRAS, 465, 4800

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, A&A, 550, A134

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A27

Prandoni, I., Gregorini, L., Parma, P., et al. 2001, A&A, 365, 392

Prandoni, I., Guglielmino, G., Morganti, R., et al. 2018, MNRAS, 481, 4548

Prescott, M., Mauch, T., Jarvis, M. J., et al. 2016, MNRAS, 457, 730

Quirrenbach, A. 2009, Experimental Astronomy, 26, 49

Rajpurohit, K., Brunetti, G., Bonafede, A., et al. 2021, A&A, 646, A135

Rajpurohit, K., Hoeft, M., van Weeren, R. J., et al. 2018, The Astrophysical Journal, 852, 65

Ramírez-Olivencia, N., Varenius, E., Pérez-Torres, M., et al. 2018, A&A, 610, L18

Reber, G. 1940, The Astrophysical Journal, 91, 621

Reber, G. 1944, The Astrophysical Journal, 100, 279

Rekier, J., Chao, B. F., Chen, J., et al. 2022, Surveys in Geophysics, 43, 149

Rengelink, R. B., Tang, Y., de Bruyn, A. G., et al. 1997, A&A, 124, 259

Resnik, D. B. & Hosseini, M. 2024, AI and Ethics

Retana-Montenegro, E., Röttgering, H. J. A., Shimwell, T. W., et al. 2018, A&A, 620, A74

Rigby, E. E., Best, P. N., & Snellen, I. A. G. 2008, MNRAS, 385, 310

Roland, J., Sol, H., Pauliny-Toth, I., & Witzel, A. 1981, A&A, 100, 7

Rosero, V., Hofner, P., Claussen, M., et al. 2016, The Astrophysical Journal Supplement Series, 227, 25

Rowan-Robinson, M. 2013, Night Vision

Rudnick, L. 2002, Publications of the Astronomical Society of the Pacific, 114, 427

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, Nature, 323, 533

Russakovsky, O., Deng, J., Su, H., et al. 2015, International Journal of Computer Vision (IJCV), 115, 211

Ryle, M. & Hewish, A. 1955, Memoirs of the Royal Astronomical Society, 67, 97

Ryle, M. & Hewish, A. 1960, MNRAS, 120, 220

Ryle, M. & Vonberg, D. D. 1946, Nature, 158, 339

Ryu, D., Kang, H., Hallman, E., & Jones, T. W. 2003, The Astrophysical Journal, 593, 599

Sabater, J., Best, P. N., Hardcastle, M. J., et al. 2019, A&A, 622, A17

Sabater, J., Best, P. N., Tasse, C., et al. 2021, A&A, 648, A2

Sadler, E. M. 2016, Astronomische Nachrichten, 337, 105

Sakelliou, I. & Ponman, T. J. 2004, MNRAS, 351, 1439

Schinckel, A. E., Bunton, J. D., Cornwell, T. J., Feain, I., & Hay, S. G. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8444, Ground-based and Airborne Telescopes IV, ed. L. M. Stepp, R. Gilmozzi, & H. J. Hall, 84442A

Schmidt, M. 1963, Nature, 197, 1040

Schmidt, M. 1968, The Astrophysical Journal, 151, 393

Schmidt, M. 1970, The Astrophysical Journal, 162, 371

Schoenmakers, A. P., de Bruyn, A. G., Röttgering, H. J. A., & van der Laan, H. 2000, MNRAS, 315, 395

Schwab, F. R. 1984, The Astronomical Journal, 89, 1076

Sereno, M. 2016, LIRA: LInear Regression in Astronomy, Astrophysics Source Code Library, record ascl:1602.006

Sexton, R. O., Secrest, N. J., Johnson, M. C., & Dorland, B. N. 2022, The Astrophysical Journal, 260, 33

Shakeshaft, J. R., Ryle, M., Baldwin, J. E., Elsmore, B., & Thomson, J. H. 1955, Memoirs of the Royal Astronomical Society, 67, 106

Shimwell, T. W., Brown, S., Feain, I. J., et al. 2014, MNRAS, 440, 2901

Shimwell, T. W., Hale, C. L., Best, P. N., et al. 2025, arXiv e-prints, arXiv:2501.04093

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, A&A, 622, A1

Singal, A. K. & Rajpurohit, K. 2014, MNRAS, 442, 1656

Sirothia, S. K., Dennefeld, M., Saikia, D. J., et al. 2009, MNRAS, 395, 269

Skipper, C. 2014, Time and channel averaging, Technical report, University of Southampton

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, The Astronomical Journal, 131, 1163

Smirnov, O. M. 2011a, A&A, 527, A106

Smirnov, O. M. 2011b, A&A, 527, A107

Smirnov, O. M. & Tasse, C. 2015, MNRAS, 449, 2668

Smith, D. J. B., Dunne, L., Maddox, S. J., et al. 2011, MNRAS, 416, 857

Snellen, I. A. G. & Best, P. N. 2001, MNRAS, 328, 897

Sob, U. M., Bester, H. L., Smirnov, O. M., Kenyon, J. S., & Russeeawon, C. 2021, MNRAS, 504, 1714

Song, X., Xu, X., & Yan, P. 2024, arXiv e-prints, arXiv:2402.15687

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Sullivan, W. T. 2009, Experimental Astronomy, 25, 107

Sutherland, W. & Saunders, W. 1992, MNRAS, 259, 413

Swarup, G., Ananthakrishnan, S., Kapahi, V. K., et al. 1991, Current Science, 60, 95

Sweijen, F., Lyu, Y., Wang, L., et al. 2023, A&A, 671, A85

Sweijen, F., Morabito, L. K., Harwood, J., et al. 2022a, A&A, 658, A3

Sweijen, F., Morabito, L. K., Harwood, J., et al. 2022b, A&A, 658, A3

Sweijen, F., van Weeren, R. J., Röttgering, H. J. A., et al. 2022c, Nature Astronomy, 6, 350

Swetz, D. S., Ade, P. A. R., Amiri, M., et al. 2011, The Astrophysical Journal Supplement Series, 194, 41

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. 2015, arXiv e-prints, arXiv:1512.00567

Tahir, J., Ganguli, S., & Rotskoff, G. M. 2024, arXiv e-prints, arXiv:2410.08194

Takizawa, M. 1999, The Astrophysical Journal, 520, 514

Tasse, C. 2014a, arXiv e-prints, arXiv:1410.8706

Tasse, C. 2014b, A&A, 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, A&A, 648, A1

The CASA Team et al. 2022, Publications of the Astronomical Society of the Pacific, 134, 114501

Thomas, N. & Davé, R. 2022, Monthly Notices of the Royal Astronomical Society, 515, 5539

Thompson, A. R., Clark, B. G., Wade, C. M., & Napier, P. J. 1980, The Astrophysical Journal Supplement Series, 44, 151

Tielens, A. G. G. M., Miley, G. K., & Willis, A. G. 1979, A&A, 35, 153

Timmerman, R., van Weeren, R. J., Botteon, A., et al. 2022a, A&A, 668, A65

Timmerman, R., van Weeren, R. J., Callingham, J. R., et al. 2022b, A&A, 658, A5

Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, Publications of the Astronomical Society of Australia, 30, e007

Tremblay, S. E., Taylor, G. B., Ortiz, A. A., et al. 2016, MNRAS, 459, 820

van Diepen, G., Dijkema, T. J., & Offringa, A. 2018a, DPPP: Default Pre-Processing Pipeline, Astrophysics Source Code Library, record ascl:1804.003

van Diepen, G., Dijkema, T. J., & Offringa, A. 2018b, DPPP: Default Pre-Processing Pipeline, Astrophysics Source Code Library, record ascl:1804.003

van Diepen, G. & Farris, A. 1994, in Astronomical Society of the Pacific Conference Series, Vol. 61, Astronomical Data Analysis Software and Systems III, ed. D. R. Crabtree, R. J. Hanisch, & J. Barnes, 417

van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2

van Weeren, R. J., Andrade-Santos, F., Dawson, W. A., et al. 2017, Nature Astronomy, 1, 0044

van Weeren, R. J., Brunetti, G., Brüggen, M., et al. 2016a, The Astrophysical Journal, 818, 204

van Weeren, R. J., de Gasperin, F., Akamatsu, H., et al. 2019, Space Science Reviews, 215, 16

van Weeren, R. J., Fogarty, K., Jones, C., et al. 2013, The Astrophysical Journal, 769, 101

van Weeren, R. J., Shimwell, T. W., Botteon, A., et al. 2021, A&A, 651, A115

van Weeren, R. J., Timmerman, R., Vaidya, V., et al. 2024, arXiv e-prints, arXiv:2410.02863

van Weeren, R. J., Williams, W. L., Hardcastle, M. J., et al. 2016b, The Astrophysical Journal, 223, 2

van Weeren, R. J., Williams, W. L., Hardcastle, M. J., et al. 2016c, The Astrophysical Journal, 223, 2

Varenius, E., Conway, J. E., Martí-Vidal, I., et al. 2016, A&A, 593, A86

Varenius, E., Conway, J. E., Martí-Vidal, I., et al. 2015, A&A, 574, A114

Vazza, F., Ettori, S., Roncarelli, M., et al. 2019, A&A, 627, A5

Venkattu, D., Lundqvist, P., Pérez Torres, M., et al. 2023, The Astrophysical Journal, 953, 157

Venturi, T., Giacintucci, S., Merluzzi, P., et al. 2022, A&A, 660, A81

Villenave, M., Ménard, F., Dent, W. R. F., et al. 2020, A&A, 642, A164

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261

Vivian, J., Rao, A., Nothaft, F., et al. 2017, Nature Biotechnology, 35, 314

Vollmer, B., Gassmann, B., Derrière, S., et al. 2010, A&A, 511, A53

Vrublevskis, A., Donerblics, M., Ryabov, B., & Bezrukovs, D. 2020, in The development and the proposed research of LOFAR-Latvia

Wayth, R. B., Lenc, E., Bell, M. E., et al. 2015, Publications of the Astronomical Society of Australia, 32, e025

Weyand, T., Araujo, A., Cao, B., & Sim, J. 2020, arXiv e-prints, arXiv:2004.01804

Wijnholds, S. J., Willis, A. G., & Salvini, S. 2018, MNRAS, 476, 2029

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. 2016, Scientific data, 3, 160018

Williams, W. L., Hardcastle, M. J., Best, P. N., et al. 2019, A&A, 622, A2

Williams, W. L., van Weeren, R. J., Röttgering, H. J. A., et al. 2016, MNRAS, 460, 2385

Willis, A. G., Strom, R. G., & Wilson, A. S. 1974, Nature, 250, 625

Willott, C. J., Rawlings, S., Blundell, K. M., Lacy, M., & Eales, S. A. 2001, MNRAS, 322, 536

Willson, M. A. G. 1970, MNRAS, 151, 1

Wilson, G., Muzzin, A., Yee, H. K. C., et al. 2009, The Astrophysical Journal, 698, 1943

Windhorst, R., Mathis, D., & Neuschaefer, L. 1990, in Astronomical Society of the Pacific Conference Series, Vol. 10, Evolution of the Universe of Galaxies, ed. R. G. Kron, 389–403

Wing, J. D. & Blanton, E. L. 2011, The Astronomical Journal, 141, 88

Ye, H., Gull, S. F., Tan, S. M., & Nikolic, B. 2022, MNRAS, 510, 4110

Ye, H., Sweijen, F., van Weeren, R. J., et al. 2024, A&A, 691, A347

Yoo, A. B., Jette, M. A., & Grondona, M. 2003, in Job Scheduling Strategies for Parallel Processing, ed. D. Feitelson, L. Rudolph, & U. Schwiegelshohn (Berlin, Heidelberg: Springer Berlin Heidelberg), 44–60

Yusef-Zadeh, F. 2012, The Astrophysical Journal Letters, 759, L11

Zheng, X. C., Röttgering, H. J. A., Best, P. N., et al. 2020, A&A, 644, A12

Zheng, X. C., Röttgering, H. J. A., van der Wel, A., & Duncan, K. 2022, A&A, 665, A114

ZuHone, J. A., Brunetti, G., Giacintucci, S., & Markevitch, M. 2015, The Astrophysical Journal, 801, 146

Popular scientific summary

Radio from the heavens

While investigating sources of noise that could interfere with telecommunications, physicist and engineer Karl Jansky unexpectedly detected in 1933 radio emission at a frequency of 20.5 MHz coming from the centre of our galaxy. With this discovery, he unknowingly opened a new window to explore the Universe at radio frequencies. Jansky's discovery was a few years later confirmed and expanded upon by engineer and amateur radio astronomer Grote Reber, who detected signals from the same direction as Jansky at 162 MHz, along with radio emissions from the Andromeda galaxy and other areas in the sky. The following post-Second World War era marked the true emergence of radio astronomy, fueled by advancements in radar technology developed during the war and a subsequent shift toward peaceful scientific endeavours.

Key discoveries

This new window to study our Universe at radio frequencies led to many remarkable discoveries. The largest and brightest objects were identified first, including galaxies and remnants of powerful star explosions within our own galaxy. Soon after, other phenomena were uncovered, such as the hydrogen line, which allows astronomers to explore the structure and dynamics of galaxies and the early universe. Another ground-breaking discovery was the cosmic microwave background, providing a snapshot of the faint afterglow of the Big Bang. Pulsars – rapidly rotating neutron stars that emit beams of radiation – were also detected, offering a precise cosmic clock and deepening our understanding of extreme physics. More recently, fast radio bursts, mysterious and powerful flashes of radio waves, have become a new exciting topic in radio astronomy, while even brown dwarfs have been detected with radio telescopes as well. The following two sections will highlight discoveries at radio wavelengths most relevant to this thesis.

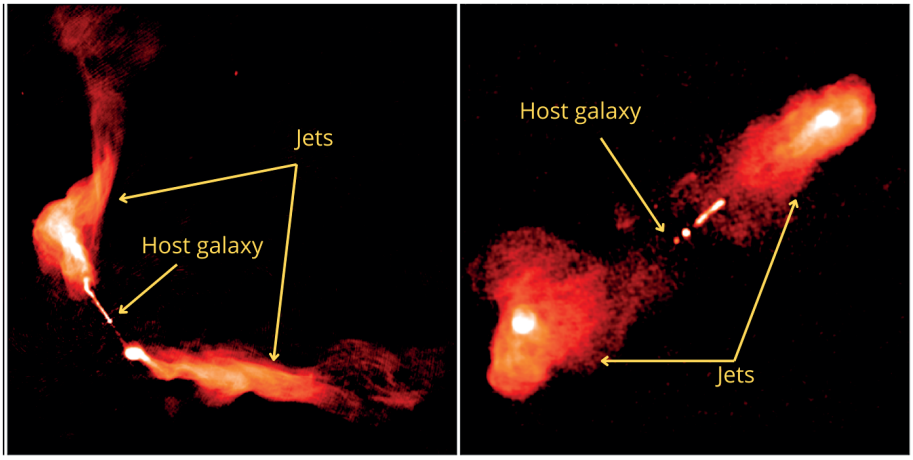


Figure I: Examples of a core-brightened jet (*left*) and edge-brightened jet (*right*) corresponding to their host galaxies. Source: Hardcastle & Croston (2020).

Radio-emitting galaxies

A major early discovery in radio astronomy was the identification of bright radio-emitting sources initially called 'quasi-stellar'. This name was given to these objects because their counterparts, detected with optical telescopes, appeared star-like. These were later commonly known as 'quasars'. Measurements of quasar distances revealed that their radio emissions originated from distant galaxies. It is now understood that their powerful radio emissions result from large amounts of material being drawn toward a supermassive black hole at the galaxy's centre. This material forms a hot, rotating disk, which radiates energy in particular at wavelengths associated with hot ultra-violet (UV) and X-ray radiation.

The radio-emitting galaxies exhibit in many cases jet-like structures emerging from their cores, as indicated in Figure I. These jets can extend far beyond their host galaxies, injecting vast amounts of energy into the surrounding mediums. This energy can disrupt the formation of new stars and so prevent the growth of galaxies. Jetted radio-emitting galaxies are classified by their morphological appearances into Fanaroff-Riley type I (core-brightened) and type II (edge-brightened). Core-brightened jets are characterised by having the brightest parts of their jet near the host galaxy while diffusing out into more plume-like structures towards at the outer edges. The edge-brightened jets exhibit the brightest parts (also called hotspots) of their jet closer to the outer edges. Figure I presents examples of both types. This difference in appearance is attributed to a combination of the power of their jets and the environments in which they reside: Core-brightened jets lack the power

to penetrate through their (dense) surrounding environment, causing their jets to diffuse outward, while the more powerful edge-brightened jets remain strong enough to push through their environment, which keeps their hotspots at the outer parts of their jets. Consequently, galaxies with core-brightened jets are typically found in dense environments such as galaxy clusters, while galaxies with edge-brightened jets are more often located in isolated regions.

Although our understanding of these radio-emitting galaxies has improved over time, the complete picture to explain the formation and evolution of their jets is still not fully understood. This is partly due to observational limitations in resolution and sensitivity, which introduced biases that have led to several revisions on the origins of their jets. This highlights the need for more comprehensive data, including larger samples of radio-emitting galaxies observed with more sensitive and high-resolution telescopes.

Radio bridges

Groups of galaxies are called galaxy clusters. These large-scale cosmic structures can consist of hundreds to thousands of galaxies bound together by gravity, making them the largest gravitationally bound objects in the Universe. They often contain not just galaxies, but also vast amounts of hot gas, dark matter, and other intergalactic material. An important process in the evolution of our Universe is the merging of galaxy clusters to form larger structures. The merging processes generate large-scale faint radio emission that can be detected by our telescopes. The source of this emission can be attributed to several physical processes, one of which involves shock-like behaviour, while the other involves randomised turbulence in the presence of magnetic fields.

More recently, faint radio emission between galaxy clusters that are about to merge (in a pre-merging phase) has been detected as well at very low radio frequencies. An example of such radio emission bridging a pair of pre-merging clusters is presented in Figure II. While some have proposed that shock-like behaviour is responsible for their origin, more recent work suggests that the emission may arise from turbulence. However, since only a few direct observations from two confirmed radio bridges have been analyzed, further studies are required to fully understand their physical origin and prevalence.

Interferometers and the Low Frequency Array

Multiple telescopes observing across the radio spectrum have been developed over the last 100 years. Among the most famous are the Very Large Array (VLA),

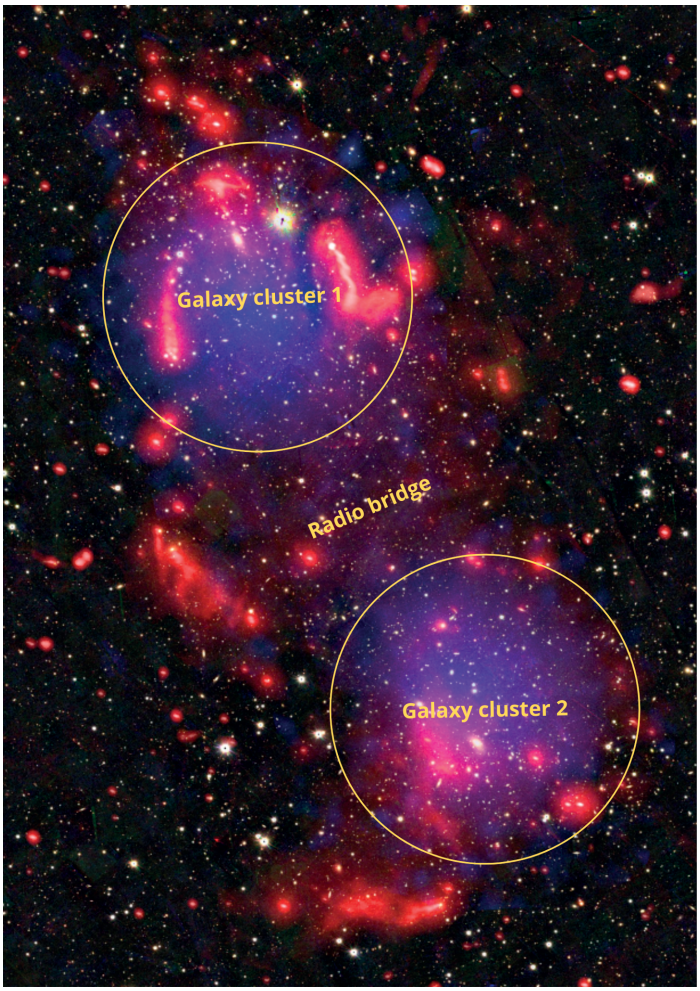


Figure II: Radio bridge in between a pair of pre-merging galaxy clusters. The red colour corresponds to radio emission, the blue colour corresponds to X-ray emission, and the background is optical data. This image is designed by F. Sweijen, using data from Chapter 2.

the Westerbork Synthesis Radio Telescope (WSRT), and the Atacama Cosmology Telescope (ALMA). These are large instruments that make use of interferometry to improve resolution. This is a technique where multiple antennas or dishes are working together to observe the same object. Combining the observed signals from different locations on Earth, allows astronomers to simulate a much larger telescope to study objects at much finer resolutions. While an instrument such as the VLA gave us high-quality radio data of our Universe since the 1980s, there was no data available at comparable quality at wavelengths larger than one meter (below about



Figure III: The LOFAR Superterp is located in Exloo, Drenthe, in the Netherlands. The word 'terp' comes from the Dutch language and refers to an artificially elevated area created by early settlers in the Netherlands, particularly in the northern regions, to protect their homes and livestock from flooding in low-lying areas. In the context of LOFAR, the Superterp is a larger, more concentrated LOFAR station located within the core area of the network, built on elevated land and surrounded by a canal to protect it from wildlife discovering this tool to study the radio universe as well. The large grey tiles cover the high-band antennas, while the smaller, more randomly distributed squares represent the low-band antennas. Source: ASTRON.

300 MHz). This frequency regime conceals valuable information about astrophysical processes related to diffuse radio sources from the previously mentioned jets and radio bridges. Fortunately, due to technological advancements and growing interest (and the availability of financial resources), this has since the 2000s changed and several interferometers observing at low radio frequencies have been developed. One of the main instruments observing below 250 MHz is the Low-Frequency Array (LOFAR).

LOFAR is an instrument equipped with two types of antennas: low-band antennas that detect frequencies between 10 and 80 MHz, and high-band antennas that detect frequencies from 110 to 240 MHz. A LOFAR station and the two antenna types are depicted in the images in Figures III and IV. None of LOFAR's components, including its antennas, have moving parts, as it uses technology to digitally steer its focus across the sky. This means that all the data processing happens within computers, making the construction of its antennas relatively inexpensive compared to other radio telescopes. The low construction costs and international



Figure IV: Different LOFAR antennas. On the left, a low-band antenna is shown with a human for scale. In the middle below, an image of a high-band antenna is displayed, which is concealed beneath the tiles on the right, with humans standing on top for scale. These pictures were taken at the PL612 station in Bałdy (Poland) by the author and his colleagues.

collaborations have contributed to the rapid expansion of LOFAR with stations across Europe. LOFAR's core is located in the Netherlands, with international stations in Germany, Poland, England, Ireland, Sweden, Latvia, France, and Finland.[16] Soon, LOFAR will also be expanded to Italy and Bulgaria. By utilizing all of these stations, LOFAR effectively simulates a radio telescope the size of the European continent, offering exceptionally high resolution in addition to its impressive sensitivity due to its dense array of antennas.

Powerful computers and advanced software

LOFAR can capture images of large areas of the sky, up to tens of square degrees in a single observation,[17] allowing it to detect many radio-emitting sources simultaneously. However, creating these large images poses significant computational challenges, particularly when combining data from all of LOFAR's international stations. The computational complexity arises from the need to correct various effects that distort low-frequency observations, one of the most significant being the ionosphere – a layer of Earth's atmosphere filled with charged particles that interfere with propagating radio waves. Correcting for these distortions requires a carefully designed strategy, and once calibration is complete, the data must be transformed into images – a process known as 'imaging' – to produce data ready for scientific

[16]The Finnish station in Kilpisjärvi operates independently and is not connected to the LOFAR network.

[17]For comparison: the diameter of the moon covers roughly 0.5 square degrees

analysis. All these steps need specialised software.

Over the past decade, most of the software development efforts have focused solely on processing data from the Dutch LOFAR stations, as this is computationally less demanding while still yielding valuable insights into the Universe. The next major step is to extend this capability to automatically process data from all LOFAR stations. This will enable us to conduct a full survey of the northern sky at a much higher resolution (see Figure V) and to target specific regions with very long exposure times by combining multiple observations of the same area, without the need for manual intervention to produce high-quality science-ready images. Each of these observations produces about 16 terabytes (TB) of data.[18] Processing these amounts of data cannot be done on standard laptops or desktops. Instead, it requires the use of large high-performance computing clusters designed to handle the immense computational load efficiently (see Figure VI).

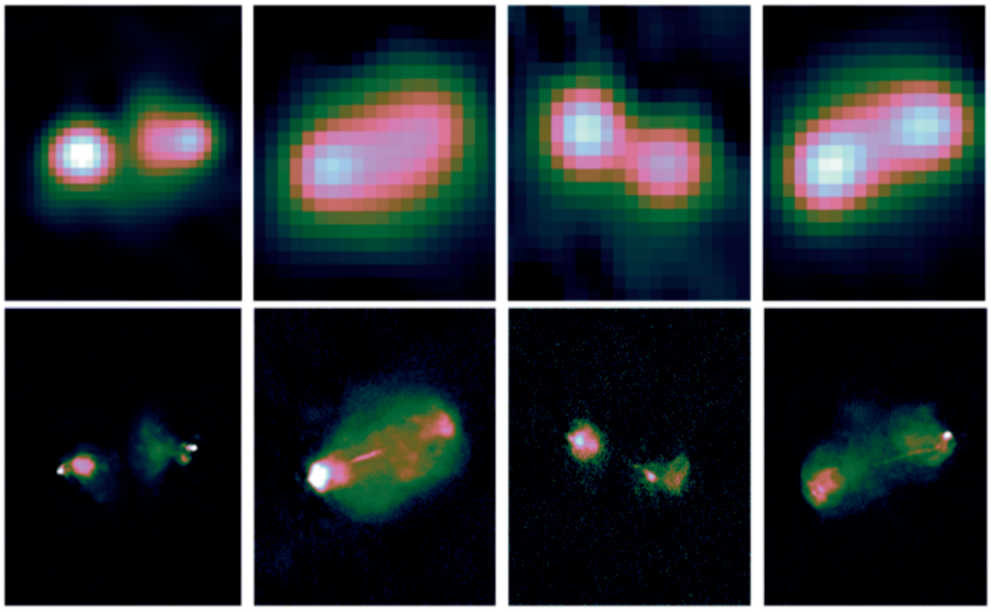


Figure V: Resolution improvement when going from Dutch-only LOFAR station imaging (first row) to imaging with the full international LOFAR (second row). Columns correspond to the same radio source. Source: de Jong et al. (2024).

[18] 16 TB is enough storage to hold around 4 million photos from a modern smartphone, 8,000 hours of HD video, or over 100 million pages of text documents. An average laptop has on average about 1 to 2 TB.

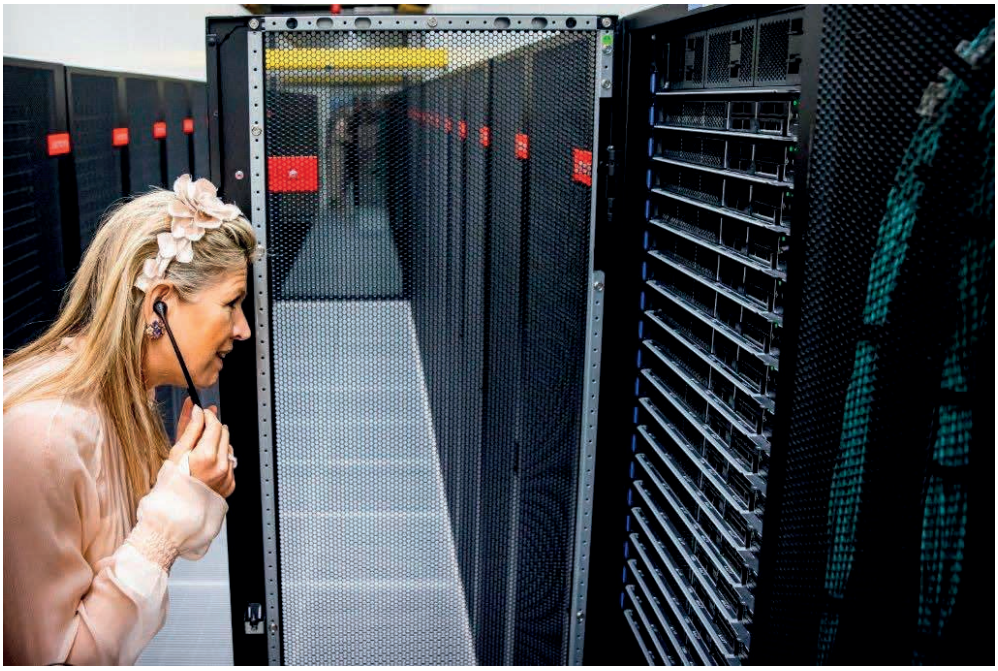


Figure VI: Queen Máxima of the Netherlands inaugurating the Snellius supercomputer in Amsterdam. This supercomputer is part of the SURF computing infrastructure. Using the GPUs of this system, we trained a neural network to detect when the calibration of our radio data has converged. Source: Het Koninklijk Huis.

This thesis

We begin this thesis in **Chapter 2** at the larger scales of our Universe, where we combine six observations from the Dutch LOFAR stations to create the deepest image to date of a radio bridge between a pair of galaxy clusters. These clusters are in the process of merging into a single galaxy cluster. At their current pre-merging stage, they exhibit radio emission that forms a bridge between them. This unique phenomenon was only recently discovered, and our work aims to follow up on it to better understand its origin, using our new deep image. Our findings suggest a turbulent re-acceleration process, where particles gain energy through interactions with turbulence in the presence of magnetic fields, causing them to emit radiation at radio frequencies.

We then step to smaller scales by gathering known core-brightened (FRI) and edge-brightened (FRII) radio-emitting galaxies with powerful jets detected by LOFAR's Dutch stations, along with their cosmic distances measured by other telescopes. This allows us to investigate in **Chapter 3** how these two types of jets

have evolved over cosmic time. We find that the earlier Universe was more densely populated by the brightest sources of both morphologies, likely due to the greater availability of gas in the past when the Universe was also smaller. However, we also observe a decline in the number density of fainter sources, a result that is not easily explained. We suggest that this could be due to the resolution limits of our data, motivating the need for higher-resolution images of our sky to better understand this trend.

Driven by the need for higher-resolution images, we focus in **Chapter 4** on processing data from four observations of a single sky area, taken with all Dutch and international LOFAR stations. Combining multiple observations enables the creation of a deeper image which is needed to enable the detection of more distant and faint radio sources.[19] However, this also presents challenges, as it provides us with four times more data to process. To streamline the data reduction strategy, we develop techniques to automatically select areas of the sky with sufficient signal to optimise our data processing. This automatic selection is crucial, as stronger signals provide more information about how the sky changes over time and frequency due to distortions, allowing us to correct better for these with specialised software. After completing all data processing, we achieved the deepest image over a 2.5 by 2.5 degrees sky area with a 20-fold improvement in resolution compared to making images with only the Dutch LOFAR stations (see Figure V), meaning that our image pixels cover 400 times smaller area than those in the data used in Chapter 3.[20]

Although we achieved in the previous chapter the production of the deepest and high-resolution image with LOFAR to date, our work revealed that the computational costs are too high to scale this approach effectively (and responsibly, given the large energy consumption). To tackle this issue, we revisit in **Chapter 5** a technique to reduce data sizes when combining multiple observations of the same sky area. This method makes use of the sidereal day – the period of approximately 23 hours and 56 minutes during which the Earth completes one full rotation around its axis. This allows us to average data samples from our observations, corresponding to a similar sidereal timestamp, such that we reduce the data volume without significant loss in information. In this way, our method enables much faster processing when converting the calibrated data into a final sky image.

Finally, in **Chapter 6**, we present and discuss additional advancements to improve the processing of LOFAR's vast data volumes. This includes quality improvements, implementing a framework for automatic processing on high-performance

[19] Similar to setting a longer exposure time on your camera to catch more light during darker evenings.

[20] This also means that the final image, with its size of 32 GB, is too large to be opened on a standard laptop.

computing systems, and utilizing artificial intelligence (AI) to automatically determine when calibration is sufficiently refined, eliminating the need for human intervention. Moreover, we optimise the most computationally intensive aspects of our workflow, making the entire process up to approximately four times faster. These improvements will ultimately allow us to process much larger and many more datasets in a more efficient and cost-effective manner.

Populairwetenschappelijke samenvatting

Radio uit de hemel

Tijdens het onderzoeken van bronnen van ruis die telecommunicatie konden verstoren, ontdekte natuurkundige en ingenieur Karl Jansky in 1933 onverwachts radiostraling op een frequentie van 20,5 MHz vanuit de richting van het centrum van onze Melkweg. Met deze ontdekking opende hij een nieuw venster om het Universum te bestuderen via radiogolven. Jansky's ontdekking werd een paar jaar later bevestigd en verder uitgewerkt door ingenieur en amateur radioastronoom Grote Reber. Hij detecteerde signalen uit dezelfde richting als Jansky op een frequentie van 162 MHz, als ook radiostraling van het Andromedastelsel en andere gebieden aan de hemel. De ontwikkeling van de radioastronomie raakte in een versnelling na de Tweede Wereldoorlog, aangedreven door vooruitgang in radartechnologie die tijdens de oorlog was ontwikkeld en een daaropvolgende aandachtsverschuiving richting meer 'vreedzame' wetenschappelijke activiteiten.

Belangrijke ontdekkingen

Dit nieuwe venster om het Universum te bestuderen via radiogolven heeft geleid tot vele belangrijke ontdekkingen. De grootste en helderste objecten werden door de eerste radioastronomen als eerste geïdentificeerd, waaronder radiostraling van sterrenstelsels en van supernovaresten binnen onze eigen Melkweg. Kort daarna werden andere fenomenen ontdekt, zoals de waterstoflijn, die astronomen in staat stelt de structuur en dynamiek van sterrenstelsels en het vroege Universum te onderzoeken. Een andere grote ontdekking was de kosmische achtergrondstraling, afkomstige van de zwakke nagloed van de Oerknal. Pulsars – snel roterende neutronensterren die bundels straling uitzenden – werden ook ontdekt, waardoor ons begrip van extreme fysica werd verdiept. Meer recent zijn ook snelle radioflitsen, mysterieuze en krachtige uitbarstingen van radiogolven ontdekt. Zelfs signalen afkomstig van bruine dwergen kunnen met radiotelescopen worden waargenomen. De volgende twee secties belichten ontdekkingen op radiogolflengten die het meest relevant zijn voor

dit proefschrift.

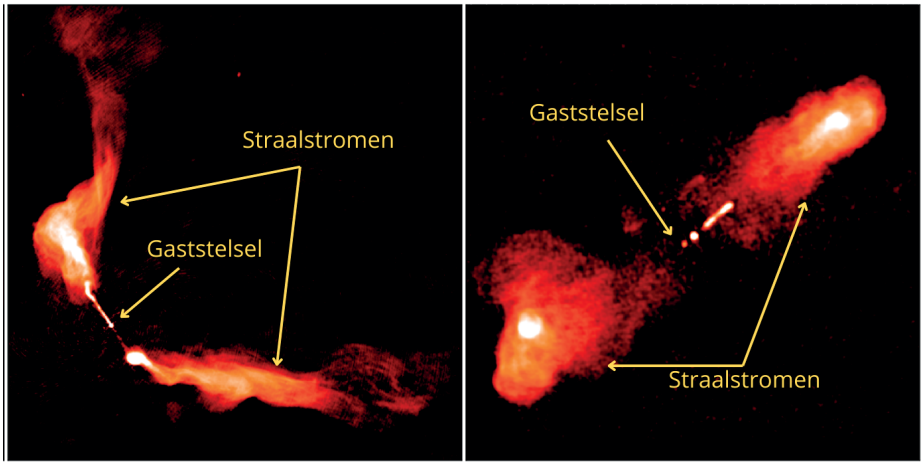


Figure I: Voorbeelden van kernheldere (*links*) en randheldere (*rechts*) straalstromen, met aanduiding van het gaststelsel en straalstromen. Bron: Hardcastle & Croston (2020).

Radiostralende sterrenstelsels

Een van de belangrijkste vroege ontdekkingen in de radioastronomie was de identificatie van heldere radiobronnen die aanvankelijk 'quasi-stellair' werden genoemd. Deze objecten kregen deze naam, omdat de geassocieerde gaststelsels, waargenomen met optische telescopen, sterachtig leken. Deze werden later bekend als 'quasars'. Het meten van de afstanden tot quasars onthulde dat de radiostraling in werkelijkheid kwam van verre sterrenstelsels. Het is nu bekend dat hun krachtige radiostraling het gevolg is van grote hoeveelheden materiaal die naar een superzwaar zwart gat in het centrum van het sterrenstelsel worden getrokken. Dit materiaal vormt een hete, roterende schijf die met name energie uitstraalt bij golflengten die geassocieerd worden met hete ultraviolet (UV) en röntgenstraling.

De radiostralende sterrenstelsels vertonen in veel gevallen straalstroomstructuren die vanuit de kern van het gaststelsel ontstaan, zoals aangegeven in Figuur I. Deze straalstromen kunnen zich ver buiten hun gaststelsels uitstrekken en enorme hoeveelheden energie in de omliggende omgeving injecteren. Deze energie kan de vorming van nieuwe sterren verstoren of zelfs de groei en ontwikkeling van andere sterrenstelsels verhinderen. Sterrenstelsels met straalstromen worden op basis van hun vorm (morfologie) geclassificeerd als Fanaroff-Riley type I (kernhelder) en type II (randhelder). Kernheldere straalstromen worden gekarakteriseerd door de helderste delen van hun straalstroom dicht bij het gaststelsel, die vervolgens overgaan in

meer pluimachtige structuren naar de buitenranden toe. Randheldere straalstromen vertonen de helderste delen van hun straalstromen dichter bij de buitenranden. Figuur I toont voorbeelden van beide types. Dit verschil in verschijning wordt momenteel toegeschreven aan een combinatie van de kracht van de straalstromen en de omgevingen waarin een sterrenstelsel zich bevindt. De kernheldere straalstromen zijn namelijk niet krachtig genoeg om door hun (dichte) omgeving heen te dringen, waardoor hun straalstromen diffuus naar buiten uitwaaieren. De randheldere straalstromen zijn echter krachtig genoeg om door hun omgeving heen te dringen en behouden zo hun helderste plekken aan de rand van de straalstroom die door het omliggende medium heendringt. Hierdoor worden sterrenstelsels met kernheldere straalstromen meestal gevonden in 'drukkere' omgevingen, zoals clusters van sterrenstelsels, terwijl sterrenstelsels met randheldere straalstromen vaker in geïsoleerde gebieden voorkomen.

Hoewel ons begrip van deze radiostralende sterrenstelsels in de loop der tijd is verbeterd, is onze kennis over de vorming en evolutie van hun straalstromen nog steeds onvolledig. Dit komt deels door waarnemingsbeperkingen in resolutie en gevoeligheid, die, na de ontwikkeling van betere instrumenten, tot meerdere herzieningen van de oorsprong van straalstromen hebben geleid. Dit benadrukt de wens voor het verzamelen van grotere steekproeven van radiostralende sterrenstelsels op basis van data van hogere kwaliteit, verkregen door gevoeligere telescopen die scherpere beelden kunnen maken.

Radiobruggen

Groepen van sterrenstelsels worden sterrenstelselclusters genoemd. Deze grootschalige kosmische structuren kunnen bestaan uit honderden tot duizenden sterrenstelsels die door de zwaartekracht bij elkaar worden gehouden, waardoor ze de grootste door zwaartekracht gebonden objecten in het Universum zijn. Ze bevatten vaak niet alleen sterrenstelsels, maar ook enorme hoeveelheden heet gas, donkere materie en ander intergalactisch materiaal. Een belangrijk proces in de evolutie van ons Universum is het samensmelten van sterrenstelselclusters om zo grotere structuren te vormen. De samenvoegingsprocessen genereren grootschalige, zwakke radiostraling die door onze telescopen kan worden waargenomen. De bron van deze straling kan worden toegeschreven aan verschillende effecten in de aanwezigheid van magnetische velden, zoals schokgolfachtig gedrag en willekeurige turbulentie.

Recentelijk is ook zwakke radiostraling gedetecteerd tussen sterrenstelselclusters die op het punt staan samen te smelten. Een voorbeeld van dergelijke radiostraling die een brug vormt tussen een paar sterrenstelselclusters wordt getoond in Figuur II. Terwijl sommigen hebben voorgesteld dat schokgolfachtig gedrag verantwoordelijk is

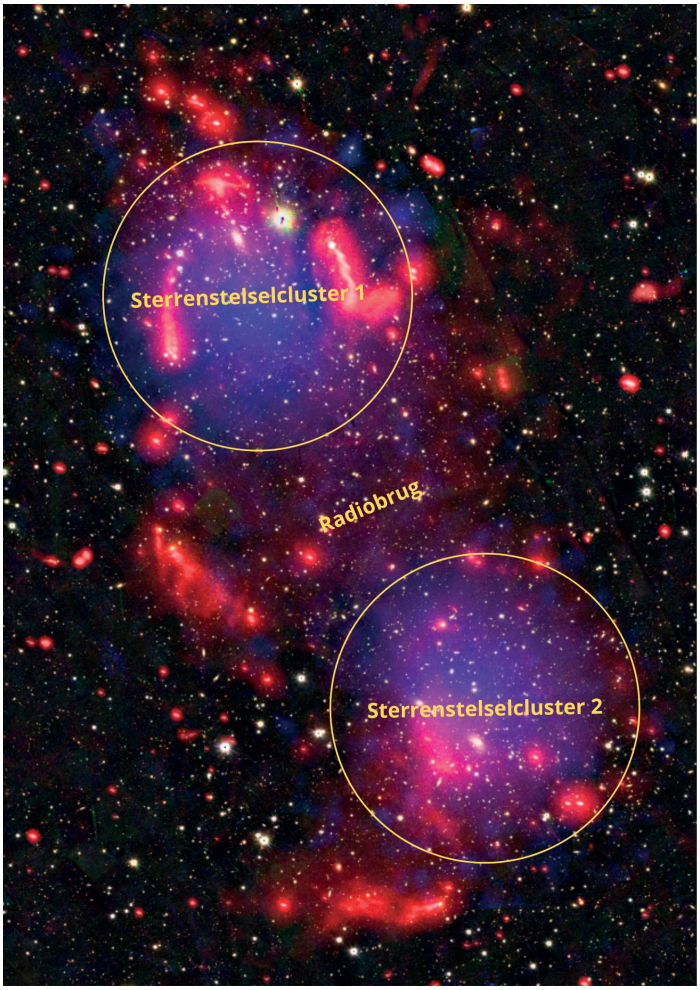


Figure II: Radiobrug tussen een paar sterrenstelselclusters die op het punt staan om samen te smelten. De rode kleur komt overeen met radiostraling, de blauwe kleur met röntgenstraling en de achtergrond toont optische data. Deze afbeelding is ontworpen door F. Sweijen, met gebruik van data uit Hoofdstuk 2.

voor hun oorsprong, suggereert recenter onderzoek dat de straling kan ontstaan door turbulent gedrag. Aangezien er tot nu toe slechts een paar directe waarnemingen van twee radiobruggen zijn geanalyseerd, zijn meer studies nodig om hun fysieke oorsprong en verspreiding volledig te begrijpen.



Figure III: De LOFAR Superterp bevindt zich in Exloo, Drenthe, in Nederland. Het woord 'terp' komt uit het Nederlands en verwijst naar een kunstmatige verhoging die werd gemaakt door vroege bewoners van Nederland, om zo hun huizen en vee te beschermen tegen overstromingen in laaggelegen gebieden. In de context van LOFAR is de Superterp een groter LOFAR station dat zich bevindt in het centrale gebied van het netwerk. Het is gebouwd op verhoogd terrein en omringd door een kanaal om het te beschermen tegen wilde dieren, die dit instrument anders mogelijk ook zouden gebruiken om het radiouniversum te bestuderen. De grote grijze tegels bedekken de HBA antennes, terwijl de kleinere, meer willekeurig verdeelde vierkanten de LBA antennes zijn. Bron: ASTRON.

Interferometers en de Low Frequency Array

In de afgelopen 100 jaar zijn er meerdere telescopen ontwikkeld die waarnemingen doen in het radiospectrum. Tot de bekendste behoren de Very Large Array (VLA), Westerbork Synthese Radio Telescoop (WSRT), en de Atacama Cosmology Telescope (ALMA). Dit zijn instrumenten die gebruikmaken van interferometrie om de resolutie te verbeteren. Dit is een techniek waarbij meerdere antennes of schotels samenwerken om hetzelfde object te observeren. Door de waargenomen signalen van verschillende locaties op aarde te combineren, kunnen astronomen een veel grotere telescoop simuleren om objecten met een veel fijnere resoluties te bestuderen. Hoewel een instrument zoals de VLA ons sinds de jaren 80 hoogwaardige radiogegevens van ons Universum heeft geleverd, waren er geen gegevens van vergelijkbare kwaliteit beschikbaar bij golflengten groter dan een meter (onder ongeveer 300 MHz). Dit radiofrequentiedomein verhult waardevolle informatie over astrofysische processen van diffusere radiobronnen van onder andere de eerder



Figure IV: Verschillende LOFAR antennes. Links is een LBA antenne te zien met een mens er naast om de schaal te verduidelijken. In het midden wordt een afbeelding van een HBA antenne getoond, die verborgen ligt onder de tegels getoond in de rechter afbeelding, waar mensen bovenop staan om de schaalgrootte aan te geven. Deze foto's zijn genomen bij station PL612 in Bałdy (Polen) door de auteur en zijn collega's.

genoemde straalstromen en radiobruggen. Gelukkig is dit sinds de jaren 2000 veranderd dankzij technologische vooruitgang, een groeiende interesse en de beschikbaarheid van financiële middelen, waardoor verschillende interferometers zijn ontwikkeld die waarnemingen doen bij lage radiofrequenties. Een van de belangrijkste instrumenten die waarnemingen onder de 250 MHz verricht, is de Low Frequency Array (LOFAR).

LOFAR is een instrument dat is uitgerust met twee soorten antennes: LBA antennes die frequenties tussen 10 en 80 MHz detecteren en HBA antennes die frequenties tussen 110 en 240 MHz waarnemen. Een LOFAR station en de twee antennetypes zijn afgebeeld in Figuur III en Figuur IV. Geen van de LOFAR componenten, inclusief de antennes, heeft bewegende onderdelen, omdat LOFAR gebruikmaakt van technologie om digitaal zijn focus over de hemel te sturen. Dit betekent dat alle dataverwerking in computers plaatsvindt en dat de constructie van de antennes relatief goedkoop is vergeleken met andere radiotelescopen. De relatief lage constructiekosten en internationale samenwerkingen hebben bijgedragen aan de snelle uitbreiding van LOFAR met stations door heel Europa. LOFAR's kern bevindt zich in Nederland met internationale stations in Duitsland, Polen, Engeland, Ierland, Zweden, Letland, Frankrijk en Finland.[21] Binnenkort zal LOFAR ook worden uitgebreid naar Italië en Bulgarije. Door gebruik te maken van al deze stations simuleert LOFAR effectief een radiotelescoop ter grootte van het

[21] Het Finse station in Kilpisjärvi opereert onafhankelijk en is niet verbonden met het LOFAR-netwerk.

Europese continent, waardoor het niet alleen een uitzonderlijk hoge resolutie biedt, maar daarnaast ook een hoge gevoeligheid dankzij het dichte netwerk van antennes.

Krachtige computers en geavanceerde software

LOFAR kan beelden vastleggen van grote delen van de hemel, tot wel tientallen vierkante graden in één enkele observatie,[22] waardoor het in staat is om tegelijkertijd veel radiobronnen te detecteren. Het creëren van deze grote beelden brengt echter aanzienlijke computationele uitdagingen met zich mee, vooral wanneer data van alle internationale LOFAR stations gecombineerd moeten worden. De computationele complexiteit komt door het feit dat verschillende verstorende effecten op de data moeten worden gecorrigeerd, voordat er een beeld kan worden gemaakt. Op lage radiofrequenties is één van de meest vervelende stoorzenders de ionosfeer – een laag in de atmosfeer van de aarde gevuld met geladen deeltjes die radiogolven verstoren. Het corrigeren van deze verstoringen vereist een zorgvuldig ontworpen strategie, en zodra de kalibratie voltooid is, moeten de gegevens worden omgezet in beelden – een proces dat 'imaging' wordt genoemd – om data klaar te maken voor wetenschappelijke analyse. Al deze stappen vereisen gespecialiseerde software.

In de afgelopen 10 jaar heeft de meeste softwareontwikkeling zich gericht op het verwerken van data van alleen de Nederlandse LOFAR stations, omdat dit computationeel minder 'duur' is, terwijl het nog steeds hoogwaardige kwaliteit data levert en zo helpt om nieuwe inzichten over ons Universum te verkrijgen. De volgende grote stap is om deze capaciteit uit te breiden naar het volledig automatisch verwerken van data van alle LOFAR stations tegelijkertijd. Dit zal ons uiteindelijk in staat stellen om een volledige scan van de noordelijke hemel uit te voeren met de hoogst mogelijke resolutie (zie Figuur V) en om specifieke regio's langdurig waar te nemen door meerdere observaties van hetzelfde gebied te combineren, zonder veel handmatig werk te hoeven verrichten om scherpe beelden uit de data te krijgen. Elk van deze observaties produceert ongeveer 16 terabyte (TB) aan data.[23] Het verwerken van zulke grote hoeveelheden data kan niet worden uitgevoerd op standaard moderne laptops of desktops. In plaats daarvan is het gebruik van grote, krachtige computernetwerken nodig, die ontworpen zijn om de enorme computationele belasting efficiënt aan te kunnen (zie Figuur VI).

[22] Ter vergelijking: de diameter van de maan beslaat ongeveer 0,5 vierkante graden.

[23] 16 TB is voldoende opslagruimte voor ongeveer 4 miljoen foto's van een moderne smartphone, 8.000 uur aan HD-video, of meer dan 100 miljoen pagina's aan tekstdocumenten. Een momenteel standaard laptop heeft doorgaans ongeveer 1 tot 2 TB aan opslagruimte.

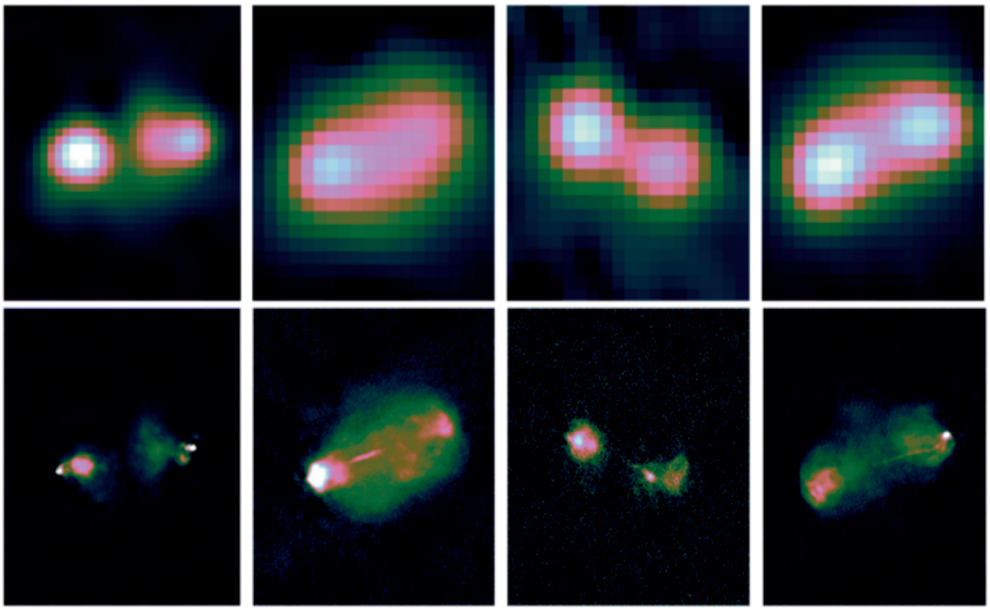


Figure V: Resolutieverbetering na het overgaan van alleen Nederlandse LOFAR-stations (eerste rij) naar de volledige internationale LOFAR (tweede rij). Elke kolom toont dezelfde radiobron. Bron: de Jong et al. (2024).

Dit proefschrift

We beginnen dit proefschrift in **Hoofdstuk 2**, waarin we zes observaties van de Nederlandse LOFAR stations combineren om het diepste beeld tot nu toe te creëren van een radiobrug tussen een paar sterrenstelselsclusters. Deze clusters bevinden zich in het proces van samensmelting tot één sterrenstelselscluster. In hun huidige fase vertonen ze radiostraling die een brug tussen hen vormt. Dit unieke fenomeen is pas onlangs ontdekt. Ons werk richt zich op verder onderzoek om de oorsprong ervan beter te begrijpen met dit nieuwe diepe beeld. Onze bevindingen wijzen op een turbulent proces, waarbij deeltjes energie winnen door interacties met turbulentie in de aanwezigheid van magnetische velden, waardoor ze straling uitzenden bij lage radiofrequenties.

Daarna richten we ons op kleinere schalen door afbeeldingen van kernheldere straalstromen (FRI) en randheldere straalstromen (FRII) van sterrenstelsels, gedetecteerd door de Nederlandse LOFAR stations, gecombineerd met data over hun kosmische afstanden, gemeten door andere telescopen. Dit stelt ons in staat om in **Hoofdstuk 3** te onderzoeken hoe deze twee typen straalstromen zich over kosmische tijd hebben ontwikkeld. We ontdekken dat het vroege Universum dichter bevolkt

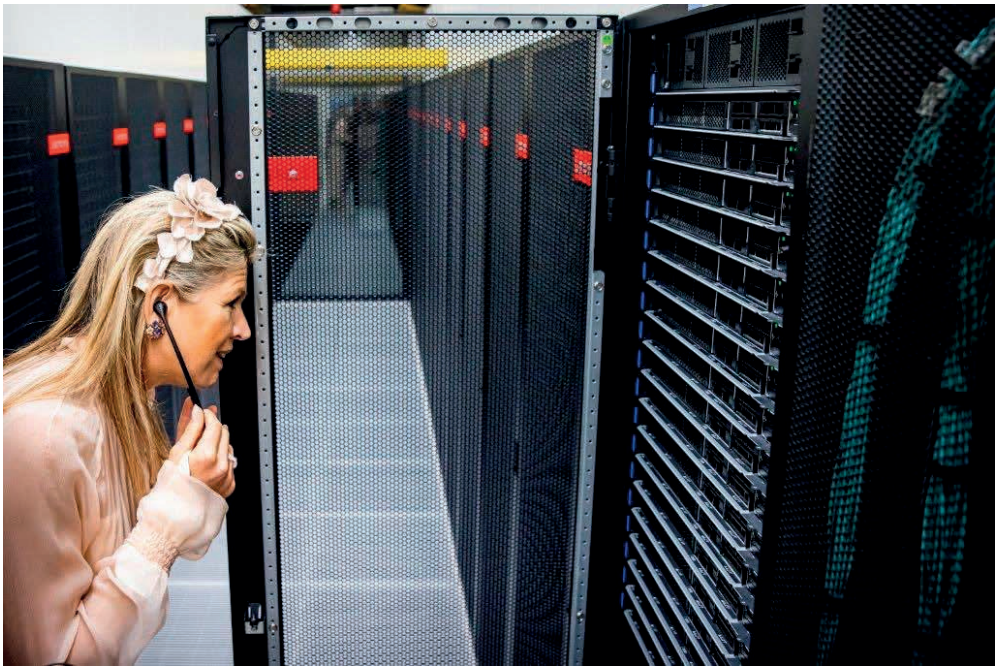


Figure VI: Koningin Máxima van Nederland tijdens de inauguratie van de Snellius-supercomputer in Amsterdam. Deze supercomputer maakt deel uit van de SURF-computinginfrastructuur. Met behulp van de GPU's van dit systeem hebben we in dit proefschrift een neuraal netwerk getraind om te detecteren wanneer de kalibratie van onze radiodata is geconvergeerd. Bron: Het Koninklijk Huis.

was met de helderste bronnen van beide type straalstromen. Dit komt waarschijnlijk door de grotere beschikbaarheid van gas in het vroegere kleinere Universum. We observeren echter onverwacht ook een afname in de dichtheid van zwakkere bronnen. Dit is een resultaat dat niet eenvoudig te verklaren is. We vermoeden dat dit komt door resolutiebeperkingen van onze data, wat de noodzaak benadrukt voor het maken van beelden met hogere resolutie van onze hemel.

Gedreven door de behoefte aan beelden met een hogere resolutie richten we ons in **Hoofdstuk 4** op het verwerken van gegevens van vier LOFAR waarnemingen van een enkel hemelgebied, genomen met alle Nederlandse en internationale stations. Het combineren van meerdere waarnemingen maakt het mogelijk om een dieper beeld te maken. Dit is nodig om meer verre en zwakke radiobronnen te detecteren.[24] Dit brengt echter ook uitdagingen met zich mee, aangezien we zo ook vier keer meer gegevens moeten verwerken, vergeleken met het maken van een beeld van

[24] Vergelijkbaar met het instellen van een langere belichtingstijd op je camera om voldoende licht op te vangen.

een enkele observatie. Om de dataverwerkingsstrategie te verbeteren, ontwikkelen we technieken om automatisch gebieden van de hemel met voldoende signaal te selecteren om zo onze dataverwerking te optimaliseren. Deze automatische selectie is cruciaal, aangezien sterkere signalen meer informatie bieden over hoe de hemel verandert over de tijd en frequentie, waardoor we verstorende effecten op de data beter kunnen corrigeren met gespecialiseerde software. Na het voltooien van de volledige gegevensverwerking verkrijgen we het diepste beeld over een hemelgebied van 2,5 bij 2,5 graden met een 20-voudige verbetering in resolutie in vergelijking met beelden gemaakt met alleen de Nederlandse LOFAR stations (zie Figuur V). Dit betekent dat de pixels in onze beelden een 400 keer kleiner gebied bestrijken dan die in de data gebruikt in Hoofdstukken 2 en 3.[25]

Hoewel we in het vorige hoofdstuk het diepste hoog-resolutie beeld met LOFAR hebben geproduceerd, zien we dat de computationele kosten te hoog zijn om deze aanpak effectief (en verantwoord, gezien het grote energieverbruik) op te schalen. Om dit probleem aan te pakken, herzien we in **Hoofdstuk 5** een 'vergeten' techniek om de datagrootte te verminderen bij het combineren van meerdere verschillende 8-uur durende waarnemingen van hetzelfde hemelgebied. Deze methode maakt gebruik van de siderische dag – de periode van ongeveer 23 uur en 56 minuten waarin de Aarde één volledige rotatie om haar as voltooit heeft. Dit stelt ons in staat om datapunten van onze waarnemingen, die elkaar herhalen op ongeveer hetzelfde siderische tijdstip, met elkaar te middelen en zo zonder veel verlies aan informatie, het datavolume te verkleinen. Onze methode maakt hierdoor een snellere verwerking mogelijk bij het omzetten van de gekalibreerde gegevens naar een definitief hemelbeeld.

Ten slotte presenteren we in **Hoofdstuk 6** aanvullende verbeteringen om de verwerking van LOFAR's enorme datavolumes te optimaliseren. Dit omvat kwaliteitsverbeteringen, het implementeren van een raamwerk voor automatische verwerking op high-performance computer systemen en het gebruik van kunstmatige intelligentie (AI) om automatisch te bepalen wanneer de kalibratie voldoende verfijnd is, waardoor menselijke tussenkomst overbodig wordt. Bovendien optimaliseren we de meest computationeel intensieve aspecten van onze strategie, waardoor het hele proces tot ongeveer vier keer sneller wordt gemaakt. Deze verbeteringen zullen ons uiteindelijk in staat stellen om veel grotere en veel meer datasets efficiënter en kosteneffectiever te verwerken.

[25] Dit betekent ook dat het uiteindelijke beeld, met een grootte van 32 GB, te groot is om te openen op een standaard laptop.

List of publications

First author

1. *Benford's Law in the Gaia Universe*
   J.M.G.H.J. de Jong, J.H.J. de Bruijne, J. De Ridder
   Astronomy & Astrophysics, Volume 642, A205, 16 pp., 2020

2. *Deep Study of A399-401: Application of Wide-Field Facet Calibration*
   J.M.G.H.J. de Jong, R.J. van Weeren, A. Botteon, J.B.R. Oonk, G. Brunetti, T.W. Shimwell, R. Cassano, H.J.A. Röttgering, C. Tasse
   Astronomy & Astrophysics, Volume 668, A107, 19 pp., 2022

3. *Cosmic Evolution of FRI and FRII Sources out to $z = 2.5$*
   J.M.G.H.J. de Jong, H.J.A. Röttgering, R. Kondapally, B. Mingo, R.J. van Weeren, P.N. Best, L.K. Morabito, M. Magliocchetti, J.B.R. Oonk, A. Villarrubia-Aguilar, F.F. Vecchi
   Astronomy & Astrophysics, Volume 683, A23, 17 pp., 2024

4. *Into the Depths: Unveiling ELAIS-N1 with LOFAR's Deepest Sub-Arcsecond Wide-Field Images*
   J.M.G.H.J. de Jong, R.J. van Weeren, F. Sweijen, J.B.R. Oonk, T.W. Shimwell, A.R. Offringa, L.K. Morabito, H.J.A. Röttgering, R. Kondapally, E.L. Escott, P.N. Best, M. Bondi, H. Ye, J.W. Petley
   Astronomy & Astrophysics, Volume 689, A80, 32 pp., 2024

5. *Unlocking Ultra-Deep Wide-Field Imaging with Sidereal Visibility Averaging*
   J.M.G.H.J. de Jong, R.J. van Weeren, T.J. Dijkema, J.B.R. Oonk, H.J.A. Röttgering, F. Sweijen
   Astronomy & Astrophysics, Volume 694, A98, 9 pp., 2025

6. *Advanced Strategy for Deep Sub-Arcsecond Wide-Field Facet Calibration with LOFAR*
   J.M.G.H.J. de Jong, L. Veefkind, R.J. van Weeren, R.J. Schlimbach, J.B.R. Oonk, D.N.G. Kampert, H.J.A. Röttgering, M. van der Wild
   To be submitted.

Contributing author

1. *Abell 0399–Abell 0401 radio bridge spectral index: First multi-frequency detection*
G.V. Pignataro, A. Bonafede, G. Bernardi, F. de Gasperin, G. Brunetti, T. Pasini, F. Vazza, N. Biava, **J.M.G.H.J. de Jong** , R. Cassano, A. Botteon, M. Brüggen, H.J.A. Röttgering, R.J. van Weeren, and T.W. Shimwell
Astronomy & Astrophysics, Volume 685, L10, 7 pp., 2024

2. *Characterization of the decametre sky at sub-arcminute resolution*
C. Groeneveld, R.J. van Weeren, E. Osinga, W.L. Williams, J.R. Callingham, F. de Gasperin, A. Botteon T. Shimwell, F. Sweijen, **J.M.G.H.J. de Jong**, L.F. Jansen, G.K. Miley, G. Brunetti, M. Brüggen, H.J.A. Röttgering
Nature Astronomy, Volume 8, p. 786-795, 2024

3. *1-arcsecond imaging of ELAIS-N1 field at 144MHz using the LoTSS survey with international LOFAR telescope*
H. Ye, F. Sweijen, R.J. van Weeren, W.L. Williams, **J.M.G.H.J. de Jong**, L.K. Morabito, H.J.A. Röttgering, T.W. Shimwell, P.N. Best, M. Bondi, M. Brüggen, F. de Gasperin, C. Tasse
Astronomy & Astrophysics, Volume 691, A347, 18 pp., 2024

4. *A hidden Active Galactic Nuclei population: the first radio luminosity functions by physical process*
L.K. Morabito, R. Kondapally, P.N. Best, B.-H. Yue, **J.M.G.H.J. de Jong**, F. Sweijen, M. Bondi, D.J. Schwarz, D.J.B. Smith, R.J. van Weeren, H.J.A. Röttgering, T.W. Shimwell, I. Prandoni
Monthly Notices of the Royal Astronomical Society: Letters, Volume 536, Issue 1, Pages L32–L37, 2025

5. *The LOFAR Two-metre Sky Survey: Deep Fields Data Release 2*
T.W. Shimwell, C.L. Hale, P.N. Best, A. Botteon, A. Drabent, M.J. Hardcastle, V. Jelić, **J.M.G.H.J. de Jong**, R. Kondapally, H.J.A. Röttgering, C. Tasse, R.J. van Weeren, W.L. Williams, A. Bonafede, M. Bondi, M. Brüggen, G. Brunetti, J.R. Callingham, K.J. Duncan, C. Horellou, S. Iyer, I. de Ruiter, K. Małek, D.G. Nair, L.K. Morabito, I. Prandoni, A. Rowlinson, J. Sabater, A. Shulevski1, D.J.B. Smith
Astronomy & Astrophysics, Volume 695, A80, 20 pp., 2025

6. *A decade of sub-arcsecond imaging with the International LOFAR Telescope* L.K. Morabito, N. Jackson, **J.M.G.H.J. de Jong**, E.L. Escott, C. Groeneveld, V.H. Mahatma, J.W. Petley, F. Sweijen, R. Timmerman, R.J. van Weeren
Springer Nature (Astrophysics and Space Science), Volume 370, article number 19, 2025

7. *The relationship between morphology and host galaxy properties of RLAGN in LoTSS DR2*
L. Clews, J.H. Croston, H. Dickinson, B. Mingo, B. Barkus, M.J. Hardcastle, **J.M.G.H.J. de Jong**
Submitted for publication in Monthly Notices of the Royal Astronomical Society.

Published software

1. **LOFAR helpers**
Python package with several tools to work with LOFAR data. This code has been developed and used for most of the chapters in this thesis.
`https://github.com/jurjen93/lofar_helpers`

2. **Advanced Astro Visualization**
Python code to produce videos and posters of fits images. It was used during a press release in 2021 to demonstrate LOFAR high-resolution imaging.
`https://github.com/jurjen93/advanced_astro_visualization`

3. **Redshifting**
Python package to move radio sources to higher redshifts in the image plane. This software has been used to simulate the appearence of FRI and FRII sources at different redshifts.
`https://github.com/jurjen93/redshifting`

4. **Sidereal Visibility Averaging**
Python package to average visibilities for similar baseline coordinates. This software is currently being used to enhance imaging when combining multiple LOFAR observations for ultra-deep imaging.
`https://github.com/jurjen93/sidereal_visibility_avg`

Curriculum Vitae

I was born on May 26, 1993, in Hulst, a city in the Dutch province Zeeland. While I had great teachers at primary school at Basisschool de Ark, my main teacher growing up was my older brother, who, being 7.5 years older, constantly inspired me with stories about history, science, and everything he learned at school himself. Navigating secondary school at Het Reynaertcollege in Hulst was challenging, largely due to the fear to speak up because of my stutter. Nevertheless, I found great joy in following mathematics classes, which was further reinforced by participating in the national Mathematics Olympiads. Eventually, I finished secondary school in 2011, but was unsure about what to do next.

Since engineering seemed to offer diverse career opportunities, I chose to pursue a bachelor's degree in mechanical engineering at Avans Hogeschool in Breda. However, during my first year, I quickly realised I had 'two left hands' – as we often say in Dutch – since I performed worse in my class in practical assignments on topics like drilling, milling, and welding. More theory-focused subjects like mathematics, mechanics, and material science appealed to me far more, which led me to decide to complete the propaedeutics and leave mechanical engineering to pursue a university degree instead.

In 2012, I started a bachelor's degree program in mathematics at the University of Utrecht, where I enjoyed in my first year courses on fundamental mathematics and added additional courses from other programs to feed my interest in philosophy, history, and physics. By the second year, I decided to pursue a full bachelor's degree in physics next to mathematics, alongside a minor in the history and philosophy of science. I completed these studies with a double bachelor's thesis under the supervision of dr. Steven Wepster. This work was focused on interpreting the ancient Indian mathematical-astronomical work "Ganita-Yukti-Bhāṣā", which had recently been translated from Malayalam to English. In my research, I sought to connect this ancient work to modern mathematics and astronomy. This research project marked my first direct academic interaction with astronomy.

After finishing my bachelor's degrees in math and physics within three years, I moved to the University of Ghent in Belgium in 2015 to pursue a master's degree in mathematics with a specialisation in mathematical (astro)physics. This was where I took a few introductory courses in astrophysics. During my second year, I went on an Erasmus+ exchange to Uppsala (Sweden) where I took a few more courses in astronomy and remotely worked under the supervision of prof. dr. Maarten Baes

on my master's degree thesis, which involved deriving analytical Fox-H functions for the Nuker model in galactic dynamics.

After a year in Sweden, I was ready for a new adventure and spent the summer of 2017 in Tunis (Tunisia), where I worked as a summer intern at Engineering Procurement and Project Management (EPPM), as part of the Belgian IAESTE program. There, I derived formulas to calculate stresses on storage tanks and pipeline constructions. After returning, I graduated with my first master's degree and began an advanced master's program in Space Studies at the University of Leuven in Belgium. This one-year program provided a comprehensive overview of the space industry and space-related research, covering topics such as space law, space biology, exoplanet research, rocket science, and general engineering. I had the privilege of learning from experts, including former astronaut Frank De Winne, who inspired me by showing that even though you have physically been in space, you can still be down to Earth. To apply the knowledge I gained across various subjects, I also took on a side job during this time, writing popular science articles for Scientias. The articles I wrote covered a wide range of topics, including mathematics, physics, astronomy, quantum computing, and space exploration.[26]

In the summer of 2018, I was fortunate to be awarded a small scholarship to study Mandarin in Tainan (Taiwan). This was an enjoyable and challenging side step out of my comfort zone, since I learned much about new cultures, while it showed me that I was not gifted with skills for learning foreign languages. A few days after returning to Europe, I began, with financial support from a scholarship from the Flemish government (foreign affairs), a three-month internship at the European Space Agency (ESA). At ESA, I did research with the brand new second data release from the Gaia space telescope under the supervision of dr. Jos de Bruijne. I applied Benford's law to identify patterns in the data corresponding to distances between stars, which led to my master's thesis and the completion of the Space studies program. This work was later converted into my first first-authored scientific publication.

Having completed 2 bachelor's degrees and 2 master's degrees, I decided that I was done with collecting university degrees for now and started with a traineeship in data science at Matrixian Group. This transitioned into a full-time role as a data scientist and machine learning engineer. During this time, I gained experience beyond science and programming as well, such as working with clients, thinking about business models, meeting strict deadlines and dependencies, and collaborating with colleagues from diverse backgrounds. When the COVID-19 pandemic hit at the start of 2020, I took in March of that year the last plane from the Netherlands to Iceland, where my girlfriend was living at the time. I ended up staying there for

[26] You can still find my articles here `https://scientias.nl/author/jurjen-de-jong`.

four months, which gave me time to reflect on my future. While I enjoyed my work as a data scientist, I missed the academic challenges...

Thankfully, I remained in contact with Anniek Gloudemans, a former colleague from my internship at ESA, who introduced me to her supervisor, prof. dr. Huub Röttgering. After a few interviews, he offered me a PhD position in Leiden, and although I had no prior experience with radio astronomy, the opportunity to work in the field of radio astronomy with a focus on the computational data processing side as well was appealing. On January 1st, 2021, under the supervision of prof. dr. Huub Röttgering, dr. Reinout van Weeren, and dr. Raymond Oonk, I embarked on my PhD journey. Despite the lockdowns keeping me home for much of the first year, I focused on mastering the calibration and imaging of LOFAR data, while studying galaxy cluster pre-mergers and radio galaxy evolution. Later on, my focus moved to pushing the wide-field sub-arcsecond imaging pipeline with LOFAR forward to open up new windows for the study of the universe at the smallest angular scales and greatest depths. All this work eventually resulted in the work presented in this thesis.

While my career has let me zigzag across the world and disciplines, I feel very fortunate to have the opportunity to continue working for the coming years as a postdoc in Leiden with Huub, Reinout, Raymond, and many others who have made the work outlined in this thesis possible.

Although stuttering had various effects on my life, I am proud to mention that next to my scientific career, I co-founded with friends the non-profit organisation 'Stamily', during the first years of my PhD. This association fosters a global community and provides a supportive safe space for people who stutter, as having a stutter can impact many aspects of people's lives. I was elected as the first chair and took on a leadership role in establishing the foundation, organizing both in-person and online events, making contact with other associations, and maintaining the website. On the day I write these words, Stamily has over 150 members from 35 different countries, has been featured in national media in the Netherlands and local media in other countries,[27] and is organizing each year several camps and weekends for people who stutter around Europe.

[27] `https://www.stamily.org/in-the-media`

Acknowledgements

The work presented in this thesis would not have been possible without the support of many people over the past four years. First and foremost, I want to thank Huub, Reinout, and Raymond. They took on the challenge of supervising a PhD student with little to no prior knowledge of radio astronomy. I am grateful for the trust you placed in me to complete this journey successfully. I am thankful to Reinout for his endless patience, support, and daily mentorship throughout these four years; to Raymond, who, despite his busy schedule, always made time to help me when I reached out with computing issues or when I just asked for advice; and to Huub for encouraging me to think critically and stay focused on the bigger picture of my research as well. I really appreciate all the help and support you three provided throughout my PhD, as well as the effort you have put into guiding me toward my future career. I am very happy that we will continue working together.

I could not have started this PhD without Anniek, who connected me with Huub. Thank you so much for this. A special thanks to my former office mate and patient colleague, Frits. Thank you for opening the heavens for high-resolution wide-field imaging with LOFAR and for always being ready to help out. I also want to thank my dear colleagues and travel companions from the first few years of my PhD: Christian, Erik, María, and Roland. I really enjoyed having the opportunity to visit my first astronomical conferences with you and explore new parts of the world. I will never forget (or perhaps partly) our dance night out in Seoul and our hunt for spotting the big five at Pilanesberg.

I would like to thank everyone from Leiden's LOFAR group with who I shared the excitement of working with data from this amazing instrument: Andrea, Bohan, Emanuele, Gabriella, George, Haoyang, Ian, James, Jinyi, Joe, Joshiwa, Lars, Luca, Marco, Martijn, Rafael, Sai, Sarah, Scott, Tong, Wendy, Xuechen, and others. Many thanks to everyone from the LOFAR VLBI working group as well: Alexander, Aleksandra, Cyril, Deepika, Emmy, Etienne, Jonny, Jort, Gulay, Henrik, Leah, Nadia, Neal, Marcin, Marco, Matthijs, Rishi, Sagar, Shane, Tom, Vijay, and many others. It was a real joy to spend weeks with you hacking LOFAR software and doing amazing things at my favourite LOFAR resolution. Especially, the week in a castle in Durham and the catacombs in Paris are experiences I will never forget. Of course, I have to thank the guys from Surf – Duncan, Lars, and Robert Jan – who have not only improved the work presented in this thesis but also enriched my coding and machine learning knowledge.

A big thanks to Michiel, Osmar, and Tim for making my experience as a teaching assistant in the radio astronomy course such a rewarding experience. Developing the Cantenna interferometer with you each year was a great learning experience and lots of fun. Thanks to Jelle and Thomas for sharing Office 460 with me. I hope our extended LOFAR brainstorming sessions were not too distracting. Rico, Tong, and Willeke, I enjoy our time together in our new office as well. Good (or perhaps bad) news: I will stay there with you. To all my Cosmos United teammates – Alfred, Amy, Andrew, Christiaan, Dario, Elia, Joey, Josh, Kayou, Kevin, Mojtaba, Murat, Osmar, and everyone else who came out (un)injured – thank you for being part of the same team during the eventful matches in the beautiful indoor football stadium from the USC. If our scores had been adjusted for age, we would have won the league after all. I also want to thank the IT group and secretaries for always being ready to help with technical, practical, or organizational issues. It is difficult to list all the names from everyone I have met over the years at Leiden Observatory and other institutes who have supported me throughout my PhD journey. But I am very thankful for all your time, help, and support, or just for spending time together during borrels, coffee breaks, and other non-scientific activities.

Mam, Pap, dank jullie wel voor de onvoorwaardelijke steun. Jullie hebben altijd alles gedaan om mij de vrijheid te geven mijn passies te volgen. Eugénie, Veerle, Justine en Jornt, bedankt voor al jullie steun en liefde. Ik ben heel dankbaar voor onze hechte broerzussenband. I also want to say ‘kiitos’ to Harri, Marja, and Henrik for providing me with a second home in Finland. I really enjoyed spending time with you every summer and winter.

Finally, and most importantly, I am so grateful for the love and support from Satu. I am so proud of you and you inspire me with your ability to never give up. Without you, I would have no idea where I would have been. At least not in Leiden, writing the last few words of my PhD thesis.

