

LEIDEN UNIVERSITY

MASTER THESIS

Systematics in LoTSS Radio Source Orientations

Author:

Erik OSINGA

Supervisor:

Prof. dr. G.K. MILEY

Dr. T.W. Shimwell

Dr. R.J. Weeren, van

*A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science*

in the

LOFAR research group
Sterrewacht Leiden

August 1, 2018

"Consistency is a virtue until it gets annoying."

Ursula K. Le Guin

LEIDEN UNIVERSITY

*Abstract*Faculty Name
Sterrewacht Leiden

Master of Science

Systematics in LoTSS Radio Source Orientations

by Erik OSINGA

This thesis tests the hypothesis that the position angles of radio galaxies are randomly oriented in the HETDEX Spring Field region. A sample of 6795 double-lobed radio galaxies is extracted from the sample of 325,694 low frequency radio sources in the first data release of the LOFAR Two-metre Sky Survey. This sample is constructed with two different automated approaches for the two separate subsets of sources: connected double lobes and isolated double lobes. We performed two statistical tests on the sample, one in two dimensions and one in three dimensions with the 3416 sources that have a redshift measurement available. After taking into consideration possible uncertainties in the position angles, we find that the subset of connected lobes with a peak flux above a threshold of 9.41 mJy/beam show a significant ($p\text{-value} < 0.001$) deviation from uniformity at angular scales between 4 and 5 degrees in the 2D analysis. While we find multiple indications that this result might be caused by biases in the dataset, the result cannot be ruled out with certainty. We find that the angular scale on which the alignment is most pronounced translates to a comoving scale of $150h^{-1}\text{Mpc}$. Supposing that the effect is not caused by systematic effects, the implied alignment of the active galactic nuclei that give rise to these radio jets suggests large scale coherence in angular momentum. If the effects seen are real, they reinforce prior evidence for the alignment of active galactic nuclei with their surrounding large scale structure inferred from polarization alignment studies.

Acknowledgements

Primarily, I want to thank my supervisors, George, Reinout and Tim, for the amazing year of radio astronomy that I was introduced to by you. George, for always sparking my excitement about the project and seeming genuinely delighted every time I came into your office. Reinout, for always being very easy to meet with, in particular when I came by many times during the crisis a few weeks before the deadline, when I thought my statistics were crumbling down. I don't know what I would have done without you. And Tim, for always being very involved with the project and interested in the progress, for example by always asking to Skype in if ever you could not attend a meeting and taking the time out of your holidays to read first drafts of this work. To all three, we have had some very lively discussions during our meetings and I often left out meetings feeling very inspired. Secretly, it is a pity that this project is now finished.

I also want to thank Martijn Oei for helping me get started on this project through the last two chapters in your radio astronomy thesis, and personal meetings. Even though we only spoke shortly, it was clear that you are a very gifted student, and your work has been very helpful throughout this project.

I also want to express my deep gratitude to Omar Contigiani, for helping me understand the contents of his thesis and helping me visualize how the heck parallel transport works. I am also very grateful that you were so kind as to provide me with your processed catalogue, so I could verify whether my implementation worked.

I would also like to thank Huub Röttgering and the entire LOFAR group. The research group has made me feel very included even though half the group meetings I had no idea what was going on. I really appreciated how much attention was given to the 'youngsters' by including us and inviting us to everything. I learned a lot in particular from the two day science project meeting, and I cannot express how amazing it was that I got a chance to observe at the IRAM telescope in the Sierra Nevada.

Finally, I would like to thank my office mates, for making it a competition to arrive at the office early. Even though I rarely won, you guys motivated me all year to stay competitive.

This research has also benefited from many software tools. The most frequently used are SAOImage DS9, TOPCAT, the Python packages SciPy, Astropy, NumPy, Matplotlib, and the text editor Sublime Text.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Large scale structure in the Universe	1
1.2 Radio galaxies	2
1.2.1 Overview	2
1.2.2 Taxonomy	3
1.3 Large scale AGN alignments	4
1.3.1 Quasars and optical polarization alignment	4
1.3.2 Radio polarization alignment	5
1.3.3 Radio structure alignment	6
2 The project	9
2.1 Introduction	9
2.1.1 The LOFAR Two-metre Sky Survey	10
2.1.2 Position angle	10
2.2 Selection of sources	12
2.2.1 Isolated radio lobes	13
Statistical motivation	14
Physical motivation	16
Decision tree	17
2.2.2 Connected radio lobes	20
2.2.3 Producing the final sample	21
2.2.4 Defining sample subsets	22
2.3 Statistics	24
2.3.1 Parallel transport	24
2.3.2 Statistical test	26
3 Results	29
3.1 Global alignment: 2D	29
3.2 Local alignment: 2D	32
3.2.1 Full sample	32
3.2.2 Flux and size bins	33
3.2.3 Effect of subsets	34
3.3 Global alignment: 3D	36
3.4 Local alignment: 3D	37
3.4.1 Full samples	37
3.4.2 Binned samples	37

4	Discussion	41
4.1	Fidelity of the results	41
4.1.1	Fitted position angles	41
4.1.2	Error propagation	42
4.2	Interpretation of the results	44
4.2.1	Global alignment	44
4.2.2	Local alignment	46
4.2.3	Comparison with previous work	49
4.2.4	Possible explanations	50
4.3	Suggestion for future work	51
5	Conclusion	55
	Bibliography	57

Chapter 1

Introduction

This chapter will discuss the relevant background of active galactic nuclei (AGN) orientation studies. It will start by discussing the large scale structure of the Universe, which may effect the orientation of galaxies. Next, it will discuss the different mechanisms that power radio sources and the classification schemes that are commonly used to classify these sources. Finally, this chapter will give a comprehensive overview of what has been discovered in the field of AGN orientations, starting with alignment of optical polarizations and progressing to radio polarization and radio structure alignments.

1.1 Large scale structure in the Universe

On the biggest scales the distribution of matter in the Universe is known to follow large scale filaments that mainly consist of dark matter (e.g., Primack, 1997). To a much smaller degree, baryonic matter, including gravitationally bound galaxies, fill the mass budget of these filaments. A large fraction (30-40%) of these baryons is found in the warm-hot intergalactic medium (Cen and Ostriker, 1999; Davé et al., 2001; Fang, Bryan, and Canizares, 2002), thus, the gravitationally bound galaxies represent only a tiny amount of the total mass in the large scale filaments (see Coil, 2013 for a recent review). These filaments fill the Universe like a Cosmic Web, interconnecting and forming a network that alternates between low density regions largely devoid of galaxies and dense thread-like structures of scales up to a few hundred megaparsec (Libeskind et al., 2018). The existence of large scale structures has been verified thoroughly, with these structures being detected in the Lick galaxy catalogue, which contained roughly a million galaxies compiled entirely by eye (Seldner et al., 1977), as well as in the present day largest redshift surveys of galaxies such as the Sloan Digital Sky Survey (SDSS) (Zehavi et al., 2002) and the Two-Degree Field Galaxy Redshift Survey (2dFGRS) (Baugh et al., 2004).

The established theoretical explanation of the existence of this large scale structure is that it originates from the small random Gaussian density fluctuations in the smooth early Universe. These initial fluctuations are observed from the anisotropies in the Cosmic Microwave Background (CMB) (e.g., Smoot et al., 1992), which consists of the photons emitted at $z \simeq 1000$ when protons combined with electrons and these photons last interacted directly with matter. Evolving due to gravity, the density fluctuations cause overdense regions to grow in mass while shrinking in size while empty regions expand to fill the bulk of the volume of the Universe (Bond, Kofman, and Pogosyan, 1996; Coil, 2013). This paradigm has been tested both indirectly by numerical N-body simulations (Benson et al., 2001) and directly by employing the two-point galaxy correlation function (Peacock et al., 2001). The two-point galaxy correlation function measures the excess probability over random of finding a pair of galaxies as a function of transverse separation and line-of-sight separation.

In an isotropic Universe this is independent of direction, but in redshift space the peculiar (proper) velocities of galaxies cause an anisotropy in the correlation function. If gravitational collapse is responsible for cosmological structure formation, this anisotropy will cause an apparent contraction of structure on large scales along the line of sight in redshift space. This is due to adjacent galaxies moving in the same direction causing coherent peculiar velocities which distort the apparent redshift (Coil, 2013). These correlation functions have been calculated extensively from various galaxy surveys (e.g., Baugh et al., 2004; Ross, Brunner, and Myers, 2007; Gaztañaga et al., 2005).

The final details of the cosmic web are a product of cosmological parameters as well as the physics of galaxy formation and evolution. Conversely, the cosmic web too influences the formation of galaxies and galaxy clusters. One example of this influence is the origin of the angular momentum of galaxies (Peebles, 1969). The tidal forces that are driving the collapsing proto-galaxies are also driving the anisotropic infall of matter (Libeskind et al., 2018). Therefore, one might expect that galaxy orientations and large-scale filaments are oriented along the same direction. This has been the subject of much, sometimes conflicting, research both in N-body simulations (e.g., Lee and Pen, 2000; Aragón-Calvo et al., 2007; Codis et al., 2012; Libeskind et al., 2013; Trowland, Lewis, and Bland-Hawthorn, 2013; Aragon-Calvo and Yang, 2014; González et al., 2017; González et al., 2017; Ganeshaiah Veena et al., 2018) and observations (e.g., Jones, van de Weygaert, and Aragón-Calvo, 2010; Tempel and Libeskind, 2013; Tempel, Stoica, and Saar, 2013; Pahwa et al., 2016; Hirv et al., 2017).

1.2 Radio galaxies

This section is divided into two different topics related to radio galaxies. First, the physical description of radio galaxies will be discussed, then we discuss often-used classification schemes.

1.2.1 Overview

According to the now universally accepted standard model of radio galaxies and quasars, the emission of these objects is powered by a central accreting supermassive black hole (Padovani et al., 2017 for a recent review). These active galactic nuclei (AGN) have very high luminosities (up to $L_{bol} = 10^{48} \text{ erg s}^{-1}$) and significant emission throughout the entire electromagnetic spectrum (Padovani et al., 2017). The dominant radio emission process for these AGN is synchrotron emission (i.e., relativistic electrons accelerating in magnetic fields). Two classes of radio (loud) galaxies were differentiated by Fanaroff and Riley (1974) using the ratio R_{FR} of the separation of the high surface brightness regions, also called hotspots, and the extent of the source up to the lowest brightness contour. Sources with $R_{FR} < 0.5$ were classified as FR I and sources with $R_{FR} > 0.5$ were classified as FR II. FR I sources are variously referred to as core-brightened or edge-darkened sources while FR II sources are commonly referred to as classic doubles, edge-brightened doubles or symmetric doubles (Miley, 1980). Unification schemes exist that attempt to consolidate different classes of AGN, such as radio galaxies, quasars, Seyfert galaxies and blazars, as one mechanism viewed under different angles (Urry and Padovani, 1995; Padovani, 1997). Figure 1.1 shows this unification model, showing that the orientation of the AGN with respect to the line of sight determines the type of object we observe.

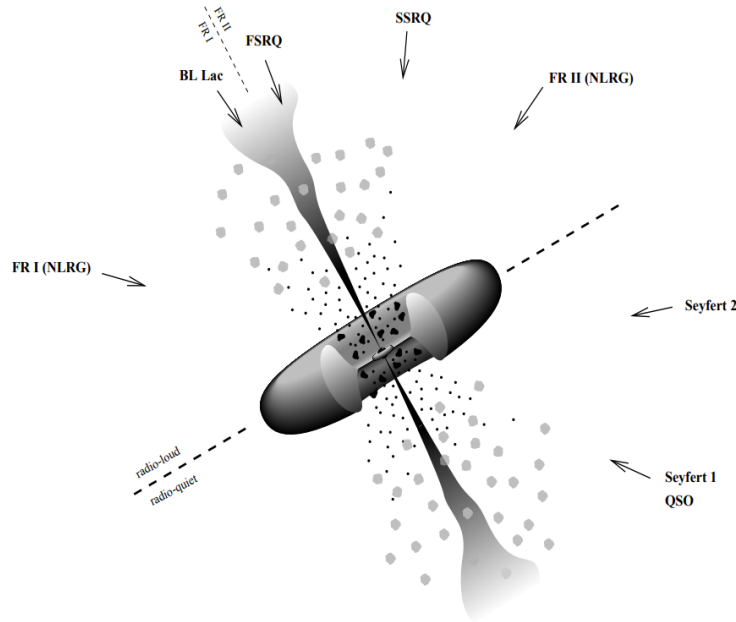


FIGURE 1.1: The unification model for AGNs. At the center a supermassive black hole is accreting galactic matter. Two jets of relativistic particles are formed along the spin axis of the nucleus, perpendicular to the plane of the accretion disk. Different objects (quasars, radio galaxies, FR I, FR II) are the result of different orientations of the observed object. Figure obtained from Torres and Anchordoqui (2004).

1.2.2 Taxonomy

The taxonomy of radio galaxies encompasses many different classification schemes since Fanaroff and Riley posed their classification in 1974. These are now discussed based on the review by Padovani et al. (2017).

Morphological Continuing from the Fanaroff-Riley morphological classification scheme, Miley (1980) classified 5 types of radio galaxies based on their appearances. Most radio sources are *classic double-lobed sources* that have bright hotspots near the edges of the lobes. These doubles show large degrees of symmetry and collinearity as the hotspots line up with the core and usually have comparable flux. *Narrow edge-darkened sources*, in contrast, have brightness distributions that are brightest near the nucleus and diminish near the edges. These sources are sometimes called double sources as well and also fall in the FR I category. *Wide double sources* are much like classic doubles but the lobes are more relaxed and thus have lower length-to-width ratios than the classic doubles. An example of such a source is NGC1316 (e.g., Cameron, 1971). *Narrow tailed sources*, also called head-tail sources, are asymmetric sources with a bright head and a diffuse tail. The flux distribution follows a tadpole like shape with the brightest component matching the nucleus of the hosting elliptical galaxy. These sources are believed to be a result of active radio galaxies moving through a dense intergalactic medium causing the remnant tail and bright head (Miley et al., 1972). *Wide-tailed sources* are an intermediate class between the narrow tailed sources and the narrow edge-darkened sources. Often called wide angle tail

sources, these usually have a curved morphology. This morphology indicates that the source interacts significantly with its environment.

Radio spectrum When multiple frequency observations are available, radio sources can also be grouped on basis of the shapes of their spectrum ($f_\nu \sim \nu^\alpha$). A division between steep ($\alpha \geq 0.5$) and flat ($\alpha < 0.5$) spectrum sources roughly corresponds to extended and compact sources (Massardi et al., 2011; Kimball and Ivezić, 2008).

Loud and quiet A long-established and often used classification of radio AGN is the division into radio loud (RL) and radio quiet (RQ) AGN (Sandage, 1965). The threshold between the two classes can either be the radio power or the ratio of the radio-to-optical flux densities (Padovani et al., 2015). It has been argued that these classes are intrinsically different objects, with RL AGN emitting most of their energy non-thermally in association with powerful radio jets and RQ AGN emitting thermally, (in)directly related to the accretion disk around the nucleus (Padovani, 2017).

High and low excitation Radio galaxies can finally also be classified on basis of their optical spectra. High- and low-excitation radio galaxies (HERGs and LERGs) are classified based on optical line fluxes and equivalent widths (e.g., Best and Heckman, 2012). HERGs have a jet fueled efficiently by cold gas in host galaxies with typically young stellar populations while LERGs are associated with inefficient flows of hot gas from the galaxy halo at low accretion rates (Best and Heckman, 2012).

Many of these classifications have some overlap. The FR I and FR II classification roughly corresponds to the LERG and HERG classification for example, but there is no one-to-one connection. Radio-loud galaxies are more often HERGs, while radio quiet galaxies are mostly LERGs (Padovani et al., 2017).

1.3 Large scale AGN alignments

In the following sections a chronological overview of what has been discovered in the field of AGN orientations will be given. Correlations between radio structure and optical polarization (e.g., Antonucci, 1983; Lister, 2001; Joshi et al., 2007), and optical polarization and radio polarization (e.g., Rusk and Seaquist, 1985; Rusk, 1988) have been found. Thus, if alignment has been found in radio structure, one might expect this alignment signal to be present in polarization as well, and vice versa. Therefore, even though this research will only consider radio structure alignment, this section will encompass the tendency for the optical polarization of quasars to be aligned, as well as the preferential orientation of radio structures and radio polarizations.

1.3.1 Quasars and optical polarization alignment

In 1998, the first evidence was found for very large scale correlations between the optical polarization angles of quasars. Hutsemekers (1998) found, in a small sample of 42 moderate to high-redshift quasars, that the quasar polarization position angles were not randomly distributed within 180 degrees. A larger sample was also constructed, that consisted of 170 quasars. No significant deviation from a uniform distribution was found in this sample, but a statistical test was developed to check

for local alignment signals, and a significant deviation from an isotropic distribution was found in subgroups of around 24 quasars. The spatial scale at which this alignment effect was observed corresponded to $\sim 1000 h^{-1}$ Mpc at redshift $z \simeq 1 - 2$.

This result was corroborated by Hutsemékers and Lamy (2001), by measuring the polarization of a new sample of quasars in the same region of the sky where alignment was previously found. The authors also found that subgroups of 20-30 quasars, in a sample of 170 polarized quasars, at redshift $z \simeq 1 - 2$ have preferentially aligned optical polarization vectors. The effect was found on the same gigaparsec scale as the previous study. Later, the result was confirmed again by Jain, Narain, and Sarala (2004) using coordinate-invariant statistics. In Bharadwaj, Bhavsar, and Sheth (2004), the authors state that the effect originates most likely due to propagation, since it is unlikely the quasars at such a large distance from one another are correlated. However, it was found that galactic extinction would not explain the redshift dependence of the signal, while supergalactic extinction required an extraordinarily large size of the supercluster.

The alignment effect was mapped in more detail in Hutsemékers et al. (2005). In a sample of 355 quasars the optical polarization angle was found to be aligned over gigaparsec regions at both $z \simeq 0.5$ and $z \simeq 1.5$. Even more peculiar, the orientations seemed to alternate along the line of sight between random and coherent on 1.5 Gpc comoving length scales. Furthermore, the mean polarization angle was found to be rotating with redshift at a rate of 30° per Gpc. This property argued against the effect being due to local systematics. An explanation has been proposed in Payez, Hutsemékers, and Cudell (2010), with the existence of an axion-like particle. The axion-photon mixing that would then occur in the intergalactic medium would affect the polarization, which could explain the large scale alignment effects (Payez, Cudell, and Hutsemékers, 2008). However, this axionlike particle explanation was ruled out by the absence of predicted circular polarization (Payez, Cudell, and Hutsemékers, 2011) leaving the observed effect to remain a mystery.

In contrast to the propagation-related explanations, the well-known correlation between quasar optical polarization and host galaxy position angles (Joshi et al., 2007; Borguet et al., 2008; Borguet et al., 2011) would imply that the quasars are aligned themselves or with respect to the large scale structure they reside in. This effect was found by Hutsemékers et al. (2014). They measured the polarization of 93 quasars that belong to large-scale quasar groups. 19 of these 93 quasars were significantly polarized, and were found to have a polarization angle either parallel or perpendicular to the large-scale structures they inhabit. Furthermore, they found that the probability of the polarization angles being randomly distributed is less than one percent. An even stronger signal is found over a scale of ~ 550 Mpc, where the probability of a uniform distribution is less than 0.1%. The existence of intrinsic alignments on such a large scale would explain the previously found quasar polarization alignments. However, it would have profound implications for the cosmological principle, since it would argue for privileged directions in the Universe on scales larger than previously found upper limits for the scale of homogeneity of $260 h^{-1}$ Mpc (Yadav, Bagla, and Khandai, 2010).

1.3.2 Radio polarization alignment

If, as suggested may be the case by the latest optical polarization studies, the galaxies themselves are aligned along large scale filaments, one would expect to see this effect in radio wavelengths as well. The alignment between optical polarization and

radio polarization is a known effect (Rusk and Seaquist, 1985; Rusk, 1988). In extension of this, the correlation between the radio structure and optical polarization is established as well (Antonucci, 1983; Lister, 2001; Joshi et al., 2007). Stockman, Angel, and Miley, 1979, for example, find out of 24 quasars that have a well-defined radio and optical polarization position angle, 20 have the two angles agreeing to within 30 degrees. For a random distribution of these angles, one would expect a third of these sources to have the angles agreeing within this range, and such the probability that this occurs by chance was found to be less than 10^{-5} . This correlation implies long term (lifetime of the radio jet, $\sim 10^6$ year) memory of direction in quasars.

As such, there have been various studies investigating alignment between radio polarizations. Joshi et al. (2007) tested radio polarization vectors of 4290 flat-spectrum sources at 8.4 GHz with flux higher than 1 mJy. No effect was found in the sample as a whole, split in half on redshift, or when looking at specific areas where alignment effects have been found before. However, Pelgrims and Hutsemékers (2015) contradicted this claim, finding significant alignments in a nearly identical sample of sources by constraining the sample on redshift. The authors claimed that considering all sources, regardless of redshift, in the specific areas where alignment was found before blurs the alignment signal. This blur supposedly caused the discrepancy between the positive findings of Pelgrims and Hutsemékers (2015) and negative findings of Joshi et al. (2007). An additionally discovered difference was that adding other source species than QSOs completely blurs the alignment. Ultimately, alignment was found at the 0.1% significance level for QSOs in groups with an angular radius of about 20 degrees (Pelgrims and Hutsemékers, 2015). It was concluded, however, that there are arguments present for both the signal being the result of a bias in the dataset and being a physical effect. The fact that the found alignments were more pronounced in 2D than in 3D argues for a bias in the dataset, while the detection of alignment present in only the QSO objects argues for a physical effect.

More evidence for alignment of quasar radio polarization with large scale structure was found in Pelgrims and Hutsemékers (2016). This study extended the optical study in the previous subsection that found alignment between polarization and large-scale structures. In a sample of 22,381 quasars with defined large quasar groups (LQGs) (Einasto et al., 2014) it was found that rich (≥ 20 members) LQGs have polarization vectors preferentially perpendicular to the major axis of the host LQG. The LQGs with between 10 and 20 members were found to have polarization vectors preferentially parallel to the major axis of the LQG while no signal was found for LQGs with less than 10 members. This strengthened the previous finding of the alignment of optical polarization of AGN with their parent LQG by Hutsemékers et al. (2014), but also added an interesting correlation. Apparently, the alignment effect depended on the richness of the large quasar group, with the orientation switching from preferentially parallel to the major axis of the LQG to preferentially perpendicular at a quasar density of $1.5 * 10^{-5} (h^{-1} \text{Mpc})^{-3}$ (Pelgrims and Hutsemékers, 2016).

1.3.3 Radio structure alignment

Taylor and Jagannathan (2016) corroborated the results of the previously mentioned AGN alignment studies, by studying the alignment of radio jet structures at 612 MHz in one square degree of the ELAIS N1 field. The probability that the position angles of the jets are drawn from a uniform distribution was found to be less

than 0.1%. At angular scales of 0.5° – 1.0° , at redshift $z = 0.9$ this translated to co-moving physical scales of $(20\text{--}40)h^{-1}\text{Mpc}$, supporting the findings in Pelgrims and Hutsemékers (2016).

A large study of radio jet position angles was carried out in Contigiani et al. (2017). Two samples of respectively 30,059 and 11,674 sources over more than 7,000 and 17,000 square degrees were investigated for alignment effects. The former sample is from the Faint Images of the Radio Sky at Twenty centimeters (FIRST) (Becker, White, and Helfand, 1995) survey and the latter from the TGSS Alternative Data Release 1 (Intema et al., 2017). Evidence at the 2% significance level was found for an alignment signal at a median angular source separation of 1.5 degrees in the FIRST sample only. Converted to physical distances, assuming the median redshift of the sample ($z = 0.47$) this was found to correspond to the range $[19, 38] h^{-1}\text{Mpc}$, again supporting the earlier findings of alignment on cluster scales. In the same study, the TGSS sample of 11,674 sources was found to be too sparsely populated to allow for a similar signal (Contigiani et al., 2017).

Chapter 2

The project

2.1 Introduction

As described in Chapter 1, it is clear that there is some compelling evidence for a preferential direction in the orientation of quasar optical polarizations on gigaparsec scales. An explanation was proposed in terms of mixing incoming photons with nearly massless axion-like particles, but this was later discarded due to the absence of an expected circular polarization signature. Earlier attempts to explain the preferential orientation involved systematic effects introduced by a global rotation of the Universe (Birch, 1982), but the evidence for this global rotation has been heavily disputed (Phinney and Webster, 1983; Birch, 1983; Bietenholz and Kronberg, 1984; Kendall and Young, 1984).

The correlation between quasar optical polarization and host galaxy position angles (Joshi et al., 2007; Borguet et al., 2008; Borguet et al., 2011) prompts one then to think that the objects themselves are aligned, perhaps along the large scale structure the quasars reside in. As such, various radio polarization studies and radio structure position angle studies have found preferential orientation of quasar axes with respect to each other or to major axes of hosting large quasar groups (Pelgrims and Hutsemékers, 2016; Taylor and Jagannathan, 2016; Contigiani et al., 2017). The state of the field appears now to have converged to the conclusion that the orientation of active galactic nuclei seems to be influenced by the large scale structure they inhabit.

In this project, we will attempt to explore this hypothesis using the position angles of radio structures in the LOFAR Two-metre Sky Survey Data Release I (Shimwell et al., submitted). First, we aim to develop an automated method to select from the 325,694 radio sources in the LOFAR Two-metre Sky Survey Data Release I those sources that are suitable to fit reliable position angles to. We will focus on selecting mainly double-lobed radio galaxies, to which a clear position angle can be fit. The data will be explained in Section 2.1.1 and the definition of the position angle, which is at the core of this research, will be explained in Section 2.1.2. The methodology to select from the data the appropriate sources and to fit position angles will be explained in Section 2.2. Thereafter, the statistical methods to find global and local deviations from randomness are explained in Section 2.3.

2.1.1 The LOFAR Two-metre Sky Survey

The LOFAR Two-metre Sky Survey (LoTSS) will be a high resolution, low frequency (120-168 MHz) survey of the entire northern sky. As of now, observations are 20% complete and the first data-release, which covers 2% of the northern sky, has been published in Shimwell et al. (submitted). This release covers 424 square degrees in the HETDEX Spring Field region (right ascension 10h45m00s to 15h30m00s and declination 45°00'00" to 57°00'00") and contains 325,694 radio sources with a signal of at least five times the noise. The images have a resolution of 6" with a median sensitivity at 144 MHz of $71\mu\text{Jy}/\text{beam}$ and a positional accuracy better than 0.2".

The catalogue of the 325,694 radio sources was created from the LoTSS images using the Python Blob Detector and Source Finder (PyBDSF) (Mohan and Rafferty, 2015). PyBDSF initially identifies islands of emission which are then decomposed into Gaussians, and these Gaussians are grouped into sources. The peak detection threshold was set as 5σ and a 4σ threshold was used for the boundaries of the detected islands. The output catalogue contains the following columns for every detected source: the source position, peak brightness, integrated flux density, source size, source orientation, the statistical errors and the local root mean square (RMS) noise. Appended to these columns are the mosaic ID, the number of pointings that contribute to the mosaic at the position of the source, and the fraction of the pointings where the source was deconvolved with the point-spread-function.

In addition to the PyBDSF-generated catalogue, a value-added catalogue is produced by identifying optical counterparts to the radio sources. Either a likelihood ratio method (e.g. Richter, 1975) or human visual classification through the LOFAR Galaxy Zoo¹ project is employed. The decision of which technique to use is made through a decision tree, detailed in Williams et al. (submitted). Using this methodology, optical counterparts have been identified for 71% of the sources in the PyBDSF catalogue. Where available, the spectroscopic redshift of the optical counterpart, taken from the Sloan Digital Sky Survey Data Release 14 (Abolfathi et al., 2018), is given in the value-added catalogue. If this was not available, the photometric redshift is estimated using a hybrid methodology based on traditional template fitting and machine learning. This is done by making use of the Panoramic Survey Telescope and Rapid Response System (PanSTARRS) (Chambers et al., 2016) database and the Wide-field Infrared Survey Explorer mission (WISE) (Wright et al., 2010). The specifics of this endeavour are detailed in Duncan et al. (submitted).

The LoTSS, when completed, will detect millions of active galactic nuclei out to high redshifts (Wilman et al., 2008) making it an exceptionally good survey for a study of radio source position angles. Already, with only a small percentage of the northern sky observed, the unprecedented radio source density combined with the high resolution and large number of available redshifts allows for a more thorough study of AGN alignment than ever before.

2.1.2 Position angle

The position angle of a source is the orientation of the source measured relative to the north celestial pole, turning positive in the eastern (increasing right ascension) direction. We will here derive the general expression for the position angle between two points, say \vec{s} and \vec{t} , on the celestial sphere. A point s (equivalently for t) at right ascension α_s and declination δ_s on the celestial sphere has position

¹<https://www.zooniverse.org>

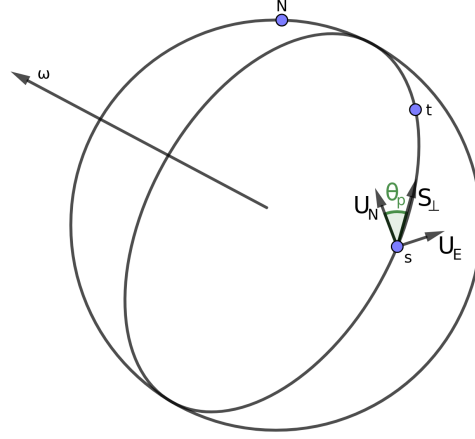


FIGURE 2.1: Illustration showing the celestial sphere, points s and t that are connected through the great circle, the unit vectors pointing north (increasing declination) and east (increasing right ascension), the great circle normal vector ω and the north celestial pole N . The position angle between source s and t , θ_p , is the angle that projects \vec{s}_\perp onto the vector \vec{u}_N .

$$\vec{s} = \begin{pmatrix} \cos \alpha_s \cos \delta_s \\ \sin \alpha_s \cos \delta_s \\ \sin \delta_s \end{pmatrix}. \quad (2.1)$$

The angular distance $\theta_{s,t}$ between point \vec{s} and \vec{t} is calculated by the dot product

$$\vec{s} \cdot \vec{t} = \cos \theta_{s,t} = \sin \delta_s \sin \delta_t + \cos \delta_s \cos \delta_t \cos(\alpha_t - \alpha_s). \quad (2.2)$$

To derive the expression for the position angle between \vec{s} and \vec{t} , we start by reconstructing the great circle between \vec{s} and \vec{t} (Figure 2.1). This great circle lies in the plane that contains the center of the sphere, \vec{s} and \vec{t} by definition, and is constructed by rotating \vec{s} by the angle $\theta_{s,t}$ around an axis $\vec{\omega}$. This axis is perpendicular to the great circle, and can thus be computed by the normalized cross product of \vec{s} and \vec{t} ,

$$\omega = \frac{1}{\sin \theta_{s,t}} \vec{s} \times \vec{t} = \frac{1}{\sin \theta_{s,t}} \begin{pmatrix} \cos \delta_s \sin \alpha_s \sin \delta_t - \sin \delta_s \cos \delta_t \sin \alpha_t \\ \sin \delta_s \cos \alpha_t \cos \delta_t - \cos \delta_s \sin \delta_t \cos \alpha_s \\ \cos \delta_s \cos \delta_t \sin(\alpha_t - \alpha_s) \end{pmatrix}. \quad (2.3)$$

We then define a right-handed coordinate system $(\vec{s}, \vec{\omega}, \vec{s}_\perp)$ with

$$\vec{s}_\perp = \vec{\omega} \times \vec{s} = \frac{1}{\sin \theta_{s,t}} \begin{pmatrix} \cos \delta_t \cos \alpha_t (\sin^2 \delta_s + \cos^2 \delta_s \sin^2 \alpha_s) - \cos \alpha_s (\sin \delta_s \cos \delta_s \sin \delta_t + \cos^2 \delta_s \sin \alpha_s \cos \delta_t \sin \alpha_t) \\ \cos \delta_t \sin \alpha_t (\sin^2 \delta_s + \cos^2 \delta_s \cos^2 \alpha_s) - \sin \alpha_s (\sin \delta_s \cos \delta_s \sin \delta_t + \cos^2 \delta_s \cos \alpha_s \cos \delta_t \cos \alpha_t) \\ \cos \delta_s (\cos \delta_s \sin \delta_t - \sin \delta_s \cos \delta_t \cos(\alpha_t - \alpha_s)) \end{pmatrix}. \quad (2.4)$$

A position \vec{r} in this basis, on the great circle, is given by $\vec{r} = \cos \theta \vec{s} + \sin \theta \vec{s}_\perp$ with $0 \leq \theta \leq 2\pi$. The position angle θ_p is then given by computing the gradient of this position vector at point \vec{s} ($\theta = 0$) and splitting the resulting direction, \vec{s}_\perp into two orthogonal directions pointing north and east in the plane tangential to the sphere at point \vec{s} . These two orthogonal directions can be computed by the partial derivatives of \vec{s} (Eq. 2.1) with respect to α and δ :

$$\vec{u}_N = \begin{pmatrix} -\sin \delta_s \cos \alpha_s \\ -\sin \delta_s \sin \alpha_s \\ \cos \delta_s \end{pmatrix}, \vec{u}_E = \begin{pmatrix} -\sin \alpha_s \\ \cos \alpha_s \\ 0 \end{pmatrix}. \quad (2.5)$$

The position angle projects the gradient direction \vec{s}_\perp onto these two vectors pointing north and east, and thus measures a positive angle from north due east,

$$\vec{s}_\perp = \cos \theta_p \vec{u}_N + \sin \theta_p \vec{u}_E. \quad (2.6)$$

With clever multiplication by the unit vectors pointing north and east we arrive at

$$\cos \theta_p = \vec{s}_\perp \cdot \vec{u}_N = \frac{1}{\sin \theta_{s,t}} (\cos \delta_s \sin \delta_t - \sin \delta_s \cos \delta_t \cos(\alpha_t - \alpha_s)), \quad (2.7)$$

$$\sin \theta_p = \vec{s}_\perp \cdot \vec{u}_E = \frac{1}{\sin \theta_{s,t}} (\cos \delta_t \sin(\alpha_t - \alpha_s)). \quad (2.8)$$

Which we can finally rewrite as the known expression for the position angle between point \vec{s} and \vec{t} on the celestial sphere

$$\tan \theta_p = \frac{\sin(\alpha_t - \alpha_s)}{\cos \delta_s \tan \delta_t - \sin \delta_s \cos(\alpha_t - \alpha_s)}. \quad (2.9)$$

2.2 Selection of sources

The first task of this research is to select from the 325,694 radio sources the subset of sources that are appropriate to fit position angles to. Point sources obviously have no clear orientation and should therefore be excluded from our sample, as do sources with an irregular shape, as in Figure 2.2. Our focus lies in selecting mainly double-lobed radio galaxies, with a clearly definable position angle. There are, in this problem, two different morphology types of these double-lobed sources. The source can either have a bridge of emission between the lobes, or can be two isolated lobes. Examples are provided in Figures 2.3 and 2.4. The doubles without a bridge of emission are found as two separate sources in the PyBDSF catalogue, and should be matched on the basis of their distance and resemblance to each other. This has been performed by the process of compiling the value-added catalogue, either by likelihood ratio or visual matching (see Section 2.1.1). However, as the value-added catalogue was not yet finished at the start of, and during a large part of this project, we implement our own algorithm to match these isolated lobes. The doubles with a bridge of emission between the lobes are classified as a single source by PyBDSF and are always fit with multiple Gaussian components, due to their extended emission. Hence, we take two separate approaches for selecting the two different types of doubles. To define a clean starting point for our selections, we will limit our sample to only those sources with a signal-to-noise ratio above 10, as is also adapted in Contigiani et al. (2017). The starting point for the selection of sources then contains 101,662 sources. The approaches for the ‘isolated lobes’ and ‘connected lobes’ sources will be explained in Sections 2.2.1 and 2.2.2 respectively.

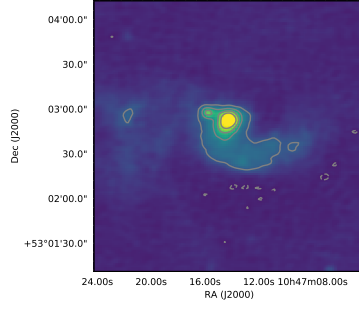


FIGURE 2.2: Example of a radio source where a clear orientation or position angle is not present. The 1σ to 5σ contours are shown in steps of 1σ . This source should be excluded from an alignment analysis sample

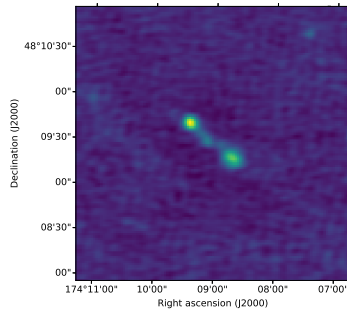


FIGURE 2.3: Example of a clearly double-lobed source with a bridge of emission between the two lobes. PyBDSF has found a single source which is fit with multiple Gaussian components.

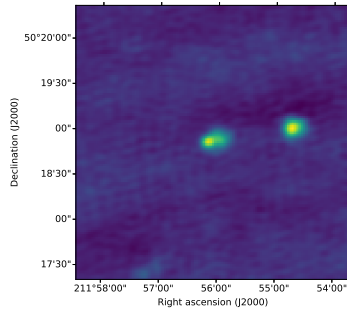


FIGURE 2.4: Example of a possible double-lobed source where there is no bridge of emission between the lobes. PyBDSF has found two single sources which could be matched based on proximity.

2.2.1 Isolated radio lobes

To find the sources that could be double-lobed sources but are classified as two separate sources by PyBDSF because there is no emission between them, the nearest neighbour of every source is crucial information. For every source in the initial sample of $S/N > 10$ sources, fitted by a single (S) or multiple (M) Gaussian component(s), the nearest neighbour is calculated. This result is filtered by taking the subset of sources that have a nearest neighbour within a certain cutoff distance C . Then, the remaining set is divided into ‘mutuals’ and ‘loners’, where mutuals indicates two sources that point to each other as nearest neighbour (mutual nearest

neighbours) and loners indicates sources that do not (i.e., source X would be a loner if it has source Y as nearest neighbour while source Y has source Z as nearest neighbour). The loners are then discarded and the mutuals are kept as potential double lobed radio galaxies. The cutoff distance C is an important parameter in this selection. Therefore, below we discuss statistical and physical motivations for this cutoff value.

Statistical motivation

When matching neighbouring sources that do not have a bridge of emission between them, it is important to know the prior probability that a source has an unrelated nearest neighbour positioned within a certain angular distance cutoff C . Therefore, we will analyze the random distribution of N sources on the celestial sphere, following Scott and Tout (1989).

We parametrize the unit sphere according to the convention commonly used in astrophysics, with spherical coordinates θ and ϕ , where θ is the polar angle and ϕ is the azimuthal angle. A uniform distribution of sources on a sphere does not translate directly to a uniform distribution of the corresponding spherical coordinates. This is due to the surface area of the sphere $dA = r d\theta \cdot r \sin \theta d\phi$ reducing in size near the poles (θ near 0 or π). We know that a uniform probability distribution f of points v on the unit sphere must obey

$$f(v) = \frac{1}{4\pi}, \quad (2.10)$$

since

$$\int_s dA = 4\pi. \quad (2.11)$$

This can be used in order to determine $f(\phi, \theta)$:

$$f(v)dA = \frac{1}{4\pi}dA = f(\phi, \theta)d\phi d\theta. \quad (2.12)$$

Therefore,

$$f(\phi, \theta) = \frac{\sin \theta}{4\pi}. \quad (2.13)$$

From Equation 2.13 we can separate the distribution into two functions for the individual spherical coordinates:

$$f(\phi) = \int_0^\pi f(\phi, \theta)d\theta = \frac{1}{2\pi}; \quad (2.14)$$

$$f(\theta) = \int_0^{2\pi} f(\phi, \theta)d\phi = \frac{\sin \theta}{2}. \quad (2.15)$$

From which we can see that ϕ is uniformly distributed but θ is not. This will produce more points around the equator and less points around the poles as is expected from the equation of the surface element of the sphere. The cumulative distribution function (CDF) $F(\theta)$ is calculated from integrating the probability density function (PDF) $f(\theta)$:

$$F(\theta) = \int_0^\theta f(\theta)d\theta = \frac{1}{2}(1 - \cos \theta). \quad (2.16)$$

Let us now take one point on the surface of the unit sphere. Due to spherical symmetry, without loss of generality we can choose the point at the pole of the sphere.

The nearest neighbour angular distance θ_C is then simply given by the difference in polar angle. The probability of any of the other $N - 1$ points to be the nearest neighbour at distance θ_C is then given by two probabilities. First, the nearest neighbour has probability $\frac{1}{2} \sin(\theta_C)$ (cf. Eq 2.15) to be in the range $(\theta_C, \theta_C + d\theta)$. Next, the other $N - 2$ sources must be further away, which corresponds to the probability $2^{2-N} (1 + \cos(\theta_C))^{N-2}$. Together, these give the PDF of a nearest neighbour at angular distance θ_C :

$$P(\theta_C)d\theta = \frac{N-1}{2^{N-1}} \sin \theta_C (1 + \cos \theta_C)^{N-2} d\theta. \quad (2.17)$$

It should be noted that this PDF for the spherical case is nearly identical to the PDF of points distributed on a plane for $N \geq 38$, as verified by Scott and Tout (1989). The PDF of the plane case is computationally much simpler, being

$$P_{plane}(\theta) = 2\pi n\theta \exp(-\pi n\theta^2), \quad (2.18)$$

with $n = N/4\pi$ now being the number density on the plane. Continuing the spherical approach, integrating Equation 2.17 yields the angular distance CDF, which is

$$Pr(\theta_C < \Theta) = 1 - \left(\cos \frac{\Theta}{2} \right)^{2(N-1)}. \quad (2.19)$$

From Equation 2.17 we can estimate the expected nearest neighbour angular separation value $\bar{\theta}_C$ (mean) and most likely value θ^* (mode). The expected value is calculated by

$$\bar{\theta}_C = \int_0^\pi \theta_C P(\theta_C) d\theta \quad (2.20)$$

which after some intermediate steps detailed in Scott and Tout (1989) produces

$$\bar{\theta}_C = \frac{\pi(2N-2)!}{2^{2N-2}(N-1)!^2} \quad (2.21)$$

or with the Stirling approximation

$$\bar{\theta}_C = \sqrt{\frac{\pi}{N}}. \quad (2.22)$$

The most probable value is found by calculating the maximum of the PDF (i.e., by solving $\frac{d}{d\theta_C} P(\theta_C) = 0$), which results in

$$\theta^* = \arccos \frac{N-2}{N-1}. \quad (2.23)$$

Furthermore, the probability that any two points on the celestial sphere are related can be estimated from the CDF. For example, if we wish to find all nearest neighbour pairs that are more than 90% likely to be part of the same source, we can select all sources with a separation less than Θ given by

$$Pr(\theta < \Theta) = 0.10. \quad (2.24)$$

With these tools at hand, we can statistically motivate a cutoff distance C for nearest neighbour matching. After selecting all sources that have a signal-to-noise ratio of at least 10, we are left with 101,662 sources in our 424 square degrees sample.

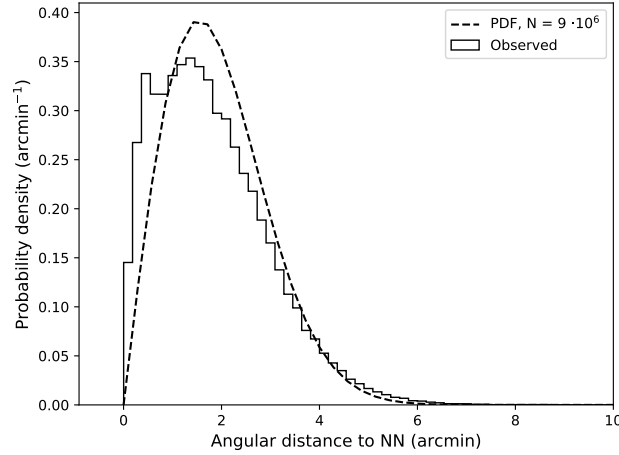


FIGURE 2.5: Histogram of the distribution of nearest neighbour angular distances in the sample of 101,662 sources with $S/N > 10$. The PDF of a uniform distribution of sources is indicated as well. An excess probability over random is present at small angular distances.

This corresponds to a source density $n \approx 240$ per square degree, which would translate to a total number of sources on the entire celestial sphere $N \approx 9.9 \cdot 10^6$. The mean nearest neighbour separation is then calculated using Equation 2.22 yielding $\bar{\theta}_C = 1.94$ arc minutes. The most probable value for the angular distance to the nearest neighbour (Eq. 2.23) is $\theta_* = 1.55$ arc minutes. For 90% statistical certainty that two sources are actually part of the same source, we find with Equation 2.24, using an approximation of the CDF

$$\frac{N-1}{4} \Theta^2 = 0.10, \quad (2.25)$$

for $\Theta = 0.69$ arc minutes.

To verify the above results, we show in Figure 2.5 the calculated nearest neighbour angular distance for the PyBDSF sources with a signal-to-noise ratio higher than 10, together with the probability density function as in Equation 2.18. Figure 2.5 shows that the PDF overlaps quite nicely with the observed angular distance, except at smaller angular distances, where the observations follow a wider distribution, shifted a little to the left. This excess probability over random at small angular distances indicates the sources that are identified as separate sources by PyBDSF but should be matched and identified as one source.

Physical motivation

The mathematical derivations provide an important, but purely statistical baseline. As we are considering physical objects, a physical motivation for the angular cutoff should also be considered. The physical separation of radio lobes typically ranges between 1 and 1000 kpc (Willott et al., 1999; Sadler et al., 2007), where the largest radio galaxies form only a small percentage of the total population. To find the distribution of radio source size in the LoTSS data, we select from the value-added catalogue all sources that have $S/N > 10$ and a redshift measurement available. The physical size can then be determined, assuming concordance cosmology with Planck 15 cosmological parameters ($H_0 = 67.8 \text{ kms}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.308$, $\Omega_\Lambda = 0.692$) (Planck Collaboration et al., 2016). Figure 2.6 shows this distribution of physical

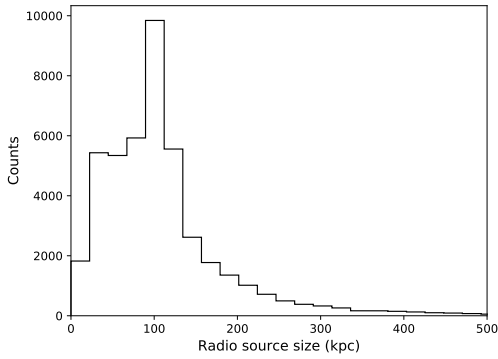


FIGURE 2.6: Radio source size distribution for the 44,226 sources in the LoTSS dataset that have an SN-ratio over 10 and a determined photometric or spectroscopic redshift. 96% of all sources are smaller than 300 kpc in extent.

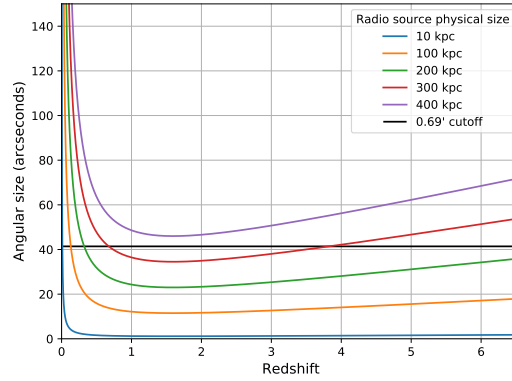


FIGURE 2.7: Observed angular size for different physical source sizes as a function of redshift. The plot is drawn up to redshift 6.6, which is the highest measured redshift in the catalogue.

sizes, from which it becomes clear that most of the sources are below 200 kpc in size, and nearly all of the sources are below 400 kpc in size. Additionally, we can plot the observed angular distance as a function of redshift, for different-sized sources up to 400 kpc. This is shown in Figure 2.7. From this figure it can be readily appreciated that for an angular size cutoff of 0.69 arc minutes, we find nearly all sources up to 100 kpc in size for all redshifts in the sample, and a large portion of sources up to 200 kpc. It should also be noted that this figure provides an upper limit for the observed angular size, since it implicitly assumes that all sources lie perpendicular to the line of sight. Thus, from combining the results of the statistical derivation, Figure 2.6 and Figure 2.7, we can safely assume an angular cutoff distance $C = 0.69'$ will not eliminate many associated radio sources or match many unrelated radio sources.

Decision tree

After selecting the nearest neighbour sources with an angular distance cutoff of $0.69'$ we are left with 8,503 mutuals, made up of two neighbouring sources. These mutuals should now be filtered based on source morphology and flux ratio between the lobes, to be certain that we are left with only clear double-lobed sources to which a position angle can be fit with low uncertainty. Therefore, the mutuals will be run through an algorithm (inspired by Oei. 2017 MSc. thesis) that finds features indicating clear double lobes as well as the position angle. The properties are easier to explain after we have explained the algorithm. An example output of the algorithm is shown in Figure 2.8 and it works as follows:

1. Create a cutout around the mutual nearest neighbour sources and center a horizontal line segment on the projected middle of the sources with radius $d_{nn}/2 + 4$ pixels. Where d_{nn} is the distance between the sources in pixels. The extra 4 pixels is chosen in an attempt to capture the entire source.

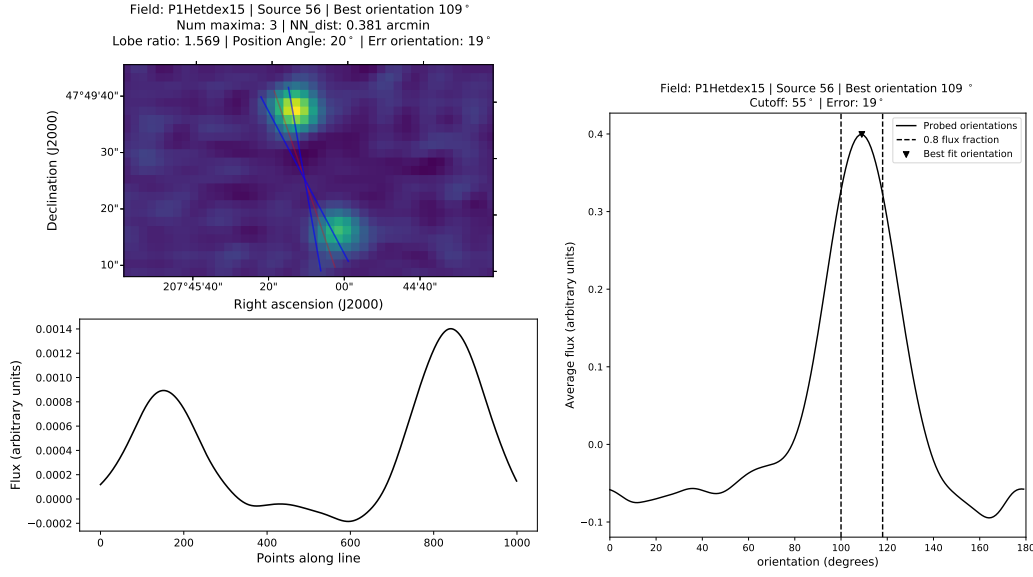
2. Rotate the line in increments of 1 degree around the center of the image and calculate the mean flux along this line. The orientation (note that this is different from the position angle) of the line that results in the maximal mean flux along the line is saved and this line is adopted as the best fitting line.
3. The local maxima in the flux along this best fitting line are found (Fig. 2.8 lower panel). Depending on the amount of maxima a position angle is calculated according to Equation 2.9 as in step (4).
4. If there is one maximum along this line a position angle is not calculated, as this does not indicate the double-lobed structure that is desired. If there are two maxima along this line, the position angle is defined as the position angle between the two maxima. And if there are more than two maxima, the position angle is defined as the position angle between the brightest maxima.

After the best orientation and the position angle are found a number of other properties are also returned. These properties are the amount of local maxima along the line, the ‘lobe ratio’ (the flux ratio between the two brightest maxima), and the ‘error’ in the optimal orientation of the line, being defined as the amount of degrees for which the mean flux along the line is more than 0.8 times the maximum mean flux. These properties are chosen as they can be used as criteria for automatically finding double-lobed sources.

The decision tree in Figure 2.11 is then constructed, based on the properties found by the algorithm. We reject immediately all mutuals that have only one maximum along the best fitting line as this is indicative of cases where PyBDSF identified a nearby neighbour that was simply noise. These are mostly single point sources that are erroneously matched to neighbouring noise. The remaining 7,445 mutuals are then split on the lobe ratio property where we take a very conservative approach to minimize the amount of incorrectly matched doubles and reject all sources that have a lobe ratio larger than two (usually, larger ratios are used, e.g., Prandoni et al., 2000; van Velzen, Falcke, and K rding, 2015; Contigiani et al., 2017). Finally, the last remaining 3,734 sources are split on the orientation error. A cutoff value is determined uniquely for every source, based on the lobe separation and sizes. The cutoff value is defined as the mean of the opening angles of the first and second radio lobe with respect to the projected center of the lobes, shown schematically in Figure 2.9. Mathematically, the cutoff value C is given by

$$C = \arctan\left(\frac{b_1}{d_{nn}/2}\right) + \arctan\left(\frac{b_2}{d_{nn}/2}\right) \quad (2.26)$$

where b_1 and b_2 are the semi-minor axes of the PyBDSF source and its nearest neighbour respectively, and d_{nn} is the distance between the sources. The semi-minor axes are used, because double lobed sources are expected to be elongated along the line connecting the two lobes. This mainly removes sources of diffuse emission that by chance have more than two flux maxima and a lobe ratio ≤ 2 . It also removes mutuals where a third source is close enough to influence the best fitting orientation of the mutual or when the sources are not comparable in size. Thus, if the orientation error, as defined before, is larger than the cutoff value, the source is excluded from the sample. Finally, this produces 3,631 mutuals that are matched on proximity and morphology and have a clear position angle.



(A) Example output after the best orientation for the line connecting the two sources is found. Three local flux maxima are found along the best fitting line (lower panel). In the upper panel, the two neighbouring sources are shown. The best fitting line is indicated by the red line, and the boundary lines for which the mean flux is 80% of the mean flux along the best fitting line are indicated by the blue lines.

(B) The mean flux along the red line in (A) versus the orientation of the line. For this source, rotating the initial horizontal line by 109 degrees gives the maximum mean flux along the line. This line is taken as the best fitting line. The error in the orientation of the line as defined in the text is 19 degrees, indicated by the dashed lines.

FIGURE 2.8: Example of the output of the algorithm to find the position angle of neighbouring sources mentioned in the text. Left (A) shows the final properties of the best fitting line. Right (B) shows the mean flux along the probed lines as a function of orientation of the line.

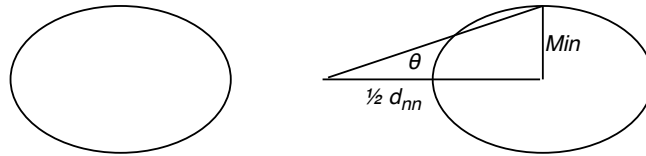


FIGURE 2.9: Illustration of the cutoff value for the orientation error of mutual nearest neighbour sources. The two ellipses represent the two radio lobes, d_{nn} is the distance between the two lobes and Min the minor axis of one of the lobes, as fit by PyBDSF. The actual cutoff value is the mean of θ calculated on both lobes, as defined in the text.

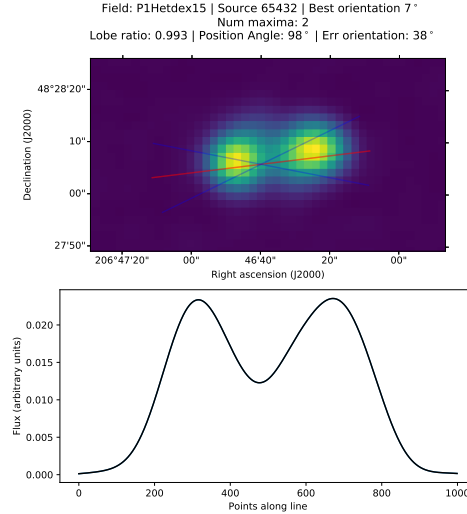


FIGURE 2.10: Example of the output of the algorithm to find the position angle of sources fit by multiple Gaussian components mentioned in the text. For this source, rotating the red line by 7 degrees gives the maximum mean flux along the line. Two local maxima are found along this line. The boundary lines for which the mean flux along the line is 80% of that of the best fitting line are indicated by the blue lines.

2.2.2 Connected radio lobes

The first selected subset of sources contains sources that are fitted by a single Gaussian (S) or multiple Gaussians (M). For the nearest neighbour matching, we did not differentiate between these. However, often when a single source is fitted by multiple Gaussians, it is an interesting source on its own. These can be double-lobed galaxies where a bridge of emission is present between the two bright lobes. For these sources, a slightly different method to calculate the position angle and the other previously defined properties is applied. Again, we start by selecting all sources that have a signal-to-noise ratio of at least 10 but this time also restrict our sample to sources that have a PyBDSF identification M. The following steps are then performed:

1. Sources that are round, defined as the semi-major axis of the source being within one arcsecond of the semi-minor axis, are removed as a position angle fit is not reliable.
2. A line segment is centered on the center of the source, with a radius that is equal to the semi-major axis of the source + 2 pixels, again to attempt to capture the whole source.
3. The line segment is rotated to find the orientation corresponding to the maximal mean flux along the line. Then the local maxima in the flux along this line are found and the position angle is calculated identically to step (4) of the algorithm in the previous subsection.

After this algorithm is executed on all selected sources, we are left with 18,522 sources that we will call ‘bridged’ sources. A decision tree is constructed for these sources as well, based on the same properties that were used for the mutual nearest neighbour sources. The cutoff value is defined slightly different for these bridged

sources since these have no nearest neighbour distance. Similar to Equation 2.26, the cutoff value is now defined as

$$C = 2 \arctan(b/a), \quad (2.27)$$

where b is the semi-minor axis of the source and a is the semi-major axis of the source. This implicitly assumes symmetric sources, which is not a bad assumption to make since the sources have by this point been propagated through the decision tree already, increasing the chance of symmetric sources.

Employing the same cuts as defined for the mutuals, we retain 11,984 out of 18,522 sources after the cut on the number of flux maxima, from which 6,576 have a lobe ratio equal to or smaller than two. Finally, cutting these sources depending on the orientation error results in 4,212 bridged double lobed sources (see Figure 2.11).

2.2.3 Producing the final sample

After the mutuals and bridged sources have been propagated through their respective algorithms and the decision tree to produce the sample of mutuals and bridged sources, it is tempting to simply concatenate the two samples to form the final sample. However, overlap is likely to exist, since the bridged sources have been formed from identification ‘M’ sources while the mutuals have been formed from both identification ‘S’ and ‘M’ sources. This overlap is easily found through comparing the source names of the two samples. We find 296 duplicate entries, which are visually inspected to be included as either a mutual, a bridged source or excluded entirely. After visual inspection, 167 of the double entries are found to be bridged sources erroneously matched to an unrelated nearest neighbour, 107 sources are found to be mutuals where one or both of the lobes was fit by multiple Gaussian components and 22 sources are excluded as these lie in the grey area between the two possibilities. This results in 6,801 sources in the concatenated sample of isolated and connected lobes.

We also found that PyBDSF in a few cases splits up large sources of emission into multiple smaller PyBDSF sources, while these actually all indicate the same source of emission. This leads to finding different sources with very similar position angle very close to each other. As this would obviously influence the signal when checking for local preferential alignment, we decide to check the distance to the nearest source for our final sample of 6,801 sources and find that 6 sources, with nearest neighbour distance ranging from 0.03 to 0.19 arc minutes, are duplicate sources. Therefore, we decide to define a duplicate entry as a source that has a nearest neighbour in the concatenated sample within 0.2 arc minutes. For all duplicate entries the nearest neighbour is removed, resulting ultimately in 6,795 sources in the final sample as shown in Figure 2.11.

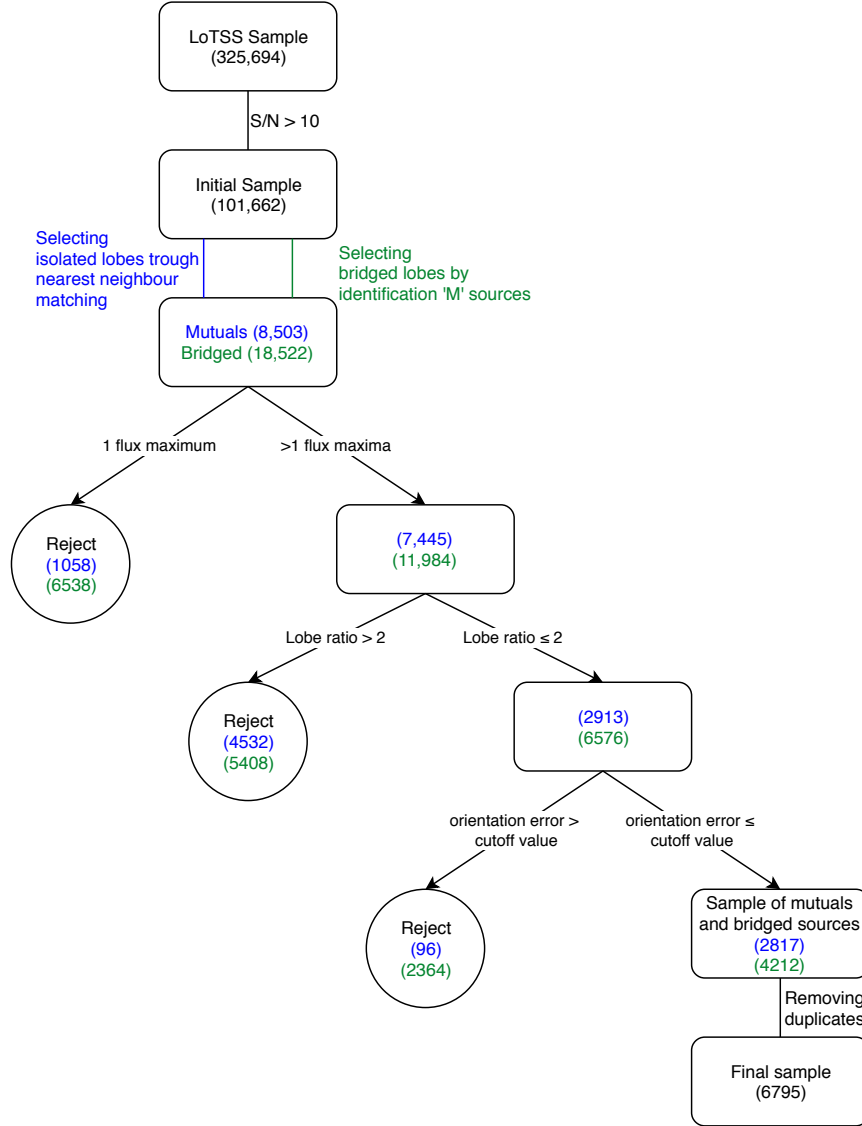


FIGURE 2.11: Decision process for the selection of the sources, aiming for clear double-lobed structures. The source counts are shown for every step along the tree, separated between the selection of isolated lobes (blue) and lobes with a bridge of emission (green).

2.2.4 Defining sample subsets

As mentioned in Section 2.2, a value-added catalogue (see also Section 2.1.1) became available during the development of the source selection process of this project. Therefore, we can now join our final sample with the value-added catalogue, providing us with information about the redshift of the sources, and the quality of our nearest neighbour matching. To recap, for the nearest neighbour matching in the value-added catalogue a likelihood ratio method was used for the small sources (less than 30"), and the LOFAR Galaxy Zoo (LGZ) was employed for large sources (more than 30") (Williams et al., [submitted](#)). In the LGZ sources were visually inspected and volunteers identified possible additional components to go with PyBDSF sources. Therefore, we can compare our sample of mutuals to the sample of matched component sources in the value-added catalogue. We find that out of 6975 sources, 6433 sources have an entry in the value-added catalogue. The missing sources might

be, as stated by Williams et al. (submitted), sources that have been grouped into a single source, spurious emission, or artefacts. In this joined table, the amount of mutuals is 2540 and the amount of connected lobes is 3893. We investigate how our 2540 matched isolated lobes compare to the sources in the value-added catalogue by checking the amount of components (i.e., PyBDSF sources), a source is made up of in to the value-added catalogue. According to the value-added catalogue, only 793 out of our 2540 mutuals are actually multiple component sources. Out of these 793 sources, we find that for 762 we have identified the correct components. Thus, as stated by the value-added catalogue 31 sources are indeed multiple component sources, but for these we have identified the wrong components and 1747 sources are single component sources that we have erroneously matched to a nearby component. Based on these results, we can now define different subsets of our sample, to pinpoint the effect that a particular selection method might have on the eventual statistical results. The different subsets are the following

Initial sample The sample consisting of the final 6795 sources as defined by our selection methods, with no additional information added. We choose to include this sample so that we are able to investigate the effect possible artefacts might have on the final result.

Value-added subset The 6433 sources out of the initial sample that have an entry in the value-added catalogue. The ‘missing’ 362 sources were either flagged as artefacts or grouped together into single sources by the workflow of the value-added catalogue.

Connected lobes subset The 3893 connected lobes with an entry in the value-added catalogue. This subset allows us to investigate alignment effects without having to take possible errors in nearest neighbour matching into account.

Isolated lobes subset The complement of the previous subset, entailing the other 2540 sources that have an entry in the value added catalogue (be it a single or multiple component source according to the value-added catalogue), but are found through nearest neighbour matching. In contrary to the previous subset, this subset allows us to investigate the effect of only proximity-matched isolated lobes.

Concluding subset The connected lobes subset plus the 762 —according to the value-added catalogue— correctly matched mutuals, for a total of 4655 sources. Due to this subset being the result of two separate methods agreeing on the matching of separate source components it has a very low chance of containing wrongly matched components. This subset thus represents a sample that contains low contamination and contains the least selection biases.

Combining the results of these five subsets will allow us to filter out signals that are caused by different selection effects. Finally, for every subset except the first, we can extract the sources that have a redshift measurement available in the value-added catalogue. This allows a study of alignment in 3D space, and provides the information to calculate the physical properties of the sources. Thereafter, we can also divide the sample into high and lower power and high and low physical size sources. The first subset can only be divided into high and low flux or angular size since it is not matched to the value-added catalogue, but it will provide a basis to compare results to other 2D alignment effect studies.

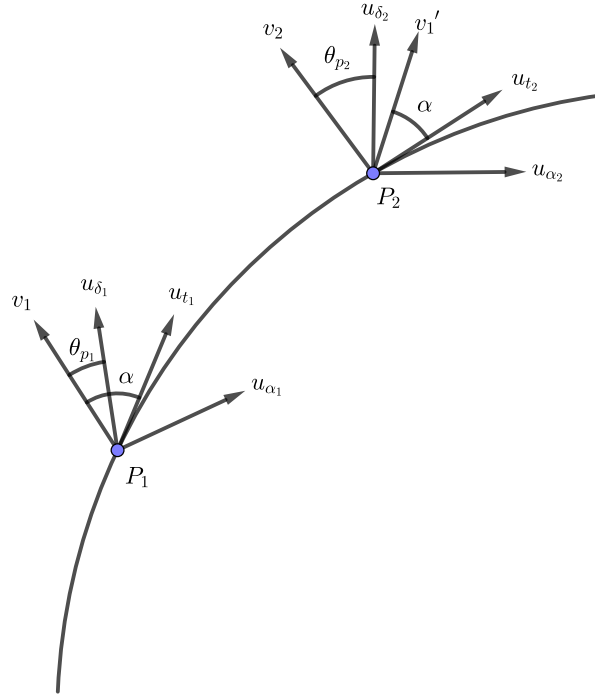


FIGURE 2.12: Illustration of parallel transport problem. Vector v_1 corresponding to a position angle θ_{p_1} and vector v_2 corresponding to position angle θ_{p_2} are shown. In order to compare v_1 to v_2 , v_1 must be parallel transported along the great circle indicated by the curve from location P_1 to location P_2 . The transported vector is indicated by v_1' and the local basis vectors are denoted by (u_δ, u_α) . In parallel transport, the angle α between the vector tangent to the sphere u_t and the vector v remains fixed. Figure adapted from Jain, Narain, and Sarala (2004).

2.3 Statistics

The second task of this research is to find whether the selected sources have a preferential position angle, either globally or locally. To determine the alignment on the sky (i.e., the celestial sphere), an appropriate statistical method must be used. In particular, it must account for the circular nature of position angles. Since a position angle of 179 degrees is very close to a position angle of one degree, standard statistical definitions such as the mean do not apply. Worse still, when considering the alignment of position angles on a sphere we cannot simply compare angles that are defined on different parts of the sphere. To tackle these problems, the notion of parallel transport and the dispersion measure are introduced in this section.

2.3.1 Parallel transport

As, by definition (Eq. 2.9), the position angle is defined with respect to the local meridian, the vectors corresponding to the position angles at different points of the celestial sphere cannot be compared directly. These vectors must be carefully transported along the great circle joining these points. Following Jain, Narain, and Sarala (2004), we will derive the method to transport a vector to a different position on the celestial sphere.

We parametrize the celestial sphere with local unit vectors $(\vec{u}_r, \vec{u}_\delta, \vec{u}_\alpha)$ which point respectively to the center of the sphere, north along the local meridian, and eastwards on the sphere. The problem will be defined as follows, we wish to compare the position angles θ_{p_1} and θ_{p_2} of sources 1 and 2 with positions P_1 and P_2 on the celestial sphere (Figure 2.12). Similar to Equation 2.6, the vector resulting from the position angle θ_{p_1} of source 1 at location P_1 is given, in terms of the local basis, by

$$\vec{v}_1 = \cos \theta_{p_1} \vec{u}_{\delta_1} + \sin \theta_{p_1} \vec{u}_{\alpha_1}. \quad (2.28)$$

According to the standard definition of the inner product, if we were to compare this vector \vec{v}_1 to another vector \vec{v}_2 with position angle θ_{p_2} defined at position P_2 on the sphere the result would be $\vec{v}_1 \cdot \vec{v}_2 = \cos(\theta_{p_1} - \theta_{p_2})$. However, this definition does not take into account the fact that the position angles are defined locally and thus this inner product is coordinate-dependent. To define a coordinate-invariant inner product we parallel transport the vector \vec{v}_1 to the position P_2 to obtain vector \vec{v}_1' . Vector \vec{v}_1' then makes an angle θ_p' with respect to the local north-pointing vector \vec{u}_{δ_2} . To find the transported angle θ_p' , let \vec{u}_s be the unit vector perpendicular to the plane containing the two radial vectors \vec{u}_{r_1} and \vec{u}_{r_2} . Thus \vec{u}_s is found by

$$\vec{u}_s = \frac{\vec{u}_{r_1} \times \vec{u}_{r_2}}{|\vec{u}_{r_1} \times \vec{u}_{r_2}|}. \quad (2.29)$$

Consider now the unit vectors \vec{u}_{t_1} and \vec{u}_{t_2} at points P_1 and P_2 , and tangent to the great circle connecting P_1 and P_2 . These vectors are given by

$$\vec{u}_{t_{1,2}} = \vec{u}_s \times \vec{u}_{r_{1,2}}. \quad (2.30)$$

Which in terms of the local basis vectors can be written as

$$\vec{u}_{t_1} = \vec{u}_{\delta_1} \cdot \vec{u}_{t_1} \vec{u}_{\delta_1} + \vec{u}_{\alpha_1} \cdot \vec{u}_{t_1} \vec{u}_{\alpha_1}, \quad (2.31)$$

where

$$\vec{u}_{\delta_1} \cdot \vec{u}_{t_1} = \frac{-\sin \delta_1 \cos \delta_2 + \cos \delta_1 \sin \delta_2 \cos(\theta_{p_1} - \theta_{p_2})}{\sqrt{1 - (\vec{u}_{r_1} \cdot \vec{u}_{r_2})^2}}, \quad (2.32)$$

$$\vec{u}_{\alpha_1} \cdot \vec{u}_{t_1} = \frac{\sin \delta_2 \sin(\alpha_2 - \alpha_1)}{\sqrt{1 - (\vec{u}_{r_1} \cdot \vec{u}_{r_2})^2}}, \quad (2.33)$$

$$\vec{u}_{\delta_2} \cdot \vec{u}_{t_2} = \frac{-\sin \delta_2 \cos \delta_1 + \cos \delta_2 \sin \delta_1 \cos(\theta_{p_1} - \theta_{p_2})}{\sqrt{1 - (\vec{u}_{r_1} \cdot \vec{u}_{r_2})^2}}, \quad (2.34)$$

$$\vec{u}_{\alpha_2} \cdot \vec{u}_{t_2} = \frac{-\sin \delta_1 \sin(\alpha_1 - \alpha_2)}{\sqrt{1 - (\vec{u}_{r_1} \cdot \vec{u}_{r_2})^2}}. \quad (2.35)$$

As \vec{v}_1 is parallel transported along the great circle to position P_2 its angle with respect to the tangent to the great circle remains fixed. Thus, to determine the angle by which the vector has turned due to this transport we consider the orientation of \vec{u}_{t_1} and \vec{u}_{t_2} with respect to the local basis at the two points where the sources lie. Call ζ_1 the angle between \vec{u}_{t_1} and \vec{u}_{α_1} and ζ_2 the angle between \vec{u}_{t_2} and \vec{u}_{α_2} . These angles are given, per definition of the inner product, by

$$\zeta_{1,2} = \arccos(\vec{u}_{\alpha_{1,2}} \cdot \vec{u}_{t_{1,2}}) \quad (2.36)$$

where the sign of the resulting angle ξ_1 or ξ_2 is equal to the sign of $\vec{u}_{\delta_1} \cdot \vec{u}_{t_1}$ or $-\vec{u}_{\delta_2} \cdot \vec{u}_{t_2}$ respectively. The transported \vec{v}_1' thus makes an angle $\theta'_{p,1} = \theta_{p,1} + (\xi_2 - \xi_1)$ defined with respect to the local coordinates in P_2 . Hence we can now define the generalized dot product between \vec{v}_1 and \vec{v}_2 as the dot product between the transported vector \vec{v}_1' and \vec{v}_2 :

$$\vec{v}_1 \odot \vec{v}_2 = \vec{v}_1' \cdot \vec{v}_2 = \cos(\theta_{p_1} - \theta_{p_2} + \xi_2 - \xi_1). \quad (2.37)$$

Here we have derived the practice for transporting position angles along a sphere which can be used in general for any problem that considers angles on a sphere. In the particular picture of comparing the difference between position angles, it makes sense to redefine the generalized inner product between two position angles as

$$(\theta_{p_1}, \theta_{p_2}) = \cos[2(\theta_{p_1} - \theta_{p_2} + \xi_2 - \xi_1)], \quad (2.38)$$

which, since the position angles range from 0 to π , assumes values between -1 and +1 where +1 expresses perfect alignment between θ_{p_1} and θ_{p_2} and -1 indicates perpendicular orientations.

2.3.2 Statistical test

To test the significance of the alignment of position angles we introduce the dispersion measure (Jain, Narain, and Sarala, 2004; Contigiani et al., 2017). The dispersion depends only on the differences between neighbouring position angles and is therefore a suitable choice when testing preferential alignment, as in this work we will not investigate in which direction the preference lies but simply whether there is a preferred direction. The dispersion of a source i as a function of a position angle θ is defined as

$$d_{i,n_n}(\theta) = \frac{1}{n_n} \sum_{k=1}^{n_n} (\theta, \theta_k), \quad (2.39)$$

where n_n is the number of nearest neighbours that is considered around source i , including the source itself and θ_k is the position angle of the respective neighbours. The generalized inner product (θ, θ_k) is defined in Equation 2.38.

The position angle θ that maximizes the dispersion around source i is analogous to the definition of the mean position angle of source i and its n_n nearest neighbours. The magnitude of $d_{i,n_n}|_{max}$ is then a measure of the dispersion around this mean. The dispersion can take a maximum value of 1, which corresponds to perfect alignment of all n_n nearest neighbours. To find the value of θ that maximizes the dispersion, we take the derivative of Equation 2.39 with respect to theta and after some intermediate steps we arrive to the following expression for $d_{i,n_n}|_{max}$:

$$d_{i,n_n}|_{max} = \frac{1}{n_n} \left[\left(\sum_{k=1}^{n_n} \cos \theta_k \right)^2 + \left(\sum_{k=1}^{n_n} \sin \theta_k \right)^2 \right]^{1/2}. \quad (2.40)$$

The statistic, to test for alignment in a sample that totals N sources, is then defined as

$$S_{n_n} = \frac{1}{N} \sum_{i=1}^N d_{i,n_n}|_{max}, \quad (2.41)$$

which is simply the average of the maximum dispersion for a number of nearest neighbours n_n calculated over all N sources in the sample. This statistic thus measures the strength of a local alignment signal throughout the whole sample of N sources considering for every source their n_n nearest neighbours.

The significance level to reject the null hypothesis that a sample of sources is randomly oriented is then given by comparing the statistic of the dataset S_{n_n} to the distribution of the statistic for simulated randomly oriented samples. In essence, it is simply the probability that one would find the calculated value of S_{n_n} (or higher), when the position angles are drawn from a uniform distribution. In mathematical terms, the significance level is found through a one-tailed significance test, which is expressed as

$$SL = 1 - \Phi\left(\frac{S_{n_n} - \langle S_{n_n}|_{MC} \rangle}{\sigma_n}\right), \quad (2.42)$$

where Φ is the cumulative standard normal distribution function. $\langle S_{n_n}|_{MC} \rangle$ and σ_n are respectively the expected value and standard deviation of S_{n_n} , expected in absence of alignment. These values can be found through numerous Monte Carlo simulations of randomly oriented sources.

Jain, Narain, and Sarala (2004) verified that for randomly oriented samples of sources, S_{n_n} is normally distributed if $N \gg n_n \gg 1$ is satisfied. The standard deviation of S_{n_n} was found to follow $\sigma_S = 0.33/\sqrt{N}$ and the mean position of S_{nn} was stated to vary with $1/\sqrt{n_n}$.

With the dispersion measure and the resulting statistic, the significance level to reject uniformity can be calculated on both a local and global scale. A sample can be checked for global uniformity by setting $n_n = N$, and local uniformity by probing different numbers of nearest neighbours. Since the number of nearest neighbours that is probed can be translated to fixing apertures with angular radii extending to the n_n -th nearest neighbour around all sources, S_{n_n} can be used to probe alignment on different angular scales. If the redshift of the sources is known, this method can be extended to nearest neighbours in 3D space and S_{n_n} will probe alignment effects dependent on physical separation.

Chapter 3

Results

This chapter will explicate the results of both a global and local search for alignment in the initial sample of selected sources. The different subsets of the initial sample, defined in the previous chapter, are added to the analysis as well, to discover which sources are causing an effect, if one is found. The chapter will begin by testing global (i.e., over the full 424 square degrees) uniformity of the position angles to give an indication of possible systematics in the different subsamples. Thereafter, we test for local alignment of the radio lobes in projected 2D angular space. Finally, an equivalent statistical analysis in 3D space is carried out by selecting only the sources with redshift estimates. Throughout this chapter, the reader should keep in mind that we use the following approximate scale for the logarithmic significance level, as first adopted in Jain, Narain, and Sarala (2004). With these statistics, the absence of alignment is indicated by significance levels (SL) higher than 0.05, thus $\log SL$ larger than -1.3. We will state that weak evidence for alignment is found for $0.003 < SL < 0.5$, thus $-1.3 > \log SL > -2.5$. Strong evidence for alignment is indicated by $\log SL < -2.5$ and very strong evidence by $\log SL < -3.5$.

3.1 Global alignment: 2D

Initially, we will investigate whether the position angles are randomly distributed over the full 424 square degree field. The distribution of position angles in the initial sample and its subsequent subsets is given in Figure 3.1. The figure shows both the distribution of locally defined position angles and position angles parallel transported to a randomly picked source position in the samples. From this figure it becomes clear that the contribution due to parallel transport is small, which is to be expected since our data is confined to a relatively small part of the sky. A systematic effect in the connected lobes subset becomes apparent from this histogram, from the visible peak around 90 degrees. This peak is not present in the isolated lobes subset, which, in contrast, seems to show a small dip between around 70 and 115 degrees.

To test the null hypothesis that the position angles are randomly distributed over the whole sample of N sources, we calculate the statistic S_{n_n} (Equation 2.41) where we set the number of neighbouring sources $n_n = N$. This will calculate the statistic S_N , checking for alignment over the whole sample of N sources. To test whether S_N is significantly higher than the statistic for randomly distributed position angles, we compare the statistic S_N with the distribution of the value $S_N|_{MC}$ expected in absence of alignment, found through Monte Carlo sampling of uniformly distributed position angles. The significance level to reject the null hypothesis is then calculated from the one-tailed test given in Equation 2.42.

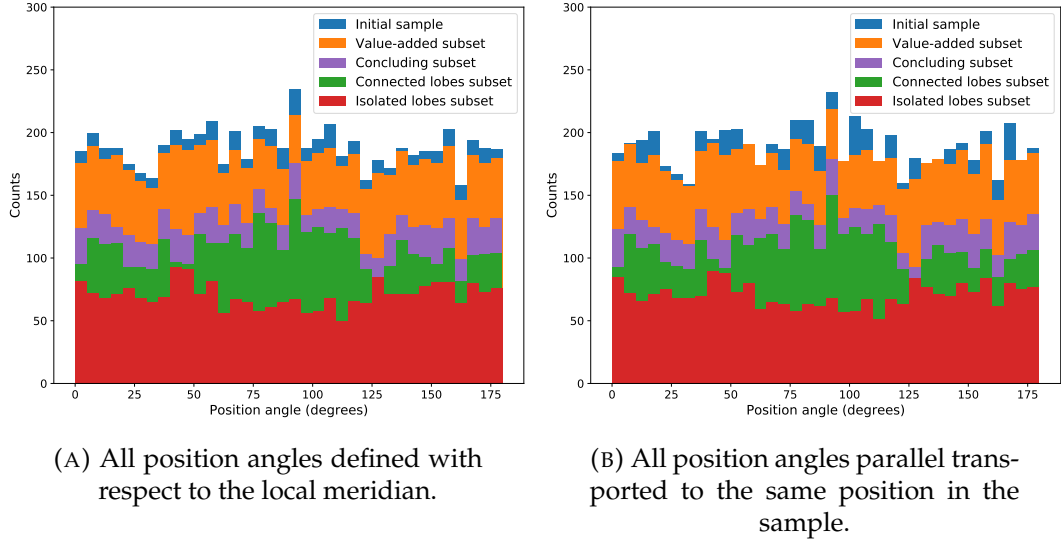


FIGURE 3.1: The distribution of position angles for all five previously defined samples, with and without parallel transport. A small peak around 90 degrees is present, originating from the connected lobes subset.

We calculate the distribution of $S_N|_{MC}$ by simulating 10,000 randomly oriented datasets of size N , generated by keeping the original source positions the same¹, but assigning each source a position angle from the uniform distribution $\mathcal{U}[0, 180)$. Since the calculations for 10,000 simulations containing the parallel transport of all $N = 6795$ sources N times to calculate $\langle S_N|_{MC} \rangle$ would be too computationally intensive, we choose to neglect parallel transport in these calculations.

The simulated distribution of $S_N|_{MC}$ is shown in Figure 3.2, where the mean $\langle S_N|_{MC} \rangle$ and standard deviation σ_S are found to be 0.011 and 0.0056 respectively. The statistic S_N corresponding to the initial sample of $N = 6795$ sources is found to be 0.020. The one-tailed significance test (Equation 2.42) with these values results in a significance level SL to reject the null hypothesis of 0.052, indicating that the null hypothesis cannot be rejected with 95% confidence.

It should, however, be reiterated that according to Jain, Narain, and Sarala (2004), the distribution of $S_N|_{MC}$ only follows a normal distribution for $n_n \ll N$. This can be visually verified from Figure 3.2, as the distribution is skewed rather than symmetrical. Due to this skewness, Equation 2.42 will overestimate the significance level when checking for global uniformity.

To investigate the overestimation due to the skewness of the distribution of $S_N|_{MC}$, we will apply the Kolmogorov-Smirnov (K-S) test to the data samples (e.g., Frommert, Durrer, and Michaud, 2012) with the same null-hypothesis of uniformity. The K-S test is better suited to check for uniformity of the whole sample as it compares the cumulative distribution function (CDF) inferred from Figure 3.2 to a uniform distribution of position angles, generated from the known equation of the uniform distribution as opposed to simulations. The K-S test finds a much larger p-value (significance level) of 18.3 per cent, indicating that we cannot reject the null hypothesis of a uniform distribution.

¹If we neglect parallel transport, the source positions do not influence the statistic when the number of nearest neighbours $n_n = N$, since this would calculate the dispersion over all sources, regardless of their position. Moreover, in this case, the statistic S_N is simply given by the maximum dispersion $d_{i,N}|_{max}$ of any source i since the maximum dispersion is then the same for all sources.

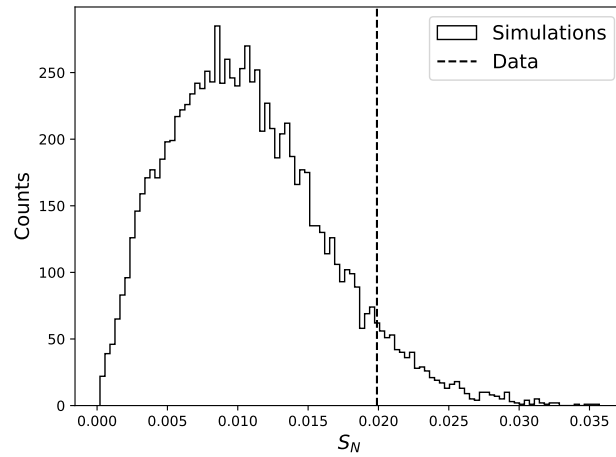


FIGURE 3.2: Distribution of the statistic S_N in 10,000 simulated datasets with $N = 6795$ sources. The calculated S_N for the initial sample is shown as the dashed line.

The equivalent results for the four subsets are presented in Table 3.1, from which it becomes clear that S_N indeed consistently overestimates the significance level due to $< S_N|_{MC} >$ following a skewed and not Gaussian distribution. However, the test still agrees trend-wise with the K-S test. According to the K-S test, the connected lobes subset shows weak evidence for a deviation from uniformity, which causes weak (1.69%) evidence for global alignment in the concluding subsample, as this subsample is largely made up of the connected lobes sources. For the other samples, at the 5% significance level we cannot reject the null hypothesis.

TABLE 3.1: Significance levels to find the current distribution of position angles in the subsamples if these were drawn from a uniform distribution. Calculated using the dispersion measure and the Kolmogorov-Smirnov (K-S) test.

Data sample	Source count	Random Sample		Data sample S_N	Significance level SL (per cent)	K-S test p -value (per cent)
		$< S_N _{MC} >$	σ_S			
Initial sample	6795	0.011	0.0056	0.020	5.20	18.3
Value-added subset	6433	0.011	0.0058	0.016	21.5	36.6
Connected lobes subset	3893	0.014	0.0074	0.054	4.58e-6	0.121
Isolated lobes subset	2540	0.018	0.0092	0.048	0.0445	6.69
Concluding subset	4655	0.013	0.0068	0.034	0.12	1.69

3.2 Local alignment: 2D

This section will present the results of the search for local alignment in 2D space. It will start by investigating local alignment in the initial sample and its different subsets. Thereafter, the initial sample is divided into four equally sized flux and size bins, which are also investigated for local alignment. For local alignment, due to the calculations increasing in complexity quickly with increasing n_n , we limit our search to $n_n = 800$. To determine the significance level to reject uniformity, the samples are compared with 1,000 simulated uniformly distributed position angle samples. In this section, these samples are generated by randomly shuffling the position angles amongst the sources, to maintain the same global position angle distribution and source positions. Unlike the previous section, here parallel transport is taken into account.

3.2.1 Full sample

The full sample can be checked for local alignment by probing the statistic S_{n_n} for different numbers of nearest neighbours. To express the statistic in terms of angular scale, a circular aperture with a radius extending to the n_n -th neighbour of every source can be drawn. By taking the median angular radius of all these apertures, the number of nearest neighbour can be translated to an approximate corresponding angular scale. For the initial sample and its subsets, this dependency is shown in Figure 3.3. As expected, the slope of this curve is dependent on the source density, showing the steepest slope for the sparsest sample, the isolated lobes, and the flattest slope for the initial sample. The significance level to reject the null hypothesis that the position angles are uniformly distributed locally is given as a function of nearest neighbours in Figure 3.4. For up to 800 nearest neighbours, the initial sample, value-added subset and isolated lobes show no significant deviation from uniformity. The connected lobes show very weak evidence to reject uniformity, with $\log SL$ barely reaching -1.3 around $n_n = 440$. Interestingly, the concluding subset seems to show slightly stronger, but still weak, evidence for alignment, with $\log SL$ reaching -1.69 at $n_n = 726$, while the other samples show no significant signal. For the concluding subset, as can be seen from Figure 3.3, $n_n = 726$ corresponds to a median angular radius of 5.42 degrees.

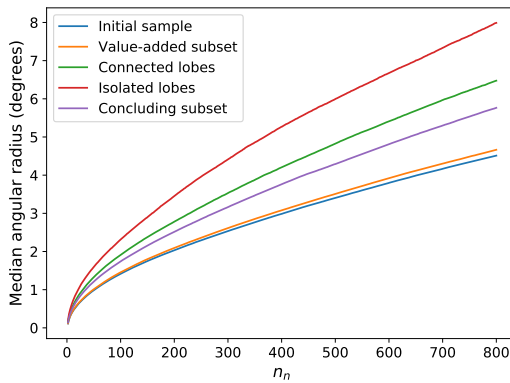


FIGURE 3.3: The median angular radius corresponding to drawing a circular aperture around every source with an angular radius bound by the n_n -th neighbour around that source. Shown for the initial sample and the different subsets.

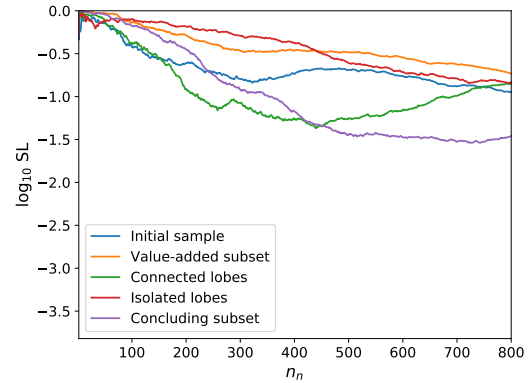


FIGURE 3.4: Logarithm of the significance level to reject uniformity as a function of the number of nearest neighbours n_n for the initial sample and the different subsets.

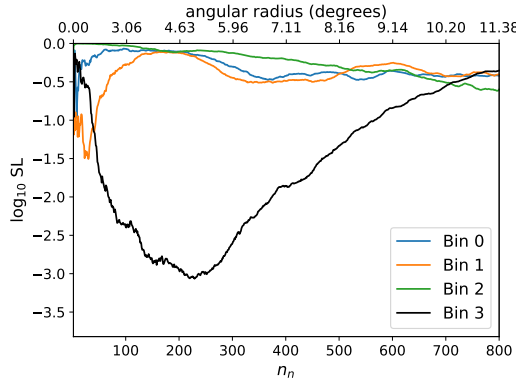


FIGURE 3.5: $\log SL$ as a function of n_n and the median angular radius around every source, for the different flux bins as defined in Table 3.2. The median angular radius is calculated for bin 3.

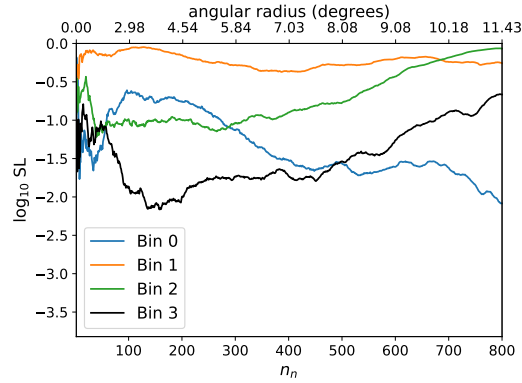


FIGURE 3.6: $\log SL$ as a function of n_n , for the different size bins as defined in Table 3.2. The median angular radius is calculated for bin 3.

TABLE 3.2: The flux and size thresholds for the bins that correspond to cutting the initial sample in four equally sized bins.

Bin number	Peak flux bins (mJy/beam)	Angular size bins (arc minute)
0	$0 < f < 1.13$	$0 < s < 0.41$
1	$1.13 < f < 2.61$	$0.41 < s < 0.55$
2	$2.61 < f < 9.41$	$0.55 < s < 0.74$
3	$9.41 < f$	$0.74 < s$

3.2.2 Flux and size bins

Cutting the initial sample on different observed characteristics may strengthen a signal that was previously obfuscated by bundling different types of sources. We choose to split the initial sample into flux and angular size bins, as these properties might crudely correspond to the redshift of the sources, if we assume most sources in our sample have similar properties. We choose four equal width bins to maintain the highest possible source count in every bin, thus keeping the sample size as large as possible. The flux cut is performed on the peak flux f of the source and the size cut on the observed angular size s of the source. Since our initial sample consists of two differently defined samples, isolated and connected lobes, the angular size and peak flux must be defined separately for these two samples. The peak flux and angular size for the connected lobes is simply defined as the observed peak flux and the major axis of the PyBDSF source respectively. For the isolated lobes, we take as the peak flux the largest of the two peak fluxes corresponding to the two matched lobes, and define the angular size as the great-circle distance between the lobes. The flux and size thresholds that correspond to splitting the initial sample into four equal width bins are given in Table 3.2. Every bin contains 1699 sources, except bin 3, which contains 1698.

Figures 3.5 and 3.6 show the logarithmic significance level as a function of the number of nearest neighbours probed or median angular radius of the corresponding apertures, for the four flux and size bins respectively. A strongly significant deviation from uniformity is found in the highest flux bin, with a significance level of

$\log SL = -3.07$ at an angular scale of 5.06 degrees. In the largest size bin weaker evidence for preferential alignment is found at a scale of 3.91 degrees, where $\log SL = -2.16$. It should be noted that this bin is defined by sources that have a size larger than $0.74'$, therefore this bin does not contain any isolated lobes, since these have been matched by setting a nearest neighbour distance cutoff of $0.69'$. The smallest sources also show a weak deviation from uniformity as n_n approaches 800. The logarithmic significance level is found to be -2.08 at 800 nearest neighbours, or 10.78 degrees.

3.2.3 Effect of subsets

To investigate the effects of the four different subsets on the signal of the highest flux and two size bins that show a local deviation from uniformity, we will apply the flux and size cuts as defined in Table 3.2 to the subsets defined in Section 2.2.4. The number of sources in every bin is then naturally not the size of the subset divided by four (as it was for the initial sample), as the flux and size distribution in every subset is different. The number of sources per subset satisfying the originally equal width flux or size cuts are given in Table 3.3. We will examine these separately per bin.

Size bins The smallest size bin is defined as all sources that have an angular size below 0.41 arc seconds. As can be viewed from Table 3.3, the number of sources in the smallest size bin varies by up to almost a factor of four depending on the subset. Thus, the median angular radius corresponding to a certain number of nearest neighbours will vary quite substantially between the different subsets. Therefore, we decide to plot the significance level only as a function of angular radius. As we calculate the significance level up to a set number of nearest neighbours $n_n=800$, the probed angular radii will vary depending on the subset. The logarithmic significance level for the smallest size bin of the four subsets is shown in Figure 3.7. The connected lobes subset seems to be the major cause of the alignment signal in the smallest size bin between angular radii of 5.5 and 7.3 degrees. The isolated lobes do show the same trend, but weaker. Interestingly, in contrast to the initial sample, where weak evidence for alignment was found on angular scales larger than 8 degrees, these subsets show stronger evidence on smaller scales. Another weak peak in significance is visible near 25 degrees, corresponding to the previously found global deviation from uniformity in Section 3.1.

The largest size bin, as stated before, contains all connected double lobed sources with a size larger than 0.74 arc minutes. In the initial sample, weak evidence for alignment was found on scales of about 4 degrees. Figure 3.8 corroborates this result by showing weak evidence at the same scale. As the three subsets are now identical, since no isolated lobes are contained in this bin, the small difference in $\log SL$

TABLE 3.3: Number of sources per subset for the given size and flux cuts.

Subset	Number of Sources		
	Angular size $< 0.41'$	Angular size $> 0.74'$	Peak flux > 9.41 mJy/beam
Value-added	1574	1612	1718
Connected lobes	411	1612	1549
Isolated lobes	1163	0	169
Concluding	876	1612	1662

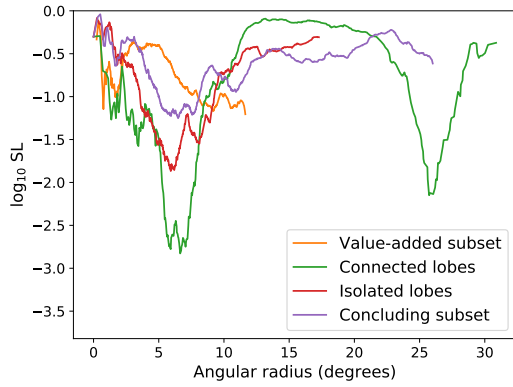


FIGURE 3.7: Log SL as a function of angular scale for the sources with size smaller than $0.41'$. The four subsets of the initial sample are shown.

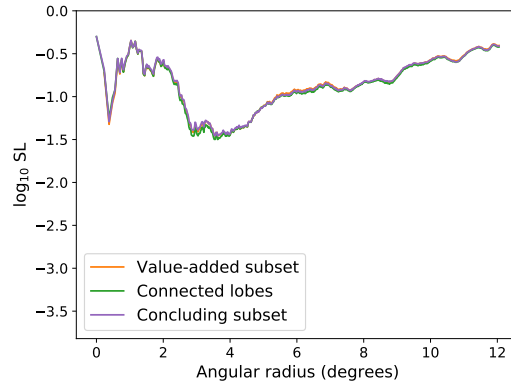


FIGURE 3.8: Log SL as a function of angular scale for the sources with size larger than $0.74'$. Three of the four subsets of the initial sample are identical in this case, as the isolated lobes subset contains no sources that satisfy this criterion.

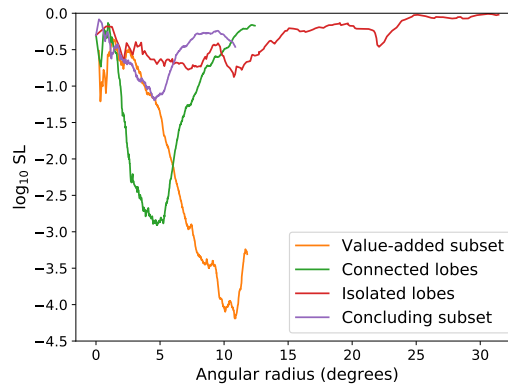


FIGURE 3.9: Log SL as a function of angular scale for the sources with peak flux brighter than 9.41 mJy/beam .

between the subsets is caused by the slightly different distributions of the simulated $S_{n_n}|_{MC}$ values.

Highest flux bin The largest flux bin is a mixture of mostly isolated lobes with only a few matched lobes. The strong departure from uniformity previously found at around 4 degrees is also found in the connected lobes subset, in Figure 3.9. This figure displays that the isolated lobes show no significant signal, but the combination of the isolated and connected lobes, the value-added subset, suddenly shows a much stronger signal at an angular scale of 10 degrees.

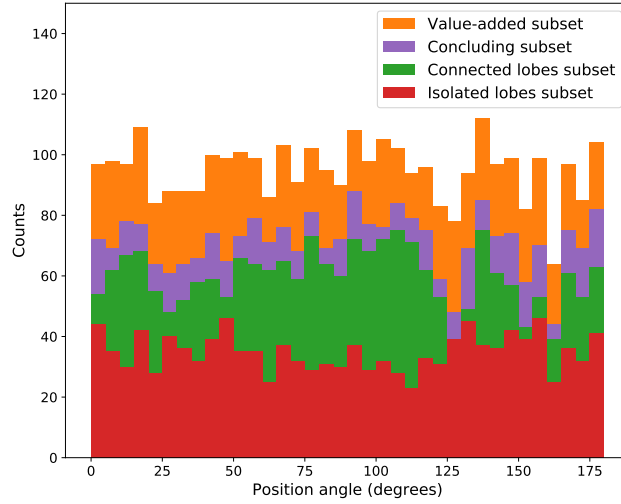


FIGURE 3.10: Distribution of position angles for the four subsamples, considering only the sources that have a redshift measurement.

3.3 Global alignment: 3D

For the 3D analysis, we cannot examine the initial sample, as it is not matched to the value-added catalogue. Therefore, a 3D analysis will only examine the four subsamples of the initial sample. We must also remove all sources that do not have a spectroscopic or photometric redshift in the value-added catalogue, reducing our source count in every subsample by about half. The new amount of sources in the subsamples for the 3D analysis are given in Table 3.4, which also shows the ratio of photometric to spectroscopic redshifts and the median redshift of the subsample. The table shows that roughly half of the redshifts are spectroscopically determined. Interestingly, the median redshift for the isolated lobes appears to be significantly lower than the median redshift for the connected lobes. The table also shows the significance level of the K-S test to reject uniformity. As becomes apparent, the connected lobe subset still shows a significant deviation from uniformity, albeit a weaker deviation than found in the 2D analysis. The other subsets do not disagree with global uniformity. The distribution of position angles is shown in Figure 3.10, which shows that the peak around 90 degrees in the connected lobes subset has disappeared. Therefore, the weak deviation from uniformity is likely associated with the significant drops at 125 and 165 degrees.

TABLE 3.4: The number of sources, number of photometric and spectroscopic redshifts and median redshift of the four subsets of the initial sample, considering only the sources that have a redshift measurement available. The p-value, calculated with the K-S test, to reject a uniform distribution of position angles is also displayed.

Data sample	Source count	Photo-z	Spectro-z	median-z	p-value (per cent)
Value-added subset	3412	2065	1347	0.57	42.2
Connected lobes	2155	1265	890	0.62	1.09
Isolated lobes	1257	800	457	0.46	25.0
Concluding subset	2564	1496	1068	0.60	25.9

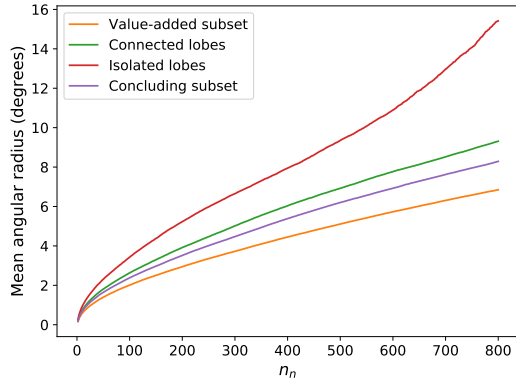


FIGURE 3.11: The median angular radius corresponding to a radius bound by the n_n -th neighbour around every source. Shown for the sources in the four subsets that have a redshift available.

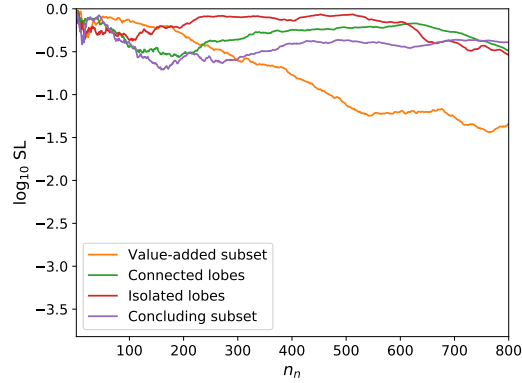


FIGURE 3.12: Logarithm of the significance level to reject uniformity as a function of the number of nearest neighbours n_n for the sources in the four subsets that have a redshift available.

3.4 Local alignment: 3D

This section will present the results of the search for local alignment in 3D space, analogous to Section 3.2. Again, parallel transport is taken into account and the samples are compared with 1,000 simulated uniformly distributed position angle samples, generated by randomly shuffling the position angles amongst the sources. To calculate the 3D source positions, the comoving distance of the sources is calculated assuming Planck 15 cosmology (Planck Collaboration et al., 2016). Each source is then assigned a position in 3D space according to their right ascension α , declination δ and comoving distance r as follows

$$\begin{aligned} x &= r \cos \alpha \cos \delta, \\ y &= r \sin \alpha \cos \delta, \\ z &= r \sin \delta. \end{aligned} \tag{3.1}$$

The nearest neighbours are then computed in 3D space according to these positions to probe for alignment on local scales.

3.4.1 Full samples

Figure 3.12 shows that the connected lobes, isolated lobes or concluding subset show no evidence for local alignment for the number of inspected nearest neighbours up to $n_n = 800$, corresponding to median angular radii above 7 degrees (Figure 3.11). In the value-added subset, a very weak ($\log SL = -1.3$) deviation from local uniformity is found at angular scales of 6 degrees.

3.4.2 Binned samples

We decide again to cut the largest sample, which is now the value-added subset, into four equally sized bins. However, since the redshift of the sources is now available, we can select on physical, rather than observed properties of the sources. Therefore, we will divide the value-added subset into four radio source power and physical size bins. The physical size is estimated from the observed angular size and measured

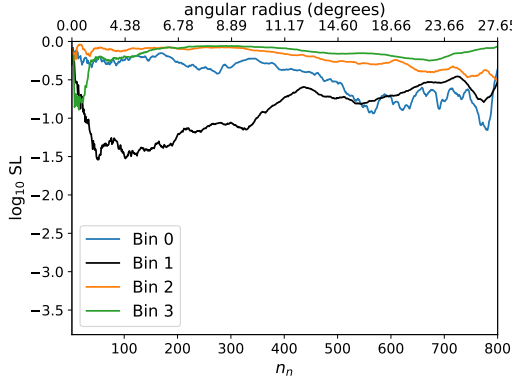


FIGURE 3.13: Log SL as a function of the number of nearest neighbours for the four power bins as defined in the text. The median angular radius is calculated for bin 1.

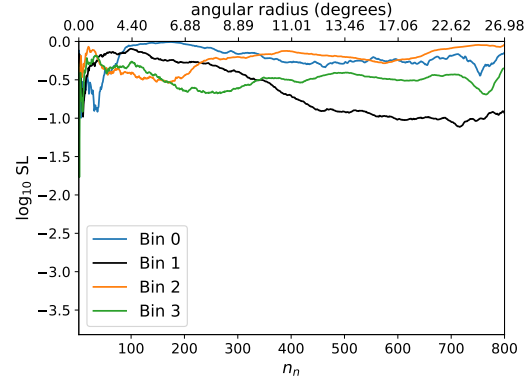


FIGURE 3.14: Log SL as a function of the number of nearest neighbours for the four physical size bins as defined in the text. The median angular radius is calculated for bin 1.

TABLE 3.5: The power and physical size thresholds for the bins that correspond to cutting the value-added subset in four equally sized bins.

Bin number	Power bins (W/Hz)	Size bins (kpc)
0	$0 < p < 3.84 \cdot 10^{24}$	$0 < s < 139$
1	$3.84 \cdot 10^{24} < p < 1.90 \cdot 10^{25}$	$139 < s < 221$
2	$1.90 \cdot 10^{25} < p < 1.00 \cdot 10^{26}$	$221 < s < 323$
3	$1.00 \cdot 10^{26} < p$	$323 < s$

redshift, assuming Planck 15 cosmological parameters (Planck Collaboration et al., 2016). The power at 150 MHz, $p_{150\text{MHz}}$, of a source with redshift z is calculated according to (e.g., Miley and De Breuck, 2008)

$$p_{150\text{MHz}} = 4\pi D_L^2 S_\nu (1+z)^{-\alpha-1}, \quad (3.2)$$

where D_L is the luminosity distance, S_ν the observed flux density at 150 MHz and α the assumed average spectral index. We assume $\alpha = -0.78$, in accordance with a previously found median spectral index for LOFAR 150 MHz observations in the Lockman Hole field (Mahony et al., 2016).

The parameters satisfying equal width bins are given in Table 3.5. Each bin contains 853 sources. With these bins, equivalently to the previous section, we plot the significance level both as a function of angular radius and as a function of nearest neighbours in Figures 3.13 and 3.14. Power bin 1 shows a weak deviation from uniformity at angular scales of 4.5 degrees, but other than this signal the subsets seem to be devoid of an alignment signal either in power or physical size bins. Therefore, we will only consider power bin 1 of the other subsets. The number of sources in power bin 1 for the different subsets are shown in Table 3.6. The significance level to reject uniformity in the other subsamples for this bin is plotted in 3.15, from which can be seen that the weak signal at 6 degrees is only present in the value-added subset, as a combined effect from the isolated and connected lobes.

TABLE 3.6: Number of sources per subset in the various bins that have a redshift measurement available. The bins are defined in Tables 3.2 and 3.5.

Subset	Number of sources		
	Power bin 1	Peak flux bin 3	Angular size bin 0
Value-added	853	947	760
Connected lobes	472	879	201
Isolated lobes	380	68	559
Concluding	618	923	443

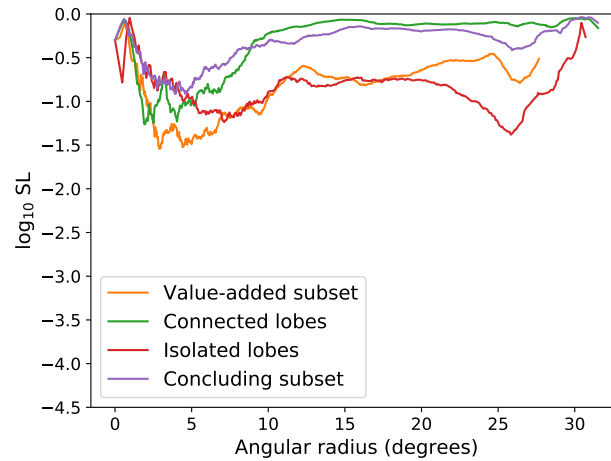


FIGURE 3.15: Log SL as a function of angular scale for the sources in power bin 1 as defined in Table 3.5.

Flux bins As the strong signal previously found in the highest peak flux bin at a scale of 4 degrees for the sample of connected lobes or 10 degrees for the value-added subset is not present anymore in any of the, until now considered, subsets in 3D space, we decide to plot $\log SL$ as in Figure 3.9, in 2D space, but only considering the sources that have a redshift measurement available. The number of sources satisfying the flux cutoff and having a redshift available are given in Table 3.6 for the different subsets. Figure 3.16 shows the logarithmic significance level to reject uniformity for the different subsets, when only considering sources that have a redshift measurement available. Its behaviour is similar to that of Figure 3.9, with strong evidence for alignment between the connected lobes on scales of 4 degrees, and strong evidence for alignment in the value-added subset on scales above 10 degrees. However, when we assign the same sources 3D positions according to their measured redshift, any signal completely disappears, and no significant evidence is found to reject local uniformity on any angular scale up to 25 degrees (Figure 3.17).

The final test is to review whether the result depends on the method of determining the redshift (i.e., spectroscopic or photometric). As photometric redshifts usually come with larger uncertainties than spectroscopic redshifts, these might misguide the nearest neighbour statistics by putting sources that should be close to each other further away. Therefore, we reproduce Figures 3.16 and 3.17, but only consider the sources that have a spectroscopic redshift available, and the two subsets that have previously shown a signal: the value-added subset and the connected lobes. This reduces the number of sources in these subsets by roughly half, from 947 and 879

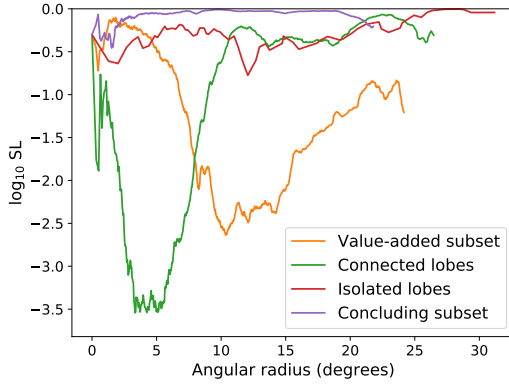


FIGURE 3.16: Equivalent to figure 3.9, but only considering sources that have a redshift measurement available.

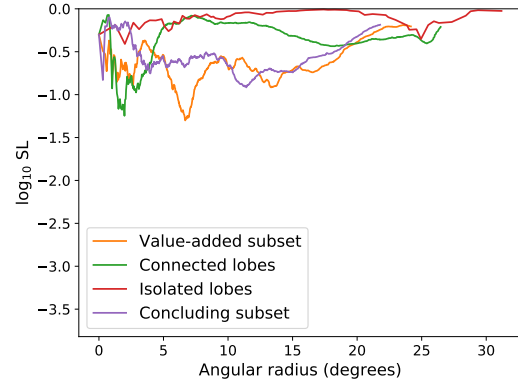


FIGURE 3.17: Equivalent to figure 3.16, but using the 3D positions of the sources as opposed to the projected 2D positions.

to 410 and 378 respectively. The significance level is plotted in Figure 3.18, which shows that in both the 2D analysis and 3D analysis of these subsets no significant signal is present.

Angular size bins Equivalently to the signal present in the flux bins, the strong signal previously found in the connected lobes subset for sources smaller than $0.41''$ seems to disappear when looking at the physical rather than angular size of sources. Therefore, in like manner, we consider from the connected lobes subset only the sources that have a redshift measurement available, and plot the significance level to reject uniformity calculated in 2D and in 3D space, to see if the added redshift information strengthens the previously found signal. The number of sources satisfying the angular size cutoff and having a redshift available are given in Table 3.6 for the different subsets. The significance level for the subsets is shown in Figure 3.19, from which becomes apparent that the signal previously found in Figure 3.7 is still present when reducing this subset to only the sources with a redshift measurement, but the signal disappears when investigating sources in 3D space. Similarly to the highest flux bin, no strongly significant signal is present when selecting only the 201 sources with spectroscopic redshift.

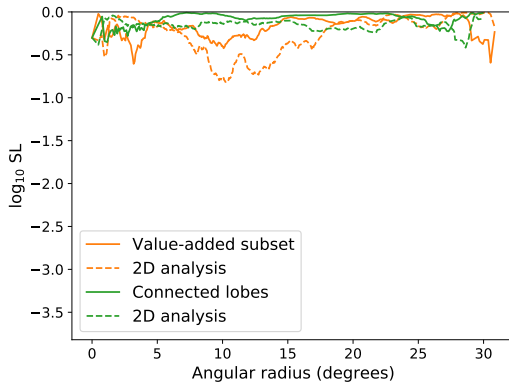


FIGURE 3.18: The logarithmic significance level to reject uniformity as a function of angular radius for the sources that have a spectroscopic redshift measurement and a peak flux higher than 9.41 mJy/beam . The dashed lines indicate a 2D analysis and the solid lines a 3D analysis.

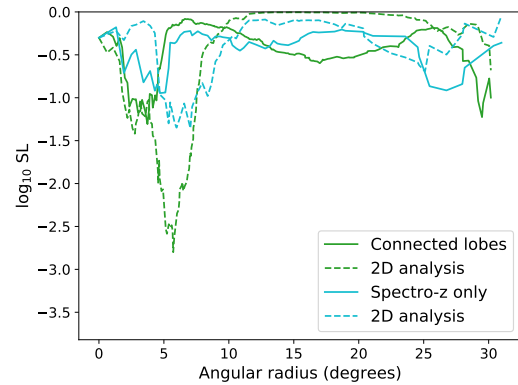


FIGURE 3.19: The logarithmic significance level as a function of angular radius for the connected lobes with angular size smaller than $0.41''$, considering only sources that have a redshift measurement available. The dashed line indicates a 2D analysis and the solid line a 3D analysis.

Chapter 4

Discussion

This chapter will discuss the results of Chapter 3. To interpret these results correctly, we must first discuss the fidelity of the results, which will be encompassed in Section 4.1. Thereafter, the results are interpreted in Section 4.2, which also includes the comparison with previous work and possible explanations for the results. Finally, suggestions for future work are given in Section 4.3.

4.1 Fidelity of the results

The fidelity of the results depends mainly on the uncertainties of the position angles that were fit to the sources. Therefore, we shall first discuss the algorithm employed to fit position angles and consequently discuss the propagation of these uncertainties in the statistical methods.

4.1.1 Fitted position angles

In this study we have implemented a method to fit the position angles of both isolated radio lobes and connected radio lobes. This method was necessary as the source finding algorithm PyBDSF only contained information about the position angle of connected radio lobes. The implemented algorithm to fit position angles has some weaknesses however, the most obvious one being that the algorithm does not not rely on the total morphology of the source, but only considers a thin line along which the maximal mean flux of the source lies. Taking into account the total morphology of the source should be a more powerful method to determine the position angles.

To examine the fidelity of our algorithm, we can compare the position angles that the algorithm has found for the connected lobes to the corresponding position angles of the sources in the PyBDSF catalogue. Since the connected lobes are correctly identified by PyBDSF as one source, the position angles calculated by PyBDSF should be comparable to the position angles that we have calculated. PyBDSF does take into account the total morphology of the source, by calculating the major axis of the source from fitting multiple Gaussian components to the islands of emission present in the image. The position angle is then computed as the angle between the major axis and the local meridian, increasing positively from north due east. Thus, if the position angles this study has computed for the connected lobes are comparable to the position angles computed by PyBDSF, we can assume our algorithm works reliably.

The distribution for the PyBDSF calculated and in this study calculated position angles of the connected lobe sources is shown in Figure 4.1. The figure shows very good agreement between the position angles calculated by PyBDSF and the position angles calculated in this study, with both distributions showing a small peak around

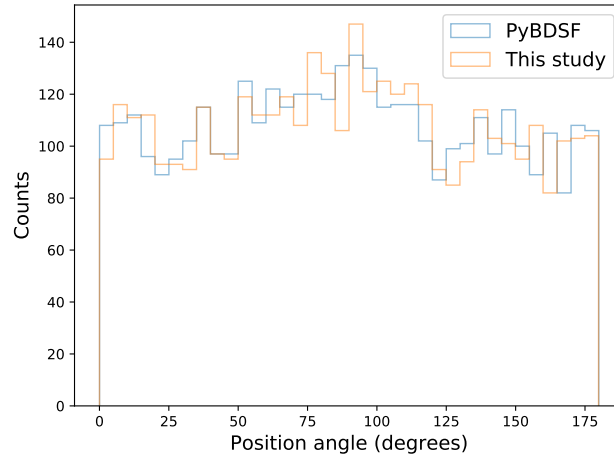


FIGURE 4.1: Distribution of calculated position angles by this study and the PyBDSF algorithm. Plotted for the subset of connected lobes.

90. On a source by source basis, we find that 81.4% of our calculated position angles agree within 5 degrees and 95.9% agree within 10 degrees with the PyBDSF position angles. Since the PyBDSF fitted position angles are not applicable for the isolated lobes that we have matched on proximity, we cannot repeat this procedure for the position angles of the matched lobes. However, matched lobes are arguably easier to fit position angles to, as the flux along the line between the sources is substantially higher when the line is oriented to cover both sources. Since the algorithm seems to agree well with PyBDSF fitted position angles for the arguably harder to fit sources, we can assume the position angle fits of the matched lobes are satisfactory as well.

4.1.2 Error propagation

This section will discuss the effect of random errors only, as possible systematics in the data are discussed along with the interpretation of the results in Section 4.2. The random error in the fitted position angles of this study should not be a large influence in the final results, as long as this error remains small enough such that the general direction of the sources is not completely changed. As this study tries to find alignment as a significant preference in the orientation of neighbouring sources, we could assume that an error of, for example, 10 degrees in the position angle would not drastically influence our statistics, since a preference for a certain direction, if one is found, would still persist. To investigate this assumption, we will determine the approximate error on the fitted position angles our algorithm has made and the effect of this error on the finally resulting significance level to reject uniformity.

In consideration of the agreement between the position angles defined by this study and the PyBDSF position angles, we can approximate the error we have made on a position angle fit by examining the PyBDSF 1σ errors on the position angles of the major axis of the connected lobes. Figure 4.2 shows the distribution of the 1σ errors, from which we can see that the large majority (80%) of sources have an error of less than 5 degrees in position angle. Thus, if we assume every source in our sample has a $1\sigma = \pm 5$ degree error we will likely overestimate the uncertainty in the fitted position angles.

To approximate the error in the final significance level as a function of a 1σ error of 5 degrees in the position angle, we must propagate this error through the statistical analysis of Section 2.3. Unfortunately, there is no straightforward procedure to

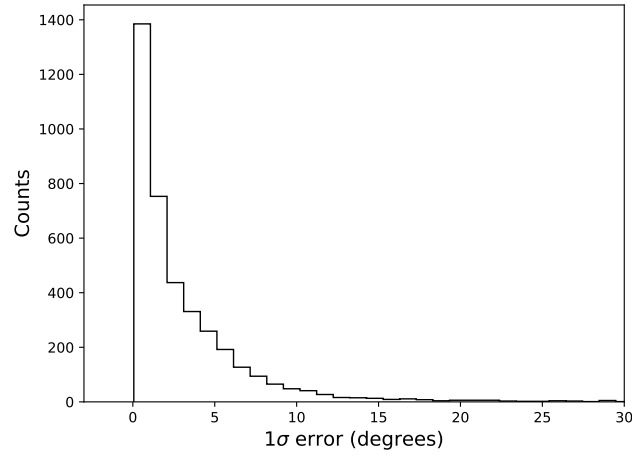


FIGURE 4.2: Distribution of the 1σ error on the position angle calculated by PyBDSF for the connected lobes subset.

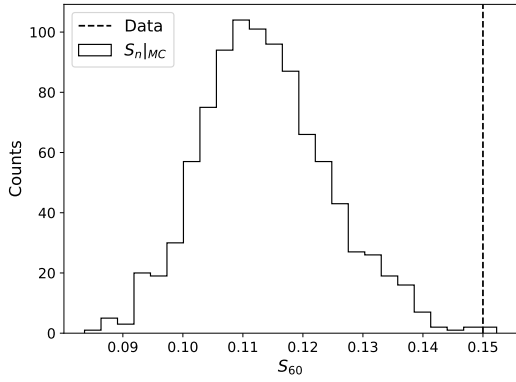


FIGURE 4.3: The distribution of the 1,000 simulated values $S_{n_n}|_{MC}$ and the highly significant value $S_{60} = 0.15$.

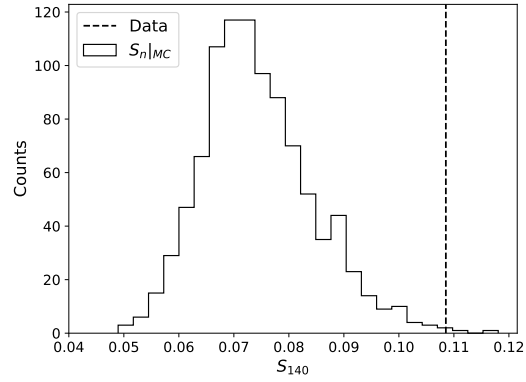


FIGURE 4.4: The distribution of the 1,000 simulated values $S_{n_n}|_{MC}$ and the highly significant value $S_{140} = 0.109$.

define the general error on the extracted significance level as a function of the error on the measured position angles. Simple error propagation can be applied up to the calculation of S_{n_n} , but becomes very complicated when a one-tailed significance level is extracted. This is due to the dependence of the significance level on the position of S_{n_n} in the distribution of $S_{n_n}|_{MC}$ (Equation 2.42). If S_{n_n} lies far away from the mean of the normal distribution, a certain change in S_{n_n} will lead to a smaller change in significance level than when S_{n_n} lies near the mean of the distribution of $S_{n_n}|_{MC}$. This is a direct effect of the cumulative normal distribution function being steepest near the mean and flattest near the edges. Moreover, considering that for every sample, $S_{n_n}|_{MC}$ is found by simulating 1,000 random datasets by randomly shuffling the position angles of the sources, the distribution of $S_{n_n}|_{MC}$ will be unique for every sample that we have considered. Therefore, we can only approximate the error on concrete subsample results and cannot give a general 1σ confidence level that will apply for different subsets.

The most interesting sample was found to be the connected lobes subsample with a peak flux threshold of 9.41 mJy/beam applied. When only considering the sources

that had a redshift measurement available, this sample rejected uniformity at a significance level of $\log SL = -3.5$ (Figure 3.16). The signal was found on an angular scale between 4 and 5 degrees, corresponding to a number of nearest neighbours in this sample between 60 and 140. Figures 4.3 and 4.4 show the distribution of the simulated data and the highly significant value of S_{n_n} for these two bounds. We will calculate the error on S_{60} , corresponding to $\log SL = -3.47$ and S_{140} , corresponding to $\log SL = -3.42$, and translate these errors to bounds on the significance values.

Assigning each position angle in this subsample a 1σ error of 5 degrees and applying standard error propagation, we find for the resulting values of S_{60} and S_{140} , 0.15 ± 0.0033 and 0.109 ± 0.0031 respectively. Taking the 1σ lower and upper bound of S_{60} and calculating the significance level of these two bounds results in the lower and upper bound logarithmic significance levels of -3.00 and -3.97 . For S_{140} the same method leads to lower and upper bound logarithmic significance levels of -2.96 and -3.92 . Thus, strong evidence to reject uniformity in this sample at scales between 4 and 5 degrees is still found after applying possible uncertainties in the position angles. However, the strength of this evidence can apparently change by a factor 10 within the assumed 1σ error bounds. We can conclude that, assuming a 1σ error of 5 degrees on the position angle of all sources, the effect of an uncertainty in the position angles is quite powerful, but the significance level does remain within the approximate bounds for strong evidence to reject uniformity adapted in Chapter 3. Thus, the resulting significance level of $\log SL = -3.5$ to reject uniformity found in Figure 3.16 should be stated with the approximate bound $\log SL = -3.5 \pm 0.5$.

As stated before, the difference in significance level for the same variation in S_{n_n} is dependent on the position of S_{n_n} , thus also dependent on the significance level itself. Therefore, we reiterate that the change of a factor 10 in significance, found for the subset considered in this section should not be applied to different subsets. We can apply the same calculation to the 3D analysis of the highest flux connected double lobed sources (Figure 3.17), where no result was found, and S_{60} is thus closer to the mean. This results in $\log SL = -0.807^{+0.172}_{-0.151}$. We also note that applying a standard deviation of 10 degrees to the position angles of the connected lobe sources in Figure 3.16 (where $\log SL = -3.5$) still does not shift strong evidence for alignment into weak evidence for alignment. Thus, the strong statistical evidence for alignment in the highest flux sources of the connected lobes perseveres after taking into account possible errors in the fitted position angles. Applying the same error propagation to the value-added subset in Figure 3.16 (where $\log SL = -2.5$) does change the statistical evidence from the strong to the weak regime.

4.2 Interpretation of the results

With the analysis of the error measure finished, we can interpret the results given in Chapter 3. Following the layout of the previous chapter, we begin by the global deviation from uniformity found in the connected lobes subset, which might indicate systematics in the data, and then treat the results of the search for local alignment.

4.2.1 Global alignment

The connected lobes subset was found to deviate from uniformity at a p-value of 0.121 per cent according to the Kolmogorov-Smirnov test. The visible peak at 90 degrees in the histogram of the position angles in Figure 3.2 might be the cause of this deviation. This peak indicated substantially more east-west oriented sources than

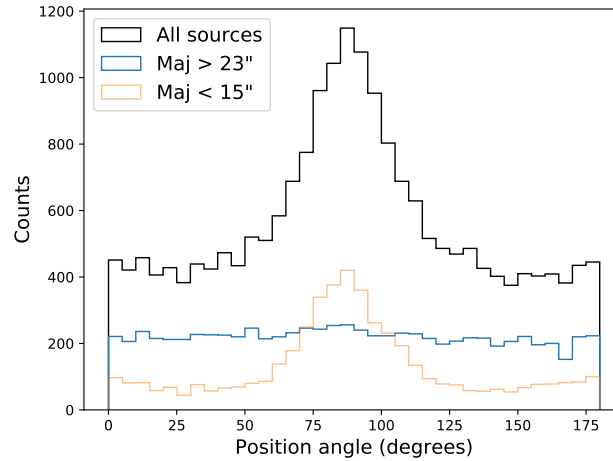


FIGURE 4.5: Distribution of PyBDSF fitted position angles to all 20,341 extended sources in the original data sample that have been fit by multiple Gaussian components. The subset of sources with a major axis smaller than 15" and the subset of sources with a major axis larger than 23" are also shown.

otherwise oriented sources. A global systematic, such as a point-spread-function anisotropy or the survey being more sensitive to sources in a certain direction, would be the most probable explanation, were it not for the fact that the isolated lobes seem to show a reversed effect, with a small decline of the number of position angles around 90 degrees. It should, however, be noted that this decline was not found to be a significant deviation from uniformity at the 5 per cent level. Nevertheless, if the peak around 90 degrees present in the connected lobes subset is caused by a systematic effect in the survey, we would expect to find this in the matched lobes subset as well. On the other hand, the fact that the peak was present strictly around 90 degrees and not another, more arbitrary angle does argue for a systematic effect. Thus, this indicates a systematic effect that might perhaps be native to only one of the subsets.

Since the peak is not present in the sources that have been matched by this study, but only in the sources that have been identified by PyBDSF as extended emission sources, the systematic effect may perhaps lie in the sensitivity of the source finder to east-west oriented sources. To test if the systematic is due to the initial source catalogue generated by PyBDSF, we can plot the distribution of PyBDSF position angles of all the sources that are fit by multiple Gaussian components (i.e., extended sources), before they have been propagated through any decision making process employed in this study. This distribution is shown in Figure 4.5, from which becomes clear that most PyBDSF sources have a position angle preferentially around 90 degrees. However, as the sources in the size subsets in Figure 4.5 show, this strong preference lies mainly in the small sized sources, since the sources with major axis larger than 23" only barely show this preference. Thus, while PyBDSF might have an innate preference to fit position angles of 90 degrees to small sources, the effect may yet lie in a global systematic unrelated to PyBDSF. This global systematic could still be the shape of the point spread function, which starts to become more important as the source size shrinks to sizes comparable to the beam size (6"). Since the peak around 90 degrees in position angles does seem to be dependent on the size of the source, this seems to be a plausible explanation, even if the matched lobes do not show this preference. Thus, the global systematic present in the connected lobes

might be due to the LOFAR primary beam in the HETDEX pointings perhaps being slightly more east-west oriented than otherwise.

As our final sample of selected sources is selected from the initial catalogue of apparently biased PyBDSF sources, some of the bias shown in Figure 4.5 is inevitably carried over to our final selected sample. The fact that Figure 3.2 does not show a peak which is as pronounced as the one in the sample of all PyBDSF extended emission sources, suggests that the decision tree, to select clear double lobed radio sources, employed in this study is able to filter out most of the biased sources that are preferentially east-west aligned. However, the deviation from uniformity still remains, showing that the source selection process can perhaps still be improved, or that the possible beam shape effect still influences our final sample.

Interestingly, Section 3.3 showed that, when selecting from the connected lobes subset only the sources that have an available redshift measurement, the distinct peak around 90 degrees disappeared (Figure 3.10). Another effect became present, however, being a significant dip around 125 degrees. The p-value to reject uniformity in the connected lobes subset was then found to be 1.09 per cent, indicating the sample still rejects the null hypothesis of a uniform distribution. Thus, selecting only the sources with available redshifts seems to clear some systematics from the connected lobes, but significant global effects still persist in this sample. However, in contrast to the peak around 90 degrees, we did not find an explanation for this anisotropy.

4.2.2 Local alignment

The analysis of local alignment depended on the contrast between the statistic S_{n_n} found for our dataset and the statistic $S_{n_n}|_{MC}$ found in absence of alignment. The statistic in absence of alignment was generated by randomly shuffling the position angles amongst the sources, to maintain the same geometry and global position angle distribution. The advantage of this method over generating position angles from the uniform distribution $\mathcal{U}[0, 180)$ is that it diminished the effect of the global systematic present in our data sample. This is due to the global systematic then also being included in the normal distribution of the statistic $S_{n_n}|_{MC}$. Therefore, as long as $n_n \ll N$, where N is the number of sources in the sample that is examined, the effect of the weak global deviation from uniformity will not have considerably impacted the result of the significance of local alignment.

First, the full sample was investigated for local alignment in angular space for up to $n_n = 800$ nearest neighbours, corresponding to an angular scale of roughly 4.5 degrees (Figure 3.3). No evidence was found to reject the null hypothesis of uniformity up to this angular scale in the initial sample, value-added subset or isolated lobes subset (Figure 3.4). The connected lobes and concluding subset were found to show weak evidence for alignment near 4.5 and 5.4 degree scales respectively. The fact that the concluding subset showed stronger evidence (log SL of -1.69 versus -1.3) than the connected lobes subset argues against the effect being caused by the systematics present in the connected lobes subset, since the concluding subset is the result of adding high-certainty matched lobes to the connected lobes subset. If this effect was caused by the systematics present in the connected lobes subset, adding unrelated matched lobes would diminish it. This evidence might be seen as the precursor for a stronger alignment effect that is obfuscated by stacking multiple unrelated sources.

To investigate this further, the sample was divided into four equal-width flux and four equal-width size bins, according to Table 3.2. Three effects were then found: the

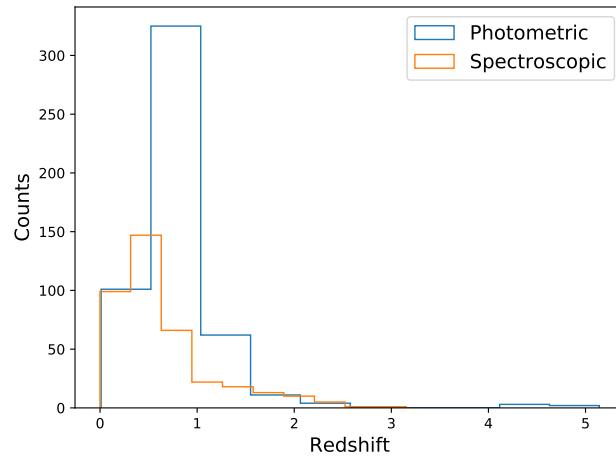


FIGURE 4.6: Redshift distribution in the connected lobes subset for the sources with a peak flux higher than 9.41 mJy/beam. Shown separately for the spectroscopic and photometric redshift calculations.

sources with peak flux higher than 9.41 mJy/beam showed a strong alignment signal between 3.5 and 6 degrees (Figure 3.5), the sources with an angular size smaller than 0.41' showed a weak deviation from uniformity at a scale of larger than 10.5 degrees (Figure 3.6) and the sources with an angular size larger than 0.74' showed weak evidence to reject uniformity at scales between 3 and 4 degrees (Figure 3.6).

The effect of the largest size sources was present on angular scales of about 4 degrees, so global systematics should not have significantly influenced this effect. However, the evidence for this effect was weak, and after investigating the subsets, the effect was barely significant when only considering the sources that were matched to the value-added catalogue (Figure 3.8). Thus we cannot, with high certainty, rule out the null hypothesis of randomly distributed position angles for the largest size sources.

The signal in the highest flux bin is the most interesting signal as it was found to be both a very significant signal, with $\log SL$ approaching -3.5 and present at angular scales where global alignment does not yet influence our statistics (Figure 3.5). When investigating this signal further, we saw that the connected lobes subset showed almost precisely the same behaviour as that of the initial sample, but oddly, when adding the matched lobes back into the statistics, the value-added subset showed very compelling evidence to reject uniformity at a different, larger angular radius (Figure 3.9). This difference might have been caused by the fact that the angular radius was not a good probe for the scale on which alignment was present. Since the two subsets have different redshift distributions, the different angular radii might probe similar physical scales. This is why the highest flux sources were also investigated in 3D space. Surprisingly, Figure 3.17 showed that any signal previously present in sources brighter than 9.41 mJy/beam had completely disappeared when applying 3D statistics. As roughly half of the measured redshifts are photometrically estimated, one might wonder whether this effect is due to the uncertainty in the photometric redshifts. To identify whether this is the case, we probed the significance level for only the sources in the value-added subset and connected lobes subset with a peak flux higher than 9.41 mJy/beam and a spectroscopic measurement of the redshift available. However, as Figure 3.18 showed, the significant signal had

disappeared now in the 2D analysis as well, leaving the problem unsolved. The disappearing of the signal might be a result of one or both of the following reasons. The source density ($\lesssim 1 \text{ deg}^{-2}$) might have dropped to low to warrant a significant signal or the dropping of preferentially higher redshift sources might have removed the alignment effect. By selecting only the spectroscopic redshift sources we are biased to lower redshift sources because the sources with an available spectroscopic redshift are generally lower redshift sources, as is shown in Figure 4.6.

The effect that the smallest size sources in the initial sample showed (Figure 3.6) could be attributed to be due to artefacts, as the value-added subset, containing nearly the same sources as the initial sample showed no such effect (Figure 3.7). Another argument for this is that the angular radius of 10.5 degrees at which this signal was present starts to approach the limit where the global alignment begins to become important again. However, in this same bin, the connected lobes subset did show a strong alignment signal at an angular scale of 5 degrees, corresponding to the scale where the highest flux sources showed alignment. This signal cannot be disregarded as it was found to be both strongly significant and present on a more credible angular scale. However, the fact that this signal is not present in the 3D analysis of this subset (Figure 3.19) argues for it being due to systematics in the dataset. Equivalently to the highest flux sources, a subset of these smallest size sources with measured spectroscopic redshift was separately investigated (Figure 3.19), but this was likewise found to be inconclusive as the strong signal had disappeared from both the 2D and 3D analyses.

In the 3D analysis of the full sample and the sample binned on the physical size and radio power at 150 MHz, no strong evidence for alignment was found. This is a surprising result, since we would expect that physical, rather than apparent, properties divide the sample into more meaningful subsets. The fact that no evidence was found in these subsets could be explained by the lower number of sources or uncertainty in the photometric redshifts. As the 3D analysis considers only roughly half of the sources that are considered in the 2D analysis, and roughly half of these sources have a photometric redshift estimate, this will evidently reduce the power of the 3D statistics. To perform a better 3D analysis of the sources that have been found to show strong alignment in the angular space, more (spectroscopic) redshift measurements are needed.

In conclusion, the fact that not only the connected lobes, but also the value added subset show evidence for local alignment, even after consideration of uncertainties in the position angles (although the value added subset then shows less strong evidence), when considering the highest flux sources in 2D analyses implies that the effect is real, and not caused by the systematic present in the connected lobes. However, the two sets do show a discrepancy in the angular radius at which the effect is present. This discrepancy might be explained by a 3D analysis, but unfortunately the effect disappears when considering the redshift of the sources. The fact that the effect disappears in the 3D analysis might be an argument for a bias in the dataset, but might also be explained by the uncertainties in the photometric redshifts. Unfortunately, the analysis including only the spectroscopic redshifts is found to be inconclusive as this indicates no evidence in both the 2D and 3D analyses. Unlike the highest flux sources, the fact that the smallest size sources only show an alignment effect in the connected lobes subset, which has previously found systematics, argues against this effect being real. These results should all be interpreted with caution, since the concluding subset, which was added to the analysis because it represented a sample with the least selection biases, did not show strong evidence for alignment in any of the investigated samples.

4.2.3 Comparison with previous work

The highest flux sources in the connected lobes subset, when analyzed in 2D space, have been shown to display convincing alignment effects on a scale of 5 degrees, even after taking into account the global systematics and possible uncertainties in the fitted position angles. The highest flux sources of the value-added subset have shown weaker, but still quite significant, alignment effects in 2D space as well. We cannot be certain yet whether the observed effect is real or a result of the particular biases of this study. To gain more insight into this, we will compare the results of this study to what previous studies of AGN alignments have found.

TABLE 4.1: Comparison of this work with previous studies. Shows the parameters of the previous studies that have found alignment effects, their significance level and the angular and physical scale of the effect assuming the median redshift of the study.

Name	Frequency	RMS (mJ/beam)	SNR	Source density (deg ⁻²)	Resolution (arcsec)	Study area (deg ²)	Significance (per cent)	Scale (deg)	Redshift (median)	Physical scale (h^{-1} Mpc)
This study ¹	150 MHz	0.071	10	2.1	6x6	424	<0.1	5	0.69	150
This study ²	150 MHz	0.071	10	2.2	6x6	424	<1.3	10	0.54	245
Contigiani ³	1.4 GHz	0.15	10	4.3	5x5	7000	<2	1.5	0.47	19-38
Taylor ⁴	612 MHz	0.01	3	53.3	6.1x5.1	1.2	<0.1	1	1	>20

¹ Only considering the connected lobes with peak flux >9.41 mJy/beam and redshift available.

² Only considering the value-added subset with peak flux >9.41 mJy/beam and redshift available.

³ Contigiani et al., 2017

⁴ Taylor and Jagannathan, 2016

⁵ Assumed redshift

Table 4.1 shows the parameters and results of this study and two previous studies of radio structure alignment (see Table for references). The physical scale corresponding to the angular scale on which alignment was found in this study is computed by assuming the median redshift of the sample. Converting the angular scale of 5 degrees to comoving distances using Planck 15 cosmological parameters (Planck Collaboration et al., 2016) yields a physical scale of $150h^{-1}\text{Mpc}$. It becomes immediately clear from Table 4.1 that the angular and physical scale of an observed alignment effect are both substantially larger than the results in the previous studies. Taylor and Jagannathan (2016) investigated an area of 1.2 square degrees, and were thus limited to finding alignment within this area. Therefore the angular scale of 1 degree for an alignment effect might be underestimated and may still be in agreement with the results of this study. Contigiani et al. (2017), however, did not contain this limitation, as they studied an area of 7000 square degrees and found an effect on scales smaller than 2.5 degrees, with the maximum alignment signal at 1.5 degrees. While the scale of the maximum effect does not agree with the angular scale of larger than 5 degrees found in this study, Contigiani et al. (2017) limited their search to angular scales below 2.5 degrees, so the signal may perhaps be present on larger scales in the FIRST survey as well.

Although the physical scale of $150\text{--}250h^{-1}\text{Mpc}$ found in this study does not agree with the physical scale of the two discussed previous studies of radio lobe alignment, it is in agreement with physical scales where other studies have found AGN alignment effects. As stated in Section 1.3.2, several studies have found that the radio polarization of quasars is preferentially aligned either perpendicular or parallel to the major axis of the surrounding large scale large quasar groups (LQGs). These effects range from distances of the order of 150 Mpc (Tiwari and Jain, 2013) to distances larger than $300h^{-1}\text{Mpc}$ (Pelgrims and Hutsemékers, 2016). The physical scales found in this study agree with the physical scale for alignment with large scale structure, while still abiding by the upper limits of homogeneity of the Universe, found to be of the order of $260h^{-1}\text{Mpc}$ by Yadav, Bagla, and Khandai (2010).

Considering the fact that the discovered signal seems to be dependent on the flux density of sources, it might be meaningful to compare the depth of the FIRST data (Becker, White, and Helfand, 1995) at 1.4 GHz where alignment was found in the Contigiani et al., 2017 study to our peak flux threshold of 9.41 mJy/beam at 150 MHz. However, it should be noted that this comparison is very crude for two reasons. The first reason is that the two studies have different resolutions (beam sizes), and the second reason is that the flux limit for a detection in the two surveys is dependent on the size of the source (i.e., the number of beams it covers), thus this crude comparison will only hold for the sources that cover a single beam (i.e., point sources) in both surveys. However, within this very crude approximation, we can translate the employed flux limit for point sources in the FIRST survey (1.5 mJy; Table 4.1) to the expected flux at 150 MHz. Assuming a spectral index of -0.78 (Mahony et al., 2016) We find that the lower bound of a point source flux $S_{1400\text{MHz}} = 1.5\text{mJy}$ translates to a flux of 8.6 mJy at 150 MHz. While this approximation is very rudimentary, it shows very rough agreement in the flux cut that is employed here and the flux threshold of the FIRST survey. This might explain why Contigiani et al., 2017 did not have to employ a flux cutoff to observe an alignment signal. A similar comparison cannot be made with the study by Taylor and Jagannathan (2016), as the SNR of 3, given in Table 4.1, is defined only as the threshold for a source detection. Visual selection of extended radio lobes was applied afterwards, and the authors did not state the flux density distribution of their final sample.

4.2.4 Possible explanations

While it is difficult to determine whether our findings are a result of the biases present in the dataset or indicate actual alignment of the sources, we can attempt to find an explanation for the results, assuming they are unrelated to present systematics. The fact that the concluding subset, which contains fairly high certainty matched lobes, as these have been independently matched by two separate methods, shows no strong alignment is rather odd. However, one can never be completely sure that all the matched isolated lobes are actually correctly matched, as there is simply no emission connecting these lobes, and they have not all been matched to an optical counterpart. The connected lobes, in contrary, do not suffer these uncertainties, since, other than identifying the emission with PyBDSF, no additional methods have been employed to find these. Thus, an explanation for the fact that the connected lobes show the strongest signal might be that these isolated lobes still contain too much incorrectly matched lobes, which blurs the alignment signal. A comparable explanation might be employed for the fact that there is only a (convincing) signal when considering the brightest sources. Sources are, in general, simply more easily identified as the brightness of the source increases. This might cause the sample of brightest sources to contain the least amount of noise. It is possible that the sources identified near the flux limit of the LOFAR survey are either more often mistaken as double-lobed sources in the selection process, or are harder to fit reliable position angles to. The addition of low-flux incorrectly picked (or fitted) double lobes might then blur an alignment signal.

We can also consider possible physical explanations for the alignment of radio jets, in case the results found in this study are verified by further studies. If the radio structures are indeed aligned over scales of $> 100h^{-1}\text{Mpc}$, an explanation may be found in the alignment of the active galactic nuclei (AGN), that power these radio structures, with the large scale structure. Evidence for the preferential orientation of AGN to their hosting structures was found in the alignment of optical polarization of

quasars and the surrounding large quasar groups (Hutsemékers et al., 2014). Since the radio structure of AGN is found to be correlated to the optical polarization of AGN (Antonucci, 1983; Lister, 2001; Joshi et al., 2007), our results might corroborate this finding, if the alignment effect is found to follow the orientation of large scale structures in the observed area of the sky.

If the AGN are indeed aligned with respect to the large scale structure they reside in, a possible cause could be the angular momentum transfer from matter as it is gravitationally collapsing during the early stages of galaxy formation. The tidal torques imparted on the collapsing halos is found to influence the spin and shape of galaxies in N-body simulations (e.g., White, 1984; Codis et al., 2012; Laigle et al., 2015). This supercluster scale mechanism may perhaps also impact the spin of the central supermassive black hole of these galaxies, resulting in AGN alignment. Another argument for this hypothesis is that the physical scale found in this study agrees quite well with the peak in the galaxy correlation function at $100h^{-1}\text{Mpc}$ predicted by the concordance model of cosmology and observed in the Sloan Digital Sky Survey (Eisenstein et al., 2005). However, the angular momentum of the active galactic nucleus and the host galaxy are found to be misaligned and generally uncorrelated (Hopkins et al., 2012), indicating that this explanation is still incomplete.

Another explanation might lie in a different effect of large scale structures on AGN. Radio-loud AGN are believed to be the result of the accretion of cold matter on the supermassive black hole, and an obvious way to accomplish this is by merging with gas-rich systems (Hardcastle, Evans, and Croston, 2007). Therefore, radio-loud AGN are believed to be associated with mergers, which is confirmed by both observations and simulations (e.g., Chiaberge et al., 2015; Croton et al., 2006). Perhaps if these mergers occur preferentially along the filaments of the large scale structure, these could orientate the central supermassive black holes in a particular way, resulting in alignment of the radio jets.

4.3 Suggestion for future work

In the statistical studies of AGN alignments, the selection method is arguably the most important part of the study. The construction of the initial sample of sources that are used to probe alignment effects is inevitably biased by the selection method employed in the study. The consequences of selection effects on the final results were carefully examined in this research, by defining multiple subsets and checking the sample for global alignment. However, this study still employs a single approach to select sources, so further studies of the same area with different selection methods could corroborate the results independently of the selection method.

Further studies should also be performed with deeper radio data as the significance of an alignment effect is found to perhaps depend on the radio source density of the sample, since removing photometric redshifts also removed the signal. This is corroborated by Contigiani et al. (2017), who found that the radio source density is an important factor in determining whether significant alignment effects can be detected. A strong alignment signal in their sample of sources disappeared when reducing the source density to 1.5 deg^{-2} by means of randomly eliminating sources. The number density of sources in the 2D analysis was comparable or higher than the number density of the sample in which Contigiani et al. (2017) previously found alignment, even after splitting the data into four equal width bins, so this should not be a problem in this study. However, for the 3D analysis, the number of sources had to be reduced by about half, since redshift information was not available for

all the sources. This led to the number density of sources dropping below that of Contigiani et al. (2017) when splitting the data into four equal width bins, and below the threshold of 1.5 deg^{-2} when only considering the highest flux sources that have spectroscopic redshift measurements available. To perform a better 3D analysis, a more complete set of redshift estimates is required, which will be provided by the upcoming WEAVE-LOFAR survey (Smith et al., 2016). In the coming years, the WEAVE-LOFAR survey will obtain more than one million optical spectra of low frequency radio sources in the LOFAR 150 MHz survey, producing spectroscopic redshifts for an unprecedented number of radio sources. The study of radio jet alignments can then be tested thoroughly in a 3D analysis, without the uncertainties of photometric redshifts or the concern for the sparsity of the sample.

For a concrete recommendation, we suggest further research to verify whether the observed effect of alignment in the sources with peak flux higher than 9.41 mJy/beam is a real effect. To verify this, spectroscopic follow-up is needed for at least some of the sources that showed an alignment effect in 2D space, as the effect had disappeared in the 2D analysis when reducing the 879 connected lobes that had redshift information available to the 378 sources that had spectroscopic redshifts available.

To identify which sources in particular should be investigated, we can examine which sources show the strongest alignment effect in 2D space. For this we can use the calculated maximum dispersion $d_{i,n_n}|_{\max}$ (Equation 2.40), which measured the significance of the alignment of a source i and its n_n nearest neighbours. We plot the maximum dispersion for every source in the sample of 879 connected lobes as a function of right ascension for $n_n = 60$ and $n_n = 140$ in Figures 4.7 and 4.8. In Figure 4.7, two clear peaks are present between 190 and 200, and between 215 and 255 degrees right ascension. The latter peak is also present in Figure 4.8. From these figures we impose the following thresholds for strong alignment: $d_{i,60} > 0.25$ and $d_{i,140} > 0.17$. Now, we can identify which sources are significantly aligned in angular space, as is shown in Figures 4.9 and 4.10. These figures both indicate that the largest contribution to the observed alignment effect is caused by the region approximately bounded by $210^\circ < ra < 230^\circ$ and $45^\circ < dec < 50^\circ$.

The region that shows the strongest alignment effect contains 89 sources, of which 40 sources have a photometric redshift, as is shown in Figure 4.11. The figure also indicates that sources in this area seem to be more preferentially east-west aligned than otherwise. Therefore, should this region be examined further, the global preference for 90 degree position angles should carefully be taken into account. In conclusion, we suggest follow-up on this region of the sky, specifically spectroscopic follow-up on the 40 double-lobed sources that, as of now, only have photometric redshifts available. It would be interesting to see if the alignment signal in a 3D analysis is then strengthened rather than weakened and if perhaps the orientation of the sources in this region is traced by a known large-scale filament.

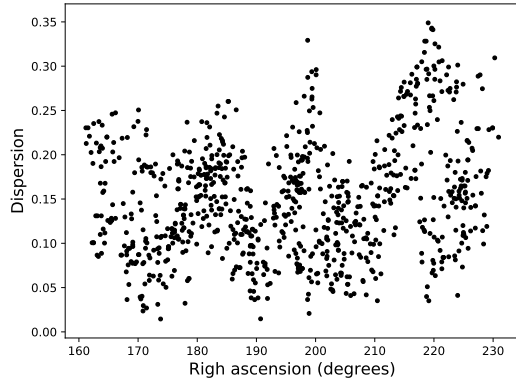


FIGURE 4.7: Scatter plot of the maximum dispersion d_{i,n_n} of the 879 high flux connected lobes. Plotted here for $n_n = 60$ as a function of right ascension.

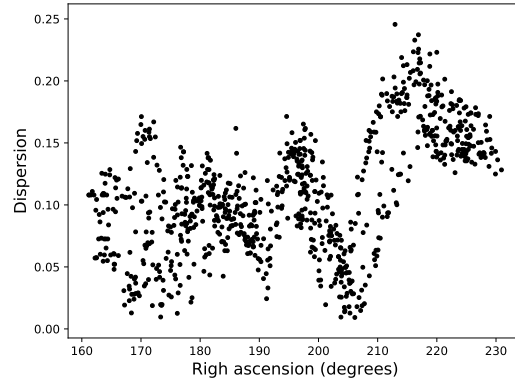


FIGURE 4.8: Scatter plot of the maximum dispersion $d_{i,140}$ as a function of right ascension of the 879 high flux connected lobes.

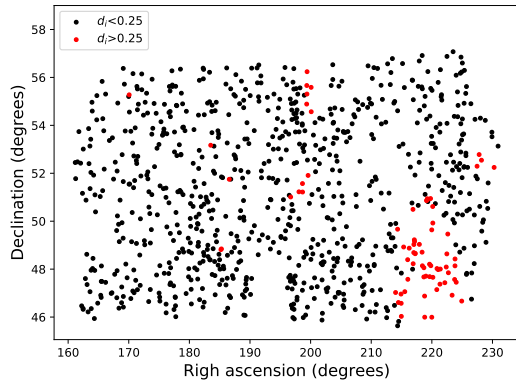


FIGURE 4.9: Scatter plot of the dispersion of the 879 high flux connected lobes plotted for $n_n = 60$ as a function of right ascension and declination.

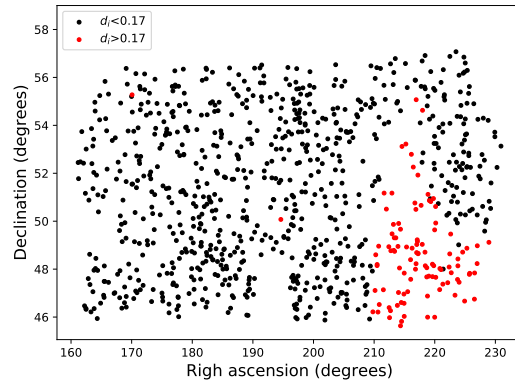


FIGURE 4.10: Scatter plot of the dispersion of the 879 high flux connected lobes. Plotted here for $n_n = 140$.

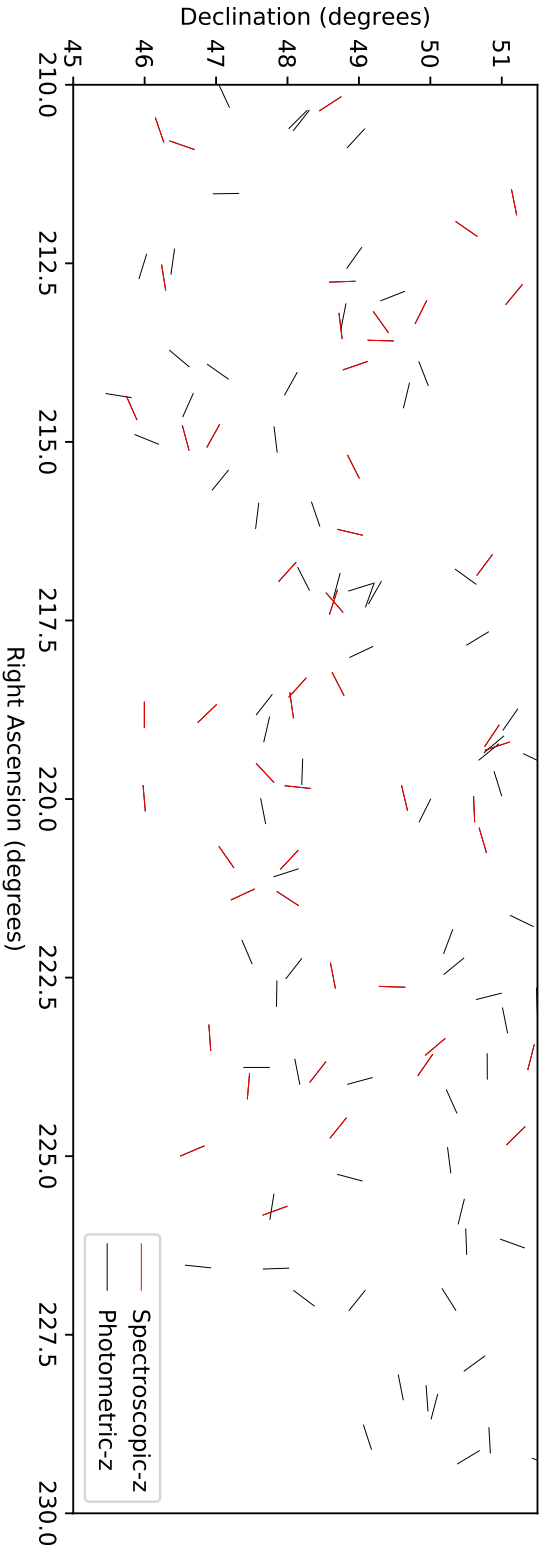


FIGURE 4.11: Position angles and locations of the sources in the LoTSS sample that are found to show the strongest alignment signal. Sources that have a measured spectroscopic redshift are indicated in red. The other sources have photometric redshifts available.

Chapter 5

Conclusion

In this thesis, we have searched thoroughly for alignment of radio galaxies in the HETDEX Spring Field region by making use of the unprecedented radio source density and large number of available redshift measurements in the first data-release of the LOFAR Two-metre Sky Survey. We have implemented a novel technique to find double-lobed sources in an automated way from the 325,694 sources initially present in the dataset. This technique approached isolated and connected radio lobes separately, matching isolated radio lobes on proximity and morphology, and analyzing connected radio lobes on morphology alone. A decision tree was employed to separate clear double-lobes from contaminants and a final sample of 6795 sources was formed. To test for alignment of position angles in this sample, the spherical nature of position angles and the effect of transporting these angles over the celestial sphere were carefully taken into account using statistics originally developed to test for alignment in polarization vectors. To be able to identify the effect of different selection methods or possible errors in different selection methods, the final sample was merged with a value-added catalogue which contained likelihood ratio matching or visual matching of isolated lobes, and redshift information of roughly half of the sources. This allowed for the division of the final sample into four subsets, defined keeping the purpose of investigating various selection effects in mind.

The final sample, together with its four subsets, was initially extensively investigated in a two-dimensional analysis. We found that the connected lobes showed a preference to be east-west oriented over the full 424 square degree sample. This preference was shown to be potentially related to the shape of the primary beam in the LOFAR pointings, as it was found to be a function of source size. Taking this preference into account in the tests for local alignment, the connected lobes with a peak flux higher than 9.41 mJy/beam were found to have the most promising alignment signal with a p -value < 0.001 to reject uniformity at an angular scale of 5 degrees. We suggest that the alignment signal apparently being a function of flux might be due to the selection method performing poorer on low flux sources.

Roughly half of the sources in the final sample had a redshift measurement available, either photometric or spectroscopic. This allowed us to perform the same analysis in three-dimensional space. No strongly significant evidence for deviation from uniformity was found in three-dimensional space, in either the full sample or the sample divided into numerous bins, including the highest flux bin. We argue that this is not due to the reduced number of sources, since a 2D analysis of the connected lobes with peak flux higher than 9.41 mJy/beam when considering only the sources with available redshift still showed a strongly significant alignment signal on an angular scale between 4 and 5 degrees. After verifying the fidelity of the fitted position angles, and (unlike previous studies) taking into consideration possible uncertainties in the fitted position angles, the p -value of this strongly significant alignment signal in 2D was found to lie in the approximate bounds $0.001 < p < 0.0001$. We find

the corresponding comoving scale for this effect to be of the order of $150 h^{-1}$ Mpc, assuming the median redshift of the sample $z = 0.69$.

The discrepancy between the two- and three-dimensional analysis might argue for the effect being due to systematics, but we suggest that the discrepancy might also be caused by the uncertainty in the photometric redshifts. Tests including only the spectroscopic redshifts are found to be inconclusive, thus further research is required to validate this. We have identified the region in which alignment is most pronounced as the region approximately bounded by $210^\circ < ra < 230^\circ$ and $45^\circ < dec < 50^\circ$ and thus suggest further studies to examine this region. The sources in this region appear to show a preferential east-west orientation, aligning with the established systematics present in the dataset. Therefore, this effect should be cautiously interpreted.

The observed comoving scale of $150 h^{-1}$ Mpc where significant alignment is present agrees well with previous studies that have found that the orientation of the active galactic nuclei that power these radio jets might be influenced by the large scale structure surrounding these objects. Other than agreeing with the previous studies, the observed scale agrees also agrees well with previously found upper limits on the scale of homogeneity of the Universe. Therefore, we proposed two physical models that might explain the observed alignment effect. These two models assumed that the large scale structure surrounding active galactic nuclei directly or indirectly effects the orientation of the active galactic nuclei, resulting in observed alignment of radio jets.

In the coming years, as the LOFAR Two-metre Sky Survey observes the full northern sky, and the WEAVE-LOFAR survey obtains more than one million optical spectra of low frequency radio sources, the amount of high quality low-frequency data will increase exponentially. We hope that future studies of radio jet alignments, making use of these impressive efforts, will be able to verify whether the observed alignment effect in this study is real.

Bibliography

- Abolfathi, B. et al. (2018). "The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment". In: *Astrophysical Journal, Supplement* 235, 42, p. 42.
- Antonucci, R. R. J. (1983). "Optical polarization position angle versus radio structure axis in Seyfert galaxies". In: *Nature* 303, p. 158.
- Aragon-Calvo, M. A. and L. F. Yang (2014). "The hierarchical nature of the spin alignment of dark matter haloes in filaments". In: *Monthly Notices of the RAS* 440, pp. L46–L50.
- Aragón-Calvo, M. A. et al. (2007). "Spin Alignment of Dark Matter Halos in Filaments and Walls". In: *Astrophysical Journal, Letters* 655, pp. L5–L8.
- Baugh, C. M. et al. (2004). "The 2dF Galaxy Redshift Survey: hierarchical galaxy clustering". In: *Monthly Notices of the RAS* 351, pp. L44–L48.
- Becker, R. H., R. L. White, and D. J. Helfand (1995). "The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters". In: *Astrophysical Journal* 450, p. 559.
- Benson, A. J. et al. (2001). "The clustering evolution of the galaxy distribution". In: *Monthly Notices of the RAS* 327, pp. 1041–1056.
- Best, P. N. and T. M. Heckman (2012). "On the fundamental dichotomy in the local radio-AGN population: accretion, evolution and host galaxy properties". In: *Monthly Notices of the RAS* 421, pp. 1569–1582.
- Bharadwaj, S., S. P. Bhavsar, and J. V. Sheth (2004). "The Size of the Longest Filaments in the Universe". In: *Astrophysical Journal* 606, pp. 25–31.
- Bietenholz, M. F. and P. P. Kronberg (1984). "Is there really evidence for universal rotation?" In: *Astrophysical Journal, Letters* 287, p. L1.
- Birch, P. (1982). "Is the universe rotating". In: *Nature* 298, pp. 451–454.
- (1983). "Is there evidence for universal rotation? (reply)". In: *Nature* 301, p. 736.
- Bond, J. R., L. Kofman, and D. Pogosyan (1996). "How filaments of galaxies are woven into the cosmic web". In: *Nature* 380, pp. 603–606.
- Borguet, B. et al. (2008). "Evidence of a type 1/type 2 dichotomy in the correlation between quasar optical polarization and host-galaxy/extended emission position angles". In: *Astronomy and Astrophysics* 478, pp. 321–333.
- (2011). "New Insights into the Quasar Type 1/Type 2 Dichotomy from Correlations between Quasar Host Orientation and Polarization". In: *Astronomical Polarimetry 2008: Science from Small to Large Telescopes*. Ed. by P. Bastien et al. Vol. 449. Astronomical Society of the Pacific Conference Series, p. 459.
- Cameron, M. J. (1971). "The structure of bright galaxies at 408 MHz". In: *Monthly Notices of the RAS* 152, p. 439.
- Cen, R. and J. P. Ostriker (1999). "Where Are the Baryons?" In: *Astrophysical Journal* 514, pp. 1–6.
- Chambers, K. C. et al. (2016). "The Pan-STARRS1 Surveys". In: *ArXiv e-prints*. arXiv: [1612.05560](https://arxiv.org/abs/1612.05560) (astro-ph.IM).
- Chiaberge, M. et al. (2015). "Radio Loud AGNs are Mergers". In: *Astrophysical Journal* 806, 147, p. 147.

- Codis, S. et al. (2012). "Connecting the cosmic web to the spin of dark haloes: implications for galaxy formation". In: *Monthly Notices of the RAS* 427, pp. 3320–3336.
- Coil, A. L. (2013). "The Large-Scale Structure of the Universe". In: *Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology*. Ed. by T. D. Oswalt and W. C. Keel, p. 387.
- Contigiani, O. et al. (2017). "Radio Galaxy Zoo: cosmological alignment of radio sources". In: *Monthly Notices of the RAS* 472, pp. 636–646.
- Croton, D. J. et al. (2006). "The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies". In: *Monthly Notices of the RAS* 365, pp. 11–28.
- Davé, R. et al. (2001). "Baryons in the Warm-Hot Intergalactic Medium". In: *Astrophysical Journal* 552, pp. 473–483.
- Duncan, K. J. et al. (submitted). "The LOFAR Two-metre Sky Survey IV . First Data Release : Photometric redshifts and rest-frame magnitudes". In: *Astronomy and Astrophysics*.
- Einasto, M. et al. (2014). "Tracing a high redshift cosmic web with quasar systems". In: *Astronomy and Astrophysics* 568, A46, A46.
- Eisenstein, D. J. et al. (2005). "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". In: *Astrophysical Journal* 633, pp. 560–574.
- Fanaroff, B. L. and J. M. Riley (1974). "The morphology of extragalactic radio sources of high and low luminosity". In: *Monthly Notices of the RAS* 167, 31P–36P.
- Fang, T., G. L. Bryan, and C. R. Canizares (2002). "Simulating the X-Ray Forest". In: *Astrophysical Journal* 564, pp. 604–623.
- Frommert, M., R. Durrer, and J. Michaud (2012). "The Kolmogorov-Smirnov test for the CMB". In: *Journal of Cosmology and Astroparticle Physics* 1, 009, p. 009.
- Ganeshaiah Veena, P. et al. (2018). "The Cosmic Ballet: spin and shape alignments of haloes in the cosmic web". In: *ArXiv e-prints*. arXiv: [1805.00033](https://arxiv.org/abs/1805.00033).
- Gaztañaga, E. et al. (2005). "Statistical analysis of galaxy surveys - II. The three-point galaxy correlation function measured from the 2dFGRS". In: *Monthly Notices of the RAS* 364, pp. 620–634.
- González, R. E. et al. (2017). "Relation between halo spin and cosmic-web filaments at $z = 3$ ". In: *Monthly Notices of the RAS* 464, pp. 4666–4672.
- Hardcastle, M. J., D. A. Evans, and J. H. Croston (2007). "Hot and cold gas accretion and feedback in radio-loud active galaxies". In: *Monthly Notices of the RAS* 376, pp. 1849–1856.
- Hirv, A. et al. (2017). "Alignment of galaxies relative to their local environment in SDSS-DR8". In: *Astronomy and Astrophysics* 599, A31, A31.
- Hopkins, P. F. et al. (2012). "Why are active galactic nuclei and host galaxies misaligned?" In: *Monthly Notices of the RAS* 425, pp. 1121–1128.
- Hutsemekers, D. (1998). "Evidence for very large-scale coherent orientations of quasar polarization vectors". In: *Astronomy and Astrophysics* 332, pp. 410–428.
- Hutsemekers, D. and H. Lamy (2001). "Confirmation of the existence of coherent orientations of quasar polarization vectors on cosmological scales". In: *Astronomy and Astrophysics* 367, pp. 381–387.
- Hutsemekers, D. et al. (2005). "Mapping extreme-scale alignments of quasar polarization vectors". In: *Astronomy and Astrophysics* 441, pp. 915–930.
- Hutsemekers, D. et al. (2014). "Alignment of quasar polarizations with large-scale structures". In: *Astronomy and Astrophysics* 572, A18, A18.
- Intema, H. T. et al. (2017). "The GMRT 150 MHz all-sky radio survey. First alternative data release TGSS ADR1". In: *Astronomy and Astrophysics* 598, A78, A78.

- Jain, P., G. Narain, and S. Sarala (2004). "Large-scale alignment of optical polarizations from distant QSOs using coordinate-invariant statistics". In: *Monthly Notices of the RAS* 347, pp. 394–402.
- Jain, Pankaj, Gaurav Narain, and S. Sarala (2004). "Large-scale alignment of optical polarizations from distant QSOs using coordinate-invariant statistics". In: *Monthly Notices of the Royal Astronomical Society* 347.2, pp. 394–402.
- Jones, B. J. T., R. van de Weygaert, and M. A. Aragón-Calvo (2010). "Fossil evidence for spin alignment of Sloan Digital Sky Survey galaxies in filaments". In: *Monthly Notices of the RAS* 408, pp. 897–918.
- Joshi, S. A. et al. (2007). "A survey of polarization in the JVAS/CLASS flat-spectrum radio source surveys - II. A search for aligned radio polarizations". In: *Monthly Notices of the RAS* 380, pp. 162–174.
- Joshi, S. A. et al. (2007). "A survey of polarization in the JVAS/CLASS flat-spectrum radio source surveys - II. A search for aligned radio polarizations". In: *Monthly Notices of the Royal Astronomical Society* 380.1, pp. 162–174.
- Kendall, D. G. and G. A. Young (1984). "Indirectional statistics and the significance of an asymmetry discovered by Birch". In: *Monthly Notices of the RAS* 207, pp. 637–647.
- Kimball, A. E. and Ž. Ivezić (2008). "A Unified Catalog of Radio Objects Detected by NVSS, First, WENSS, GB6, and SDSS". In: *Astronomical Journal* 136, pp. 684–712.
- Laigle, C. et al. (2015). "Swirling around filaments: are large-scale structure vortices spinning up dark haloes?" In: *Monthly Notices of the RAS* 446, pp. 2744–2759.
- Lee, J. and U.-L. Pen (2000). "Cosmic Shear from Galaxy Spins". In: *Astrophysical Journal, Letters* 532, pp. L5–L8.
- Libeskind, N. I. et al. (2013). "The velocity shear tensor: tracer of halo alignment". In: *Monthly Notices of the RAS* 428, pp. 2489–2499.
- Libeskind, N. I. et al. (2018). "Tracing the cosmic web". In: *Monthly Notices of the RAS* 473, pp. 1195–1217.
- Lister, M. L. (2001). "Parsec-Scale Jet Polarization Properties of a Complete Sample of Active Galactic Nuclei at 43 GHz". In: *Astrophysical Journal* 562, pp. 208–232.
- Mahony, E. K. et al. (2016). "The Lockman Hole project: LOFAR observations and spectral index properties of low-frequency radio sources". In: *Monthly Notices of the RAS* 463, pp. 2997–3020.
- Massardi, M. et al. (2011). "The Australia Telescope 20 GHz (AT20G) Survey: analysis of the extragalactic source sample". In: *Monthly Notices of the RAS* 412, pp. 318–330.
- Miley, G. (1980). "The structure of extended extragalactic radio sources". In: *Annual Review of Astron and Astrophys* 18, pp. 165–218.
- Miley, G. and C. De Breuck (2008). "Distant radio galaxies and their environments". In: *Astronomy and Astrophysics Reviews* 15, pp. 67–144.
- Miley, G. K. et al. (1972). "Active Galaxies with Radio Trails in Clusters". In: *Nature* 237, pp. 269–272.
- Mohan, N. and D. Rafferty (2015). *PyBDSF: Python Blob Detection and Source Finder*. Astrophysics Source Code Library.
- Padovani, P. (1997). "Unified schemes for radio-loud AGNs: recent results". In: *Memorie della Societa Astronomia Italiana* 68, p. 47. eprint: [astro-ph/9701074](https://arxiv.org/abs/astro-ph/9701074).
- (2017). "On the two main classes of active galactic nuclei". In: *Nature Astronomy* 1, 0194, p. 0194.
- Padovani, P. et al. (2015). "Radio-faint AGN: a tale of two populations". In: *Monthly Notices of the RAS* 452, pp. 1263–1279.

- Padovani, P. et al. (2017). "Active galactic nuclei: what's in a name?" In: *Astronomy and Astrophysics Reviews* 25, 2, p. 2.
- Pahwa, I. et al. (2016). "The alignment of galaxy spin with the shear field in observations". In: *Monthly Notices of the RAS* 457, pp. 695–703.
- Payez, A., J. R. Cudell, and D. Hutsemékers (2008). "Axions and polarisation of quasars". In: *American Institute of Physics Conference Series*. Ed. by J. Cugon, J.-P. Lansberg, and N. Matagne. Vol. 1038. American Institute of Physics Conference Series, pp. 211–219.
- (2011). "Can axionlike particles explain the alignments of the polarizations of light from quasars?" In: *Physical Review D* 84.8, 085029, p. 085029.
- Payez, A., D. Hutsemékers, and J. R. Cudell (2010). "Large-scale coherent orientations of quasar polarisation vectors: interpretation in terms of axion-like particles". In: *American Institute of Physics Conference Series*. Ed. by D. B. Tanner and K. A. van Bibber. Vol. 1274. American Institute of Physics Conference Series, pp. 144–149.
- Peacock, J. A. et al. (2001). "A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey". In: *Nature* 410, pp. 169–173. eprint: [astro-ph/0103143](#).
- Peebles, P. J. E. (1969). "Origin of the Angular Momentum of Galaxies". In: *Astrophysical Journal* 155, p. 393.
- Pelgrims, V. and D. Hutsemékers (2015). "Polarization alignments of quasars from the JVAS/CLASS 8.4-GHz surveys". In: *Monthly Notices of the RAS* 450, pp. 4161–4173.
- (2016). "Evidence for the alignment of quasar radio polarizations with large quasar group axes". In: *Astronomy and Astrophysics* 590, A53, A53.
- Phinney, E. S. and R. L. Webster (1983). "Is there evidence for universal rotation". In: *Nature* 301, p. 735.
- Planck Collaboration et al. (2016). "Planck 2015 results. XIII. Cosmological parameters". In: *Astronomy and Astrophysics* 594, A13, A13.
- Prandoni, I. et al. (2000). "The ATESP radio survey. II. The source catalogue". In: *Astronomy and Astrophysics, Supplement* 146, pp. 41–55.
- Primack, J. R. (1997). "Dark Matter and Structure Formation in the Universe". In: *ArXiv Astrophysics e-prints*. eprint: [astro-ph/9707285](#).
- Richter, G. A. (1975). "Search for optical identifications in the 5C3-radio survey. II - Statistical treatment and results". In: *Astronomische Nachrichten* 296, pp. 65–81.
- Ross, A. J., R. J. Brunner, and A. D. Myers (2007). "Higher Order Angular Galaxy Correlations in the SDSS: Redshift and Color Dependence of Nonlinear Bias". In: *Astrophysical Journal* 665, pp. 67–84.
- Rusk, R. (1988). "The Brightness and Polarization Structure of Compact Radio Sources". In: *The Impact of VLBI on Astrophysics and Geophysics*. Ed. by M. J. Reid and J. M. Moran. Vol. 129. IAU Symposium, p. 161.
- Rusk, R. and E. R. Seaquist (1985). "Alignment of radio and optical polarization with VLBI structure". In: *Astronomical Journal* 90, pp. 30–38.
- Sadler, E. M. et al. (2007). "Radio galaxies in the 2SLAQ Luminous Red Galaxy Survey - I. The evolution of low-power radio galaxies to $z \sim 0.7$ ". In: *Monthly Notices of the RAS* 381, pp. 211–227.
- Sandage, A. (1965). "The Existence of a Major New Constituent of the Universe: the Quasistellar Galaxies." In: *Astrophysical Journal* 141, p. 1560.
- Scott, D. and C. A. Tout (1989). "Nearest neighbour analysis of random distributions on a sphere". In: *Monthly Notices of the RAS* 241, pp. 109–117.

- Seldner, M. et al. (1977). "New reduction of the Lick catalog of galaxies". In: *Astronomical Journal* 82, pp. 249–256.
- Shimwell, T. W. et al. (submitted). "The LOFAR Two-metre Sky Survey II. First Data Release". In: *Astronomy and Astrophysics*.
- Smith, D. J. B. et al. (2016). "The WEAVE-LOFAR Survey". In: *SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*. Ed. by C. Reyl   et al., pp. 271–280.
- Smoot, G. F. et al. (1992). "Structure in the COBE differential microwave radiometer first-year maps". In: *Astrophysical Journal, Letters* 396, pp. L1–L5.
- Stockman, H. S., J. R. P. Angel, and G. K. Miley (1979). "Alignment of the optical polarization with the radio structure of QSOs". In: *Astrophysical Journal, Letters* 227, pp. L55–L58.
- Taylor, A. R. and P. Jagannathan (2016). "Alignments of radio galaxies in deep radio imaging of ELAIS N1". In: *Monthly Notices of the RAS* 459, pp. L36–L40.
- Tempel, E. and N. I. Libeskind (2013). "Galaxy Spin Alignment in Filaments and Sheets: Observational Evidence". In: *Astrophysical Journal, Letters* 775, L42, p. L42.
- Tempel, E., R. S. Stoica, and E. Saar (2013). "Evidence for spin alignment of spiral and elliptical/S0 galaxies in filaments". In: *Monthly Notices of the RAS* 428, pp. 1827–1836.
- Tiwari, P. and P. Jain (2013). "Polarization Alignment in Jvas/class Flat Spectrum Radio Surveys". In: *International Journal of Modern Physics D* 22, 1350089, p. 1350089.
- Torres, D. F. and L. A. Anchordoqui (2004). "Astrophysical origins of ultrahigh energy cosmic rays". In: *Reports on Progress in Physics* 67, pp. 1663–1730.
- Trowland, H. E., G. F. Lewis, and J. Bland-Hawthorn (2013). "The Cosmic History of the Spin of Dark Matter Halos within the Large-scale Structure". In: *Astrophysical Journal* 762, 72, p. 72.
- Urry, C. M. and P. Padovani (1995). "Unified Schemes for Radio-Loud Active Galactic Nuclei". In: *Publications of the Astronomical Society of the Pacific* 107, p. 803.
- van Velzen, S., H. Falcke, and E. K  rding (2015). "Nature and evolution of powerful radio galaxies at $z \sim 1$ and their link with the quasar luminosity function". In: *Monthly Notices of the RAS* 446, pp. 2985–3001.
- White, S. D. M. (1984). "Angular momentum growth in protogalaxies". In: *Astrophysical Journal* 286, pp. 38–41.
- Williams, W. L. et al. (submitted). "The LOFAR Two-metre Sky Survey III . First Data Release : optical identifications and value-added catalogue". In: *Astronomy and Astrophysics*.
- Willott, C. J. et al. (1999). "The emission line-radio correlation for radio sources using the 7C Redshift Survey". In: *Monthly Notices of the RAS* 309, pp. 1017–1033.
- Wilman, R. J. et al. (2008). "A semi-empirical simulation of the extragalactic radio continuum sky for next generation radio telescopes". In: *Monthly Notices of the RAS* 388, pp. 1335–1348.
- Wright, E. L. et al. (2010). "The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance". In: *Astronomical Journal* 140, pp. 1868–1881.
- Yadav, J. K., J. S. Bagla, and N. Khandai (2010). "Fractal dimension as a measure of the scale of homogeneity". In: *Monthly Notices of the RAS* 405, pp. 2009–2015.
- Zehavi, I. et al. (2002). "Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data". In: *Astrophysical Journal* 571, pp. 172–190.