Investigating Polarization Determination in LOFAR Lightning Observations

Bachelor's Project Astronomy

October 2024

Student: Mingcan Wang (s3547183)

First supervisor: Prof. Olaf Scholten

Second examiner: Prof. Mariano Mendez

Abstract

TRI-D imaging, a three-dimensional interferometric beamforming technique, is currently used in lightning research. In this thesis, we process signal data obtained through the Low Frequency Array (LOFAR) using TRI-D imaging, and analyze the reliability of the 3D polarization data generated by TRI-D interferometry. Our results show that the distribution of polarization directions deviates from the expected uniform distribution. To explore the factors influencing these imaging results, we modified the antenna function and calibration files, analyzing their effects across different locations. The analysis reveals significant variations in their effects for the different locations. Finally, we analyzed the relationship between the relative positions of the antennas and the polarization distribution characteristics of the background radiation. The results demonstrate a strong correlation between antenna placement and the deduced polarization data distribution.

Contents

1 Introduction

Lightning, a form of electrical discharge, is commonly believed to originate from interactions between ice particles and hail within clouds [1]. These interactions result in the transfer of positive and negative charges to lighter ice particles and heavier hailstones, respectively. This charge separation, driven by convective forces within the cloud, leads to the formation of a strong electric field, which ultimately culminates in a lightning discharge [1].

During the lightning discharge process, plasma channels known as leaders propagate within the thundercloud, forming high-temperature, conductive pathways. There are different types of leaders, including positive and negative leaders, each exhibiting distinct propagation characteristics [2]. Negative leaders propagate in a discrete, stepwise manner, referred to as "steps", which produce significant radio pulses in the 30–300 MHz frequency range [3, 4]. The scale of these steps is highly dependent on altitude. In regions far from the ground and beneath the cloud base, the time intervals between steps are approximately 50 microseconds [1]. At altitudes above 8km huge steps are observed covering distances of 300 m [5]. However, closer to the ground, both the step length and time intervals decrease, typically to about 10 meters and 10 microseconds, respectively [3, 6]. In addition, at lower altitudes, negative leaders develop more pronounced branching and increasingly complex pathways. Structurally, negative leaders consist of two main components: a highly conductive core with a radius of several centimeters and a less conductive corona sheath with a radius of approximately 10 meters [7]. The corona sheath emits a visible glow during discharge.

In contrast, positive leaders, which propagate positive charge, have a more continuous and smooth propagation path. Research has shown that positive leaders exhibit a less distinct step structure and produce minimal electromagnetic radiation [8]. Recent studies have identified plasma structures, referred to as "needles," which are typically perpendicular to the channel, as the primary sources of radio emissions from positive leaders [2, 9].

After the plasma channel is established, it gradually cools and becomes unstable, resulting in strong current pulses [10]. These pulses do not create new channels but instead travel rapidly along the pre-established positive leader channels. Recent studies have observed that the speeds of these leaders vary between 5×10^6 and 3×10^7 meters per second [7, 11], which approaches 10% of the speed of light. As they propagate, these pulses emit visible light and broadband radio waves, and the intensity of their very high-frequency (VHF) radiation decreases as their speed declines [7]. In some instances, these pulses travel far enough to enter negative leader channels. Depending on whether these channels connect with the ground, the leaders are classified as either dart leaders or dart-stepped leaders when they reach the ground, transferring large amounts of negative charge. If the leaders do not connect with the ground, they are referred to as recoil leaders/streamers or K-leaders, also known as retrograde leaders [11]. The propagation of these leaders can sometimes be smooth, but they may also display step structures [1]. Typically, their speed decreases during propagation, although occasional acceleration has been observed [7]. The mechanisms behind these variations in speed and behavior remain unclear.

Despite the fact that lightning is a common geophysical phenomenon, with approximately 9 million discharges occurring globally each day [1]. Many aspects of the physical mechanisms behind lightning remain poorly understood, as studying lightning presents numerous challenges.

First, although lightning is frequent, its occurrence in terms of location and time is random, and the relevant timescales for study range from nanoseconds to seconds [1]. Second, because lightning typically occurs within clouds, optical observations are often limited by cloud cover, making research difficult for an extended period. With advancements in observational techniques and resolution, more options have become available, such as very high frequency (VHF) imaging (30 to 300 MHz) [6]. Thanks to the excellent

penetrative ability of radio waves, VHF imaging is less affected by cloud cover compared to traditional optical methods, allowing for long-distance observation.
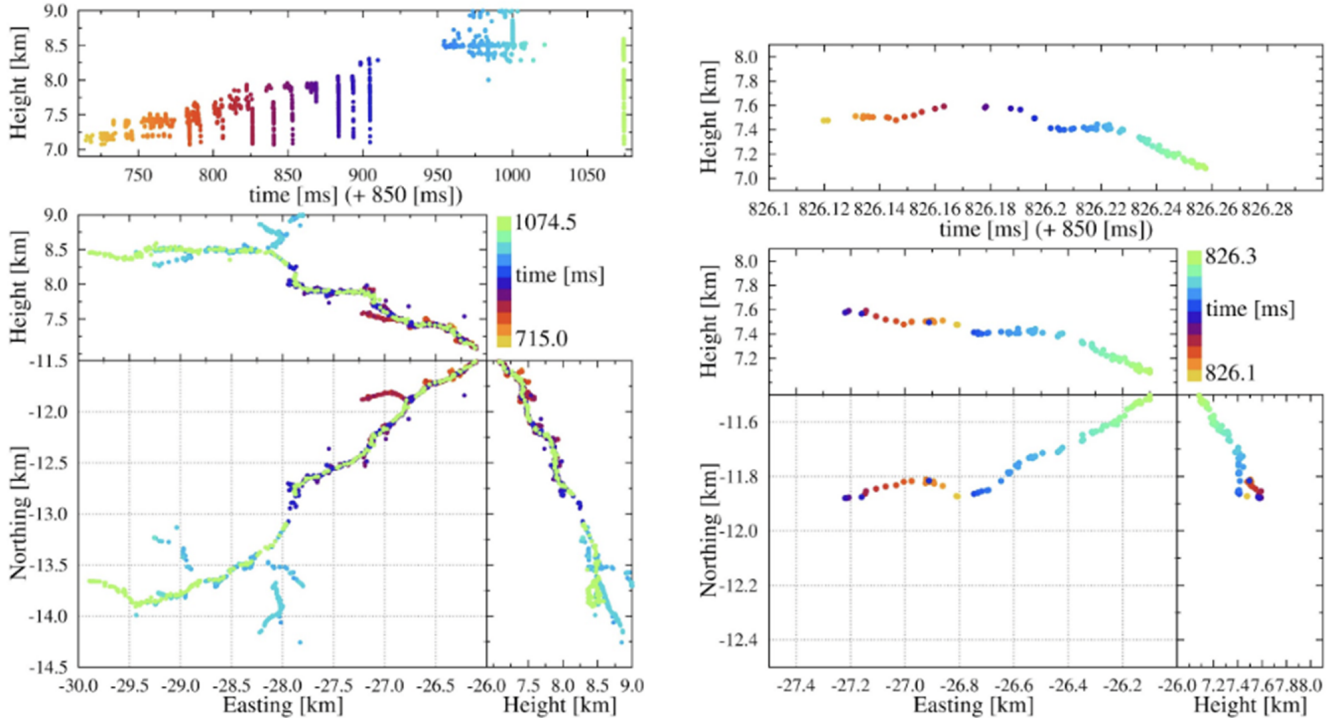

Figure 1.1: Positions and Timings of part of a lightning flash showing some dart leaders[O. Scholten, personal comm.].

Figure 1.1 shows a typical plot of some leaders in a lightning flash as made by the LOFAR lightning imaging group, which includes both spatial and temporal information of the sources emitted by the leaders. The left panel displays data for 2287 sources, all within a time interval of 300 ms. The upper part of the left panel represents the time-height graph, while the lower part presents the 3D spatial information, with colors indicating different times.

Looking more closely at the left panel of Figure 1.1, the time-height graph reveals distinct vertical lines of points spaced at intervals, indicating significant height changes occurring over extremely short time scales. This pattern is a hallmark feature of dart leaders. For example, around 1075 ms, there is a group of green points that exhibit substantial height variation. From the spatial plot below, we can see that this dart leader changes altitude by approximately 1.5 km and horizontal distance by around 5 km in a very short period, reflecting the high speed characteristic of dart leaders.

In addition to these vertical lines, the time-height graph also shows some more scattered points, which likely correspond to needle activity along the positive leader. These scattered sources suggest the presence of smaller, less organized structures typical of needle formations.

The right panel of Figure 1.1 focuses on a dart leader near 826 ms, providing both temporal and spatial details. Unlike the left panel's broader 360 ms time scale, this set of plots zooms in to a 0.2 ms window to reveal finer details. The position plot shows that the dart leader initially moves westward from approximately -26.8 km to -27.2 km in the easting direction, then quickly reverses course, moving back east to -26.8 km, before continuing to move steadily in the northeast direction. Notably, during the initial westward and then eastward motion, there is minimal height change compared to the subsequent movement.

4

The details shown in the right panel are ones that cannot be observed in the left panel due to the significant difference in time scales between the two side of Figure 1.1, with approximately an 1800 times' difference. This clearly demonstrates the point we discussed earlier about the vastly different time scales involved in lightning research. By studying lightning at different time scales, we can uncover distinct characteristics of the phenomenon. Additionally, this highlights the need for precise temporal and spatial resolution when analyzing lightning events, as different aspects of lightning behavior become apparent only at certain scales of observation.
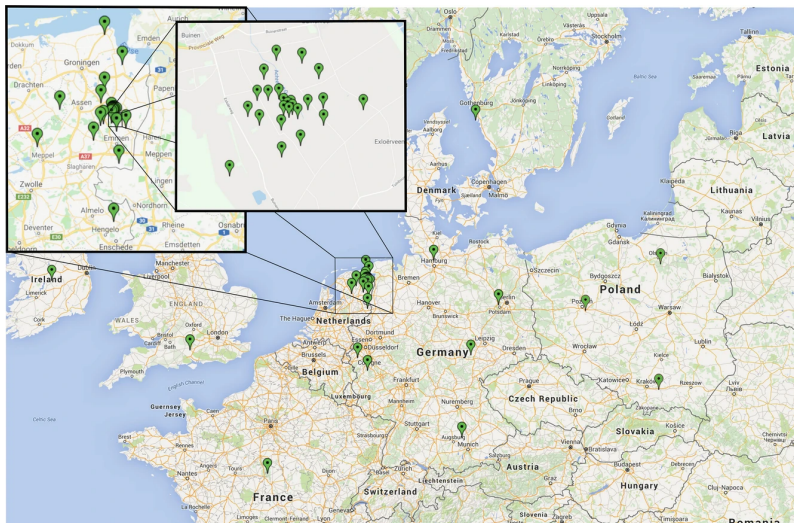
1.1 LOFAR

Figure 1.2: Layout of the LOFAR stations [12]

The Low-Frequency Array (LOFAR) is a radio telescope consisting of thousands of antennas, including Low Band Antennas (LBAs) and High Band Antennas (HBAs), distributed across Europe [12]. In lightning research, we use the Low Band Antennas, which are inverted V-shaped dipole antennas operating in the frequency range of 10–90 MHz [13]. Each station comprises 96 LBAs, arranged in pairs with each dipole oriented perpendicular to the other. These are referred to as the X-dipole, aligned in the northeast-southwest direction, and the Y-dipole, aligned in the northwest-southeast direction [6].

LOFAR's stations are distributed across several European countries, with the majority located in the Netherlands. Figure 1.2 illustrates the distribution of all LOFAR stations, showing that they span distances of up to thousands kilometers. In the top-left corner of the figure, the layout of LOFAR stations in the Netherlands, including its central region, is magnified for clarity. The research discussed in this essay exclusively uses stations situated in the Netherlands. As depicted in the magnified portion of Figure 1.2, the central region features a particularly dense array of antennas, covering a radius of approximately 300 meters. The total baseline length of the array in the Netherlands extends to around 100 kilometers [13]. This high density of antennas, combined with the large baseline, enables LOFAR to achieve a wide detection range and high imaging precision, making it exceptionally well-suited for lightning research.

1.2 TRI-D imaging technique

First, we will introduce how the impulsive imager localizes a source. The process begins by selecting a reference antenna and choosing a segment of the time trace of the signal from that source at the reference

antenna. This segment is then compared to the signals received by other antennas, allowing the calculation of the time differences in the pulse's arrival at different antennas [2, 13]. A chi-square fit is then used to determine the position of the source [13].

In contrast, this essay uses the TRI-D method, which is a more precise, though computationally intensive, imaging technique compared to impulsive imaging. In TRI-D, space is divided into many voxels, and the signals from all antennas are coherently added for each voxel. Using the Jones matrix, the signal intensities measured by the X and Y dipoles are transformed into the transverse polarization of the incoming radiation as received by the antennas, allowing each antenna to "see" the source from different angles [14].

It is important to note that, since our study focuses on lightning, which is relatively close to most antennas, the imaging technique used here is known as near-field imaging. This is different from far-field imaging, where the signal can be approximated as a plane wave. In near-field imaging, the polarization signal received by each antenna depends on the antenna's orientation relative to the source. The dependence of the antenna output signal, the measured voltages, on the polarization and angle of incidence of the signal on the antenna is expressed by the antenna function.

To extract information about the source, a pointlike dipole is used as a model source, and in the frequency domain, a formalism has been developed to perform coherent summation, allowing the extraction of the signal source's intensity and polarization direction [14]. This method is highly complex. Its basic principle involves treating the polarization direction and intensity of the point source as free parameters and then applying a chi-square criterion to fit the signal traces recorded by each antenna. Afterward, the results are transformed back into the time domain using a Fourier transform. The intensity and direction of the model source for each voxel are determined over time, with time slices created at 100 ns intervals. The voxel with the strongest source is then identified, and interpolation is used to determine the exact location of the strongest source within that voxel [6].

1.3 Polarization

TRI-D interferometry is capable of extracting 3D polarization information of the source, which significantly aids in the study of different types of leaders. By analyzing the polarization data, it is possible to infer the three-dimensional motion of charges, providing detailed insights into the geometric structure, length, and branching patterns of charge channels. This information is crucial for enhancing our understanding of lightning leader dynamics.

In this context, to assess the reliability of the 3D polarization data obtained through TRI-D interferometry, this study focuses on analyzing the polarization direction of background radiation. The rationale for this approach lies in the theoretical expectation that the polarization direction of background radiation should be uniformly distributed. By examining this, we can determine the consistency and accuracy of the polarization data produced by the system.

After analyzing the polarization data of the background sources, we found that the distribution was not uniform, contrary to theoretical expectations. To further investigate the underlying causes of this inconsistency, we made adjustments to the antenna function and calibration files. By examining how these adjustments impacted the final results, both in terms of the method and degree of influence, we aimed to identify or eliminate certain factors that could be contributing to the observed discrepancies. This process is crucial for ensuring the reliability and accuracy of the 3D polarization data obtained through TRI-D interferometry.

2 Modification of the Antenna Function

As previously described, the antenna function describes how the output signal, that is, the measured voltages, depends on the polarization and angle of incidence of the incoming signal as it reaches the antenna. This section focuses on the modification of the LOFAR antenna function as used in the TRI-D imager for lightning research to investigate its effects. Before making adjustments, we will provide a detailed explanation of how we use the TRI-D imaging and discuss the current issues with this imaging, which highlights the reasons for understanding the effects of modifications to the antenna function.

Since we will be using TRI-D imaging to capture and analyze data across different times and locations, we will establish a unified reference framework for consistency. Position will be measured relative to the LOFAR-core, the Superterp. Time will be referenced relative to the start of the flash, i.e. this is a time trace before the time of the flash, and thus background. Firstly, we used TRI-D imaging to calculate the background data of a region located at E = -32 km, N = -4 km, and h = 6.5 km at t = -218 ms (for convenience, we refer to this region as Region A). We then analyzed the polarization directions of all background sources within this region, resulting in the distribution shown in Figure 2.1. We use θ representing the zenith angle and ϕ representing the azimuth angle. In addition, to clearly distinguish between position and polarization coordinates, the subscript "p" to represent polarization coordinates and the subscript "a" for the angles from which the radiation from the source arrives at the antenna. East corresponds to $\phi_a = 0°$ and North corresponds to $\phi_a = 90°$.

In Figure 2.1, we used $\cos\theta_p$ instead of θ_p directly, this is to ensure that the corresponding volume under unit coordinates is the same, i.e., the unit solid angle Ω:

$$d\Omega = \sin\theta \, d\theta \, d\phi = -d\cos\theta \, d\phi,$$

Thus, for the background noise, we expect the polarization to be uniform distributions in $\phi \in (-180°, 180°]$ and $\cos\theta_p \in [-1, 1]$
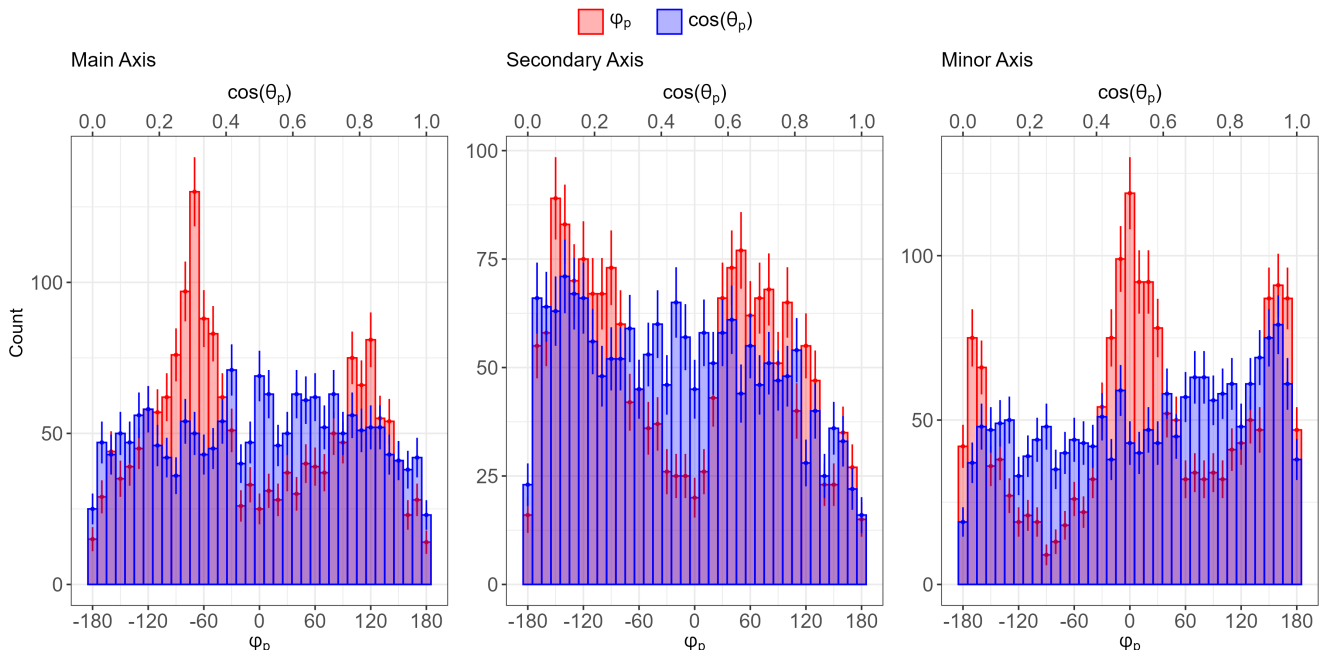

Figure 2.1: Polarization directions distribution for the three axes of the dipole distribution for each slice (see text). Shown is $\cos\theta_p$ (top scale, blue, binwidth=0.025) and ϕ_p (bottom scale, red, binwidth=9°).

The currents in a lightning flash, within LOFAR's frequency band, can be thought of as oscillating dipoles, each oscillating back and forth along a line with a specific orientation. During the TRI-D imaging process, as described in the paper [14], time is divided into 100 ns slices, with each slice containing 20 time samples where for each a source dipole direction is calculated. For the distribution of these dipoles, the program performs a principal value decomposition and represents the polarization direction using three axes: the main, secondary and minor axes. Since each section of data contains 2650 slices, for each axis, we can analyze the distribution of these slices across two polarization directions. In this study, we are most interested in the polarization axis with the highest intensity, which we refer to as "Main Axis".

In Figure 2.1, we present three plots, each corresponding to the polarization direction distribution of one of the three polarization axes. Red represents the distribution of ϕ_p, and blue represents the distribution of $\cos(\theta_p)$. Since the scales of ϕ_p and $\cos(\theta_p)$ differ, the lower x-axis represents the various ϕ_p angles with a bin width of $9°$, while the upper x-axis corresponds to the different $\cos(\theta_p)$ values with a bin width of 0.025. In Figure 2.1, it can be seen that the $\cos\theta_p$ distribution is uniform, but the ϕ_p distribution shows a strikingly uneven pattern. This is unusual, as in theory, the background sources should always have a uniform distribution. The objective of this thesis is to explore the reasons behind this surprising result.

One possible explanation is that antennas located farther from the source observe it at a relatively large zenith angle, where the antenna gain is small, poorly understood, and more susceptible to external factors, such as vegetation growth around the station. These influences may reduce the accuracy of the results. To investigate whether these factors are indeed affecting the polarization distribution, we can modify the antenna function to reduce the weight of the distant antennas. By recalculating the region after this adjustment, we can determine whether the uneven polarization distribution arises from such errors.

After understanding the reasons for modifying the antenna function, we will next describe the methods used for adjusting the antenna function. The antenna function is multiplied by a θ_a-dependent function to ensure that the phase remained unchanged, where θ_a here represents the zenith angle of position relative to each antenna, i.e.,

$$\text{New Antenna Function} = F(\theta_a) \times \text{Original Antenna Function} \tag{1}$$

Here, $F(\theta_a)$ is the function we choose to modify the antenna function. We do not modify the ϕ_a dependence of the antenna function that closely resembles that of a dipole. There are two functions are chosen for adjustment:

Function f:
$$f(\theta_a \mid a) = cos(a\theta_a) \tag{2}$$

Function g:
$$g(\theta_a \mid c, d, k) = \frac{a}{1 + exp(k(\theta_a - d))} + b \tag{3}$$

where
$$a = (1 - c) \div \left(\frac{1}{1 + exp(-kd)} - \frac{1}{1 + exp(k(90° - d))} \right)$$
$$b = 1 - \frac{a}{1 + exp(-kd)}$$

For changing the $g(\theta_a)$ in a more explicit way, c,d,k is used as the input parameter instead of a,b,k,d, then convert these parameter to a and b by the equation above. Here, c is the result at $\theta_a = 90°$; d determines how long is the flat part; k determined how smooth the function is, as k larger, the smoother of the function.
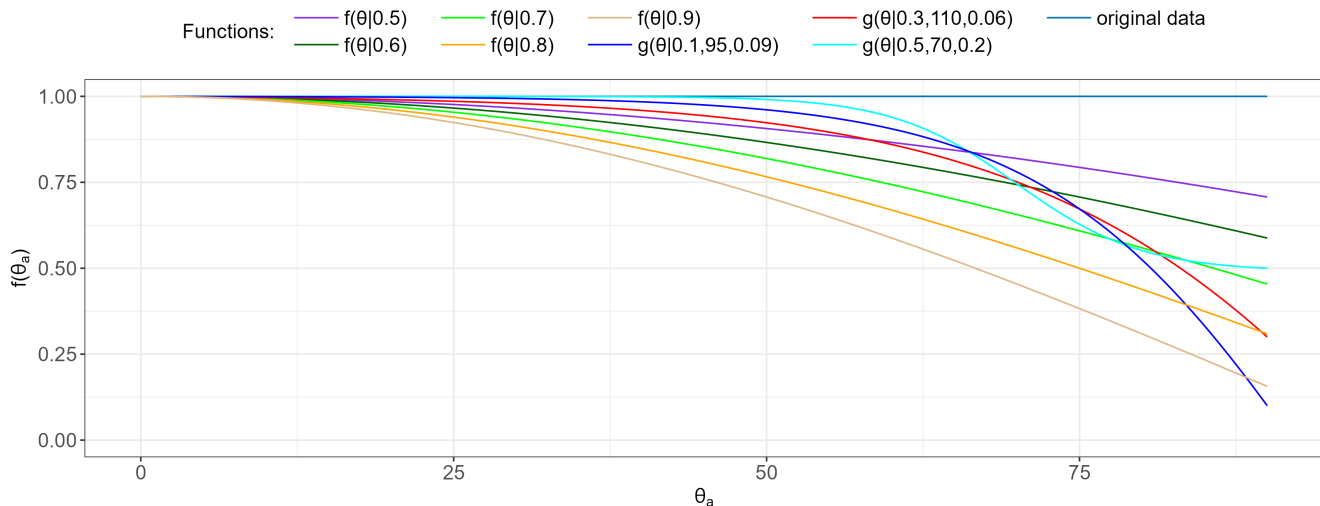
Figure 2.2: The functions f (eq.2) and g (eq.3) with different parameters as used in this work.

Compared to $f(\theta_a \mid a)$, the function $g(\theta_a \mid c, d, k)$ offers greater flexibility in modifying the gradient across different regions of θ_a, as shown in the Figure 2.2. This flexibility allows us to adjust the values in the tail sections for various θ_a intervals according to research needs, enabling a more detailed investigation of how different θ_a ranges impact the final results. However, since the program uses Legendre polynomials for interpolation, rapidly changing antenna functions can cause problems with the code. Therefore, when selecting the parameters for $g(\theta_a \mid c, d, k)$, it is crucial to avoid configurations that make the function too steep. This precaution will help maintain the stability of the interpolation process and prevent potential issues in the code, ensuring smooth execution.

The adjusted antenna function was applied to both background data and data containing sources. We select different locations and different time. This sections will discuss background data of Region A at different time, background data of Region B, and source data of Region C.

2.1 Background Analysis on Region A

Firstly, the background data around t = -218 ms will be discussed. Here, the time t is measured relative to the start of the flash, so negative time refers to the period before the flash. During this time, the atmosphere is relatively quiet, allowing only the background to be observed.

As shown in Figure 2.3, the shape of the intensity distribution is similar across different functions and axes: all exhibit a peak, with the right side of the curve decreasing more slowly than the left. Distributions corresponding to antenna functions that decrease more rapidly with increasing θ_a show a lower peak and a broader spread. This change is expected because the antenna function is multiplied by θ_a-dependent different functions with a maximum value of 1, meaning it is scaled by different values between 0 and 1 depending on θ_a. Consequently, the overall antenna function is reduced. To match the energy in the data, the reconstruction process has to increase the power of the radio sources to account for the total energy in the data.

Figure 2.4 shows the azimuth angle ϕ_p dependence of the main axis, and Figure 2.5 shows the distribution for the secondary and tertiary axis. These figures use a bin width of 10 to count the total number of sources in each interval, and then plots this as a line graph. The shaded region below the line represents
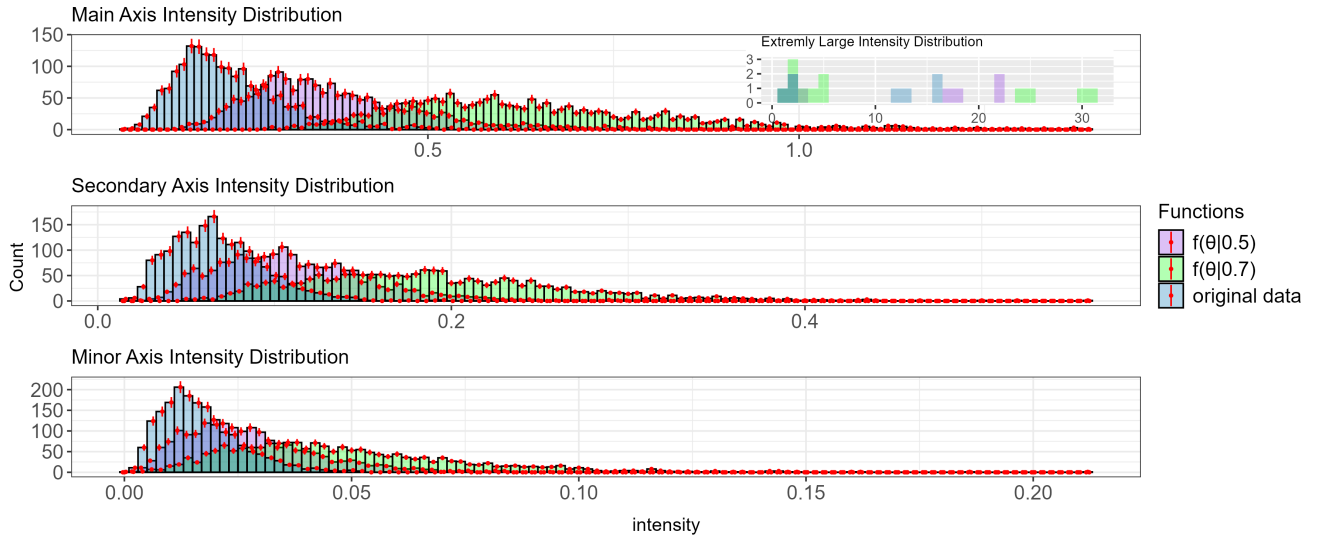
Figure 2.3: Intensity distribution along three axes, with different colors representing the use of various antenna functions. Intensity is in arbitrary units. The inset in the main axis panel shows the frequency of intensity values greater than the x-axis limit (1.39).
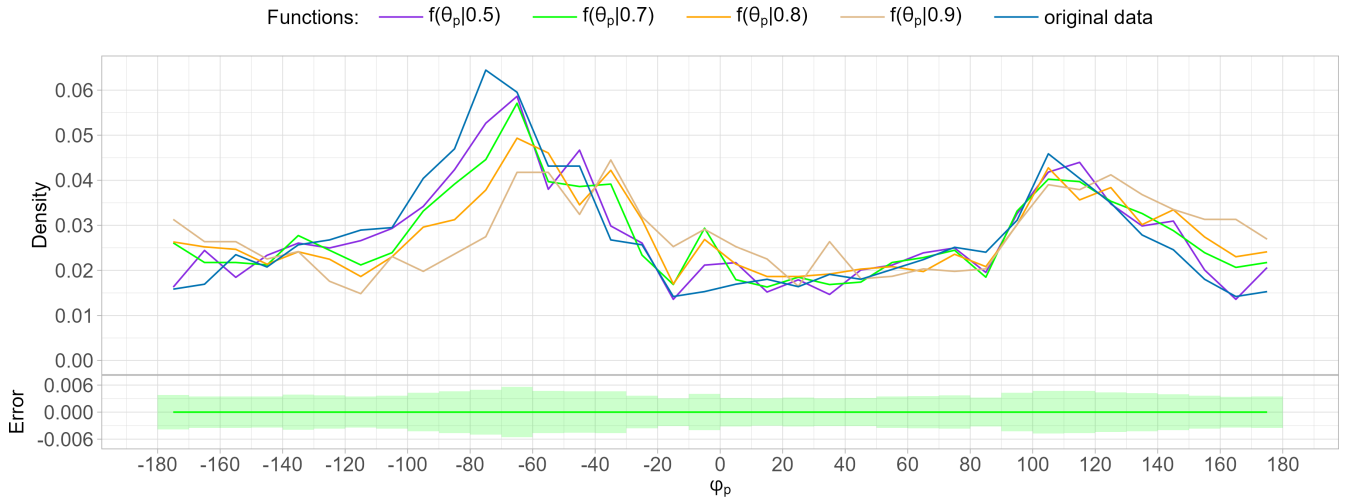

Figure 2.4: Azimuth angle ϕ_p distribution for the main axis (bin width = 10). Different colors represent various antenna functions. The green bar at the bottom corresponds to the error bar for the results calculated with $f(\theta_a \mid 0.7)$ (green line).

the error, which is calculated using

$$E_c = \sqrt{N+1},$$

$$E_d = \frac{\sqrt{N+1}}{T},$$

where N is the number of sources at the corresponding angle, T is the total number of sources, E_c is the error for count and E_d is the error for density. This shaded error region provides a visual representation of the uncertainty in the distribution. Since the errors for all functions are very similar, we only plot the error for the polarization distribution calculated for the function $f(\theta_a \mid 0.7)$ as a reference.

As mentioned earlier, the polarization direction ϕ_p along the main axis is uneven, with distinct peaks at $-70°$ and $110°$. After changing the antenna function, the distribution shows only slight changes. The

10

peak near $-70°$ shifts slightly to the right, with a slight decrease in the number of sources, while there is a small increase in the number of sources around $170°$. A more detailed analysis of these changes will be provided in the following sections.
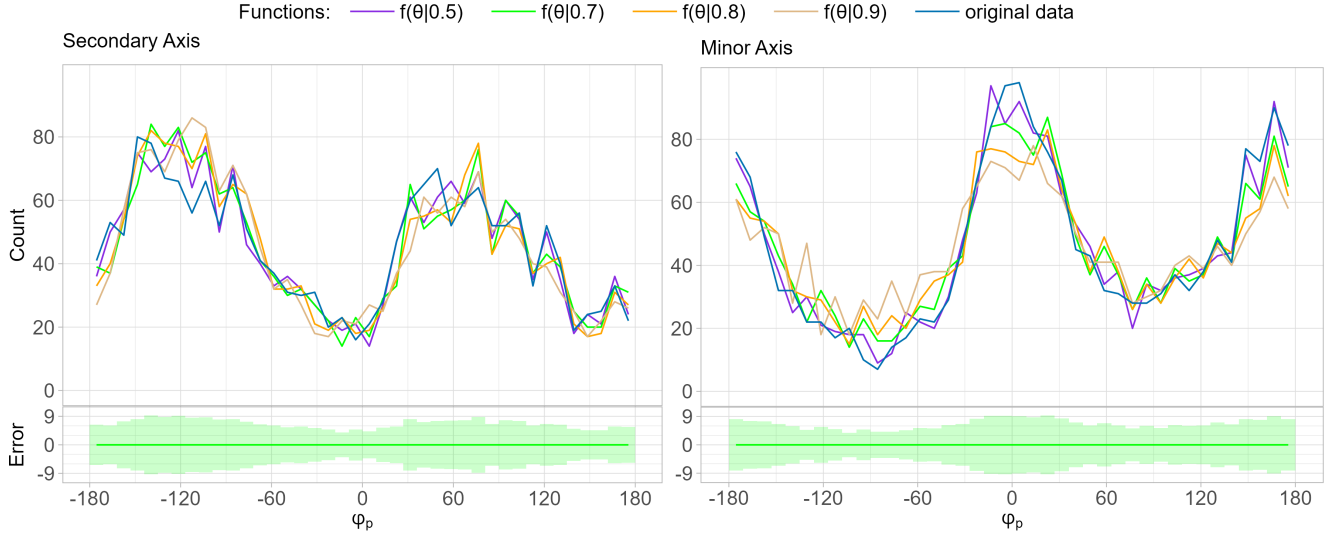

Figure 2.5: Azimuth angle ϕ_p distribution for the secondary and minor axes (binwidth = 9). Different colors represent the use of various antenna functions. The green bar at the bottom represents the error bar for the results calculated with $f(\theta_a \mid 0.7)$ (green line).

In Figure 2.5, the secondary axis has a concentrated distribution around ϕ_p equals to $-130°$ and $50°$, and the tertiary axis exhibits significant peaks at $-180°$, $0°$, and $180°$. Neither of their distributions was significantly affected by the change in the antenna function.

Since only the main axis showed noticeable changes, Figure 2.6 was plotted. On the left panel, it displays the ϕ_p distribution of the main axis, while the right side uses different colors to compare the various functions used to modify the antenna function. First, it can be concluded that regardless of whether function $f(\theta_a \mid a)$ or $g(\theta_a \mid c, d, k)$ is used, the overall trend matches the observations made in Figure 2.4, with differences mainly in the magnitude of the change. Compared to the original data, the modified antenna functions show a significant decrease in the distribution between $-120°$ and $-40°$, while other parts exhibit a slight increase. When looking at $f(\theta_a \mid a)$ specifically, the larger the decrease in $f(\theta_a \mid a)$ as θ_a increases, the greater the corresponding change in the ϕ_p distribution.

When comparing $f(\theta_a \mid a)$ and $g(\theta_a \mid c, d, k)$, and first observing $f(\theta_a \mid 0.7)$ (green line) and $g(\theta_a \mid 0.3, 110, 0.06)$ (red line), we can see that compared to $f(\theta_a \mid 0.7)$, $g(\theta_a \mid 0.3, 110, 0.06)$ decreases more slowly at first but then drops more rapidly toward the end. By analyzing their corresponding ϕ_p distributions, we observe that even though $f(\theta_a \mid 0.7)$ remains much lower than $g(\theta_a \mid 0.3, 110, 0.06)$ for most of the θ_a range, only surpassing $g(\theta_a \mid 0.3, 110, 0.06)$ around $\theta_a = 85°$, the change in ϕ_p distribution for $g(\theta_a \mid 0.3, 110, 0.06)$ is more significant than for $f(\theta_a \mid 0.7)$. Other $g(\theta_a \mid c, d, k)$ functions show similar behavior. This could be because the region under investigation is located at $\theta_a = 78.6°$ relative to the reference antenna and its surrounding antennas, making the program more sensitive to changes in the antenna function around this angle. Therefore, for $g(\theta_a \mid c, d, k)$ functions that begin to decrease significantly after $\theta_a = 75°$, the impact on the ϕ_p distribution is more pronounced.

From Figure 2.6, we can also observe a leftward shift in the peak located at $\phi_p = -70°$. Furthermore, for different functions, the extent of this leftward shift is proportional to the reduction in peak height, suggesting that this shift is unlikely to be caused by errors. A similar, though less pronounced, shift can be seen around $\phi_p = -110°$. The systematic nature of these changes in the distribution will be explored
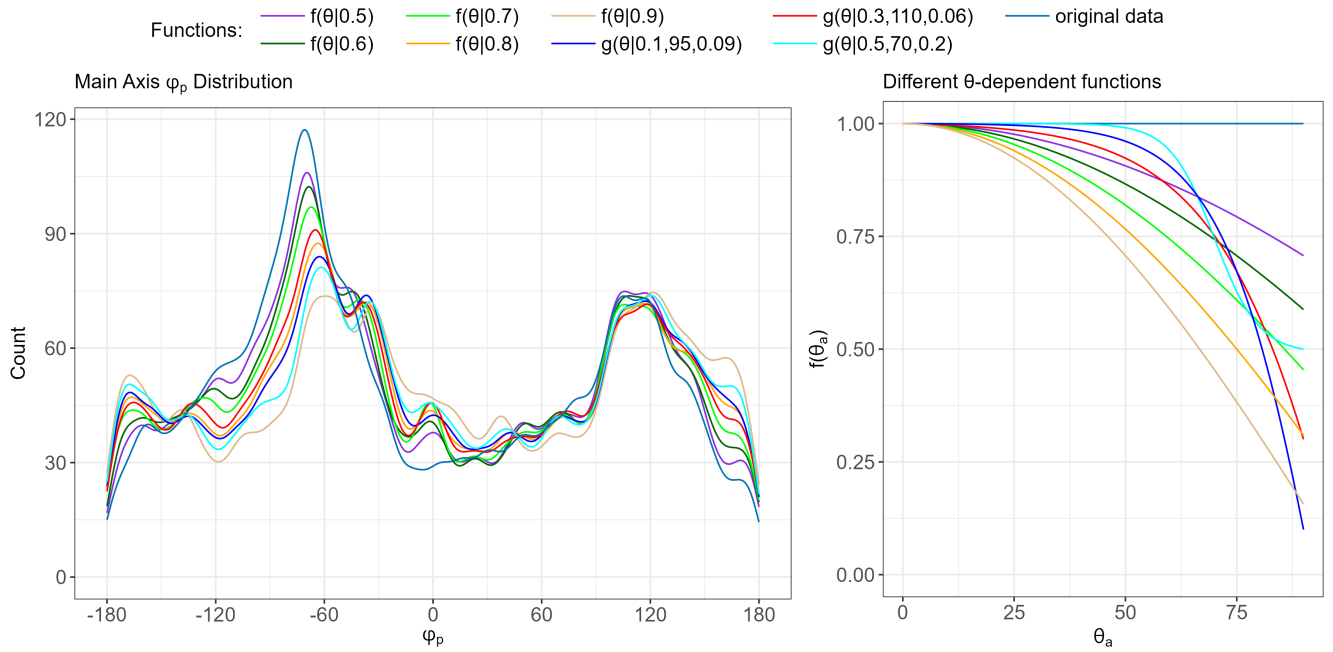
11

Figure 2.6: The left panel shows the smoothed azimuth angle ϕ_p distribution for the main axis, while the right panel displays all the θ_a-dependent functions used to generate the antenna functions. Different colors indicate the different antenna functions (repeated from Figure 2.2 for easier interpretation of the results).

further in subsequent sections, where we will investigate additional regions to gain a deeper understanding of these patterns.
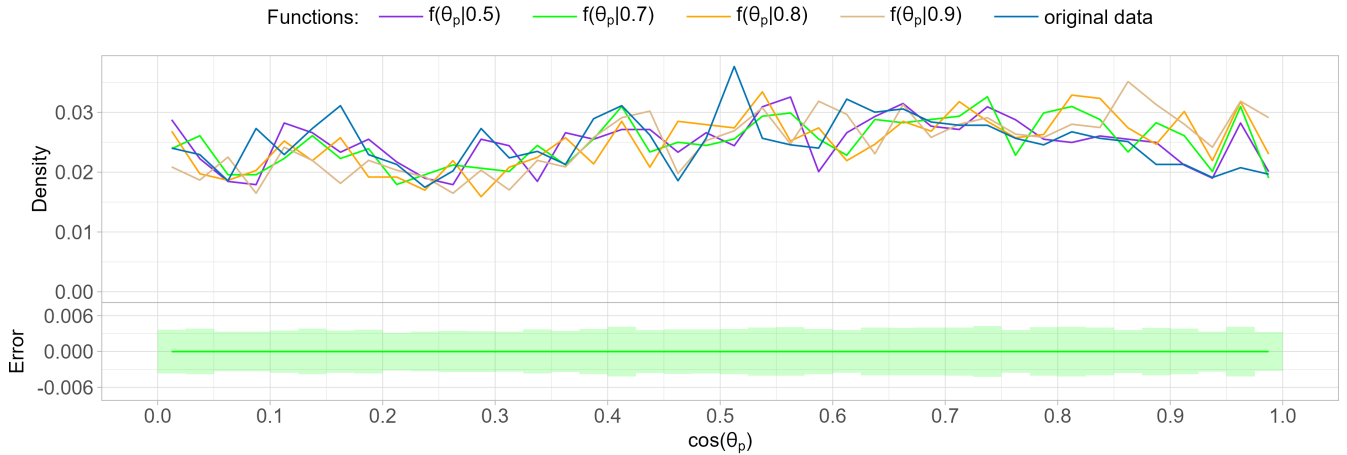

Figure 2.7: Zenith polarization angle $\cos\theta_p$ distribution for the main axis (binwidth $= 0.025$). Different colors represent the use of various antenna functions. The green bar at the bottom corresponds to the error bar for the results calculated with $f(\theta_a \mid 0.7)$ (green line).

As shown in Figure 2.7, the $\cos(\theta_p)$ distribution is very close to a uniform distribution. Looking at the line corresponding to the original antenna function (blue), it shows no significant fluctuations, with most points oscillating within the error margin of density $= 0.022$. However, there is a noticeable peak between $\cos(\theta_p) = 0.5$ and 0.7, which exceeds the error range. For the data generated with the modified antenna functions, there is no such pronounced peak in the same range ($\cos(\theta_p) = 0.5$ to 0.7), but we

observe that the count on the left side of the plot is slightly lower than on the right. At $\cos(\theta_p) = 1$, a pattern emerges: as the tail of the antenna function becomes smaller, the corresponding count increases. However, this difference is subtle, and further investigation into other regions is necessary to draw more robust conclusions.
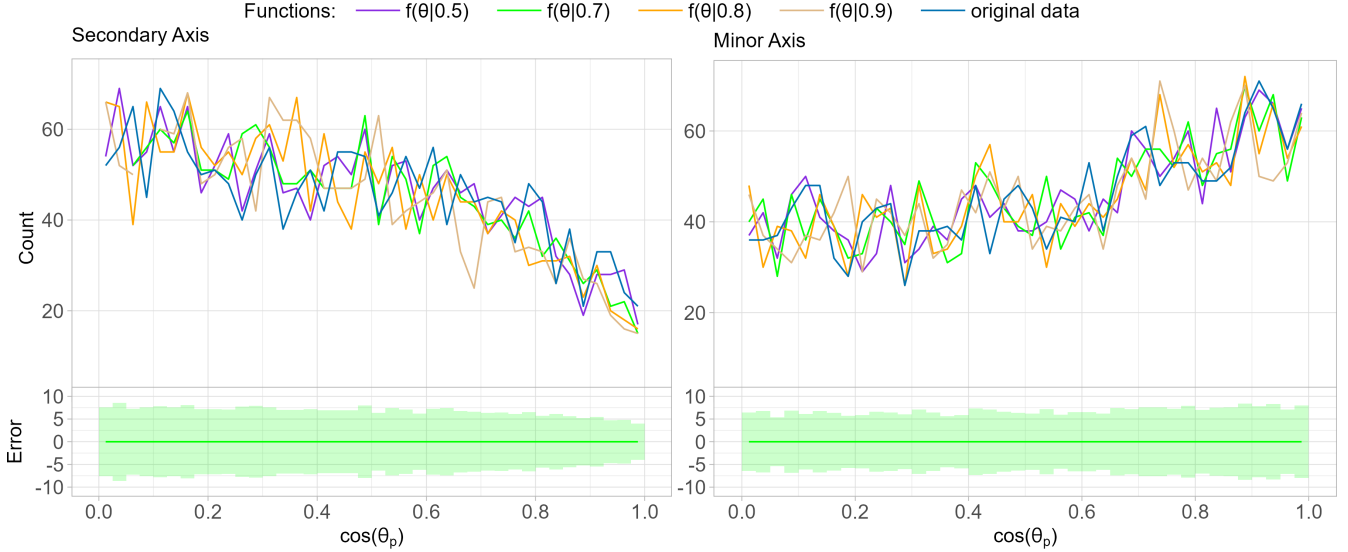

Figure 2.8: Zenith polarization angle $\cos\theta_p$ distribution for the secondary and minor axes (binwidth = 0.025). Different colors represent the use of various antenna functions. The green bar at the bottom represents the error bar for the results calculated with $f(\theta_a \mid 0.7)$ (green line).

Figure 2.8 presents the $\cos\theta_p$ distribution for the secondary and minor axes. Similar to the ϕ_p distribution, the impact of different antenna functions on the secondary and tertiary axes is difficult to discern from the figure and is almost negligible.

To further understand the distributions of ϕ_p and $\cos\theta_p$, the results of the Kolmogorov-Smirnov (K-S) tests for ϕ_p and $\cos\theta_p$ will be discussed. The K-S test is commonly used to determine whether two samples come from the same distribution. Specifically, for two samples with n and m elements respectively, each following the distributions functions $F(x)$ and $G(x)$, then,

$$D_{n,m} = \sup_x |\hat{F}_n(x) - \hat{G}_m(x)|, \tag{4}$$

where $\hat{F}_n(x)$ and $\hat{G}_m(x)$ are the empirical cumulative distribution functions(eCDFs) of $F(x)$ and $G(x)$ [15]. The output includes the D value and p value, both ranging from 0 to 1. As shown above, the D value represents the maximum difference between the eCDFs of the two samples. A smaller D value indicates a smaller maximum difference between the eCDFs, suggesting that the two distributions are more similar. The p value indicates the probability that the functions $F(x)$ and $G(x)$ show an even larger difference, i.e., a greater D value, assuming that the two are different samplings from the same distribution. Therefore, a smaller p value suggests a higher probability that the two samples come from different distributions.

Table 1 and Table 2 displays the results of Kolmogorov-Smirnov tests of ϕ_p and $\cos\theta_p$ conducted between the original data and other different functions. It can be seen that for the major axes, the D value is generally small, with most values below 0.05. This is consistent with the observations in Figure 2.6 and Figure 2.7, where the differences between the distributions corresponding to different antenna functions are not substantial. However, interestingly, for the main axis, although the D values are generally small, the p-values are also significantly low. This is particularly evident for functions like $f(\theta_a \mid 0.9)$. This suggests that the polarization direction data of the main axis obtained using different antenna functions

13

Table 1: Kolmogorov-Smirnov Test between the polarization directions obtained from modified antenna functions and those from the original antenna function for the Main Axis

Main Axis	$f(\theta_a \mid 0.5)$	$f(\theta_a \mid 0.6)$	$f(\theta_a \mid 0.7)$	$f(\theta_a \mid 0.8)$	$f(\theta_a \mid 0.9)$	$g(\theta_a \mid 0.3, 110, 0.06)$	$g(\theta_a \mid 0.5, 70, 0.2)$
D value (ϕ_p)	3.16×10^{-2}	4.15×10^{-2}	5.26×10^{-2}	7.49×10^{-2}	1.14×10^{-1}	6.83×10^{-2}	9.68×10^{-2}
p value (ϕ_p)	3.19×10^{-1}	8.48×10^{-2}	1.24×10^{-2}	7.03×10^{-5}	0.00	3.94×10^{-4}	1.00×10^{-7}
D value ($\cos\theta_p$)	2.16×10^{-2}	2.93×10^{-2}	4.15×10^{-2}	5.66×10^{-2}	6.28×10^{-2}	4.92×10^{-2}	5.76×10^{-2}
p value ($\cos\theta_p$)	7.86×10^{-1}	4.12×10^{-1}	8.42×10^{-2}	5.73×10^{-3}	1.50×10^{-3}	2.40×10^{-2}	4.69×10^{-3}

Table 2: Kolmogorov-Smirnov Test between the polarization directions obtained from modified antenna functions and those from the original antenna function for the Secondary Axis

Secondary Axis	$f(\theta_a \mid 0.5)$	$f(\theta_a \mid 0.6)$	$f(\theta_a \mid 0.7)$	$f(\theta_a \mid 0.8)$	$f(\theta_a \mid 0.9)$	$g(\theta_a \mid 0.3, 110, 0.06)$	$g(\theta_a \mid 0.5, 70, 0.2)$
D value (ϕ_p)	1.67×10^{-2}	1.91×10^{-2}	2.60×10^{-2}	3.39×10^{-2}	4.26×10^{-2}	3.28×10^{-2}	3.57×10^{-2}
p value (ϕ_p)	9.60×10^{-1}	8.90×10^{-1}	5.65×10^{-1}	2.43×10^{-1}	7.34×10^{-2}	2.79×10^{-1}	1.94×10^{-1}
D value ($\cos\theta_p$)	2.23×10^{-2}	2.37×10^{-2}	3.22×10^{-2}	5.07×10^{-2}	6.18×10^{-2}	4.59×10^{-2}	5.54×10^{-2}
p value ($\cos\theta_p$)	7.50×10^{-1}	6.80×10^{-1}	2.98×10^{-1}	1.84×10^{-2}	1.89×10^{-3}	4.24×10^{-2}	7.41×10^{-3}

are statistically significant. To facilitate a clearer analysis, these values for the main axis are plotted in Figure 2.9.
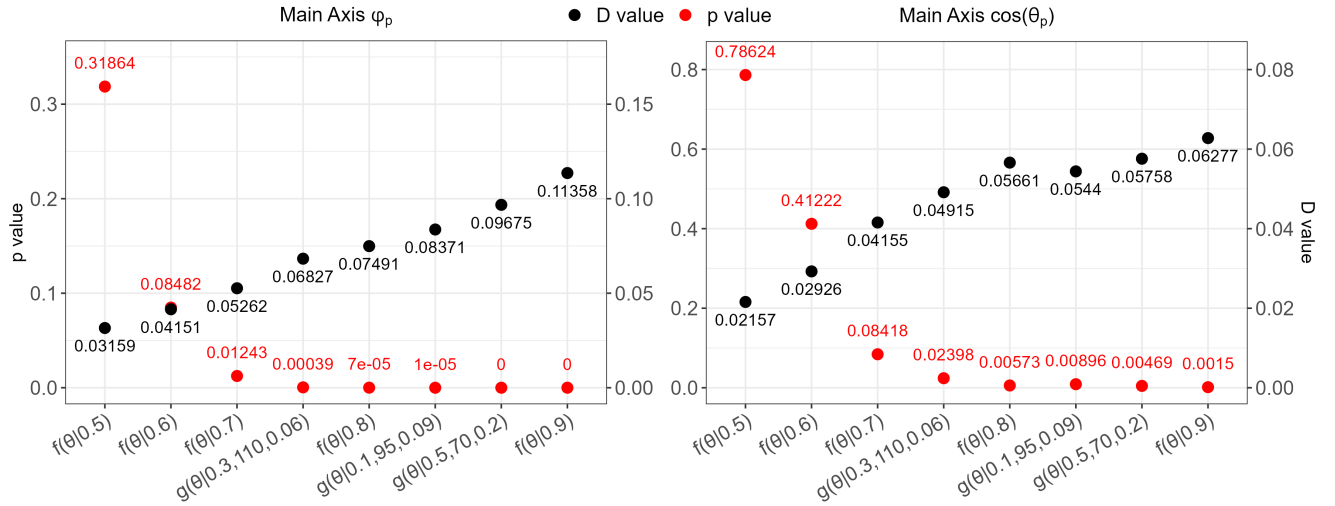

Figure 2.9: Kolmogorov-Smirnov Test between polarization data reconstructed by the original antenna functions and other functions for the main axis. The left panel shows ϕ_p data, and the right panel shows $\cos(\theta_p)$ data. The p-values (red dots, left y-axis) and D-values (black dots, right y-axis) are displayed.

Figure 2.9 plot the p values and D values obtained by Kolmogorov-Smirnov tests of the original data with all other data for the main axis. By examining the D values, we can draw the same conclusions as before. Firstly, the D values for $\cos\theta_p$ are generally smaller than those for ϕ_p. As shown in Figure 2.6 and Figure 2.7, the distribution of $\cos\theta_p$ exhibits less variation compared to the distribution of ϕ_p, remaining consistently close to the original data. Secondly, it is also observed that as the tail of the function becomes lower, the D value gradually increases, which aligns with the trend seen in the distribution plots. Another notable observation is that the D value and p value exhibit opposite trends. The p value represents the probability of obtaining data with a greater deviation than the current observation, implying the likelihood of observing the same or a larger D value under the null hypothesis. Therefore, it is understandable that a smaller D value corresponds to a larger p value.

However, even the largest D value of 0.11 is not particularly significant, yet most of the p values for the corresponding functions are much smaller than 0.05. This is an interesting result. Consequently, it can

14

be concluded that although the data generated by all the selected functions have similar distributions, they originate from entirely different distributions.
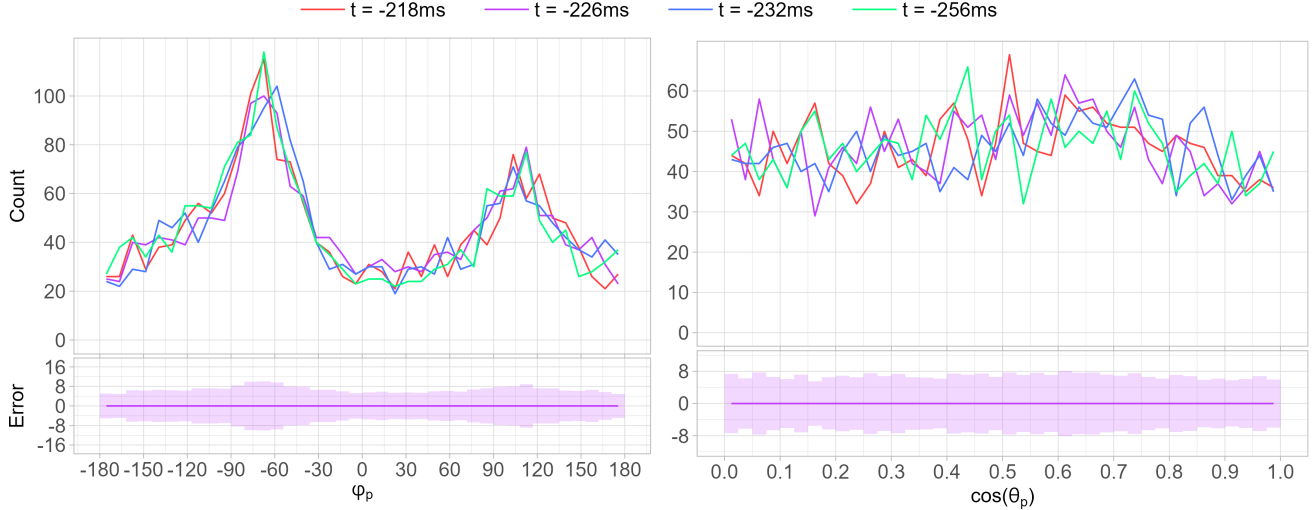
2.1.1 Time Dependence

Figure 2.10: Polarization angle distribution for the main axis. The left panel shows the azimuth angle ϕ_p (binwidth $= 9$), and the right panel shows the zenith angle $\cos(\theta_p)$ (binwidth $= 0.025$). Different colors represent different time intervals. The purple shaded region indicates the error for $t = -226$ ms.

In the previous analysis, we discussed the polarization direction distribution of Region A when the time was around -218 ms (as defined at the beginning of this chapter, referring to 218 ms before the onset of the lightning flash). To examine whether time affects the distribution, we plotted the distributions at other time points within this region in Figure 2.10. It is evident that there is no significant change in the distribution. To further support this observation, we conducted a Kolmogorov-Smirnov Test between the data at t = -218 ms and three other time's data. The results, presented in the Table 3, show that the maximum D value is only 0.03, and the p-values are not particularly low. This indicates that the polarization direction distribution remains consistent across different times. Therefore, it can be concluded that time does not affect the distribution.

Table 3: Kolmogorov-Smirnov Test between the main axis polarization directions in Region A at -218 ms and other time

Main Axis	t = -226 ms	t = -232 ms	t = -256 ms
D value (ϕ_p)	0.030	0.030	0.027
p value (ϕ_p)	0.382	0.379	0.508
D value ($\cos\theta_p$)	0.023	0.027	0.029
p value ($\cos\theta_p$)	0.708	0.524	0.441

2.2 Background Analysis in Region B

Similar to the previous subsection, we will analyze the polarization direction of the background sources in region B (E=-30km, N = 5km, h= 5km) using different antenna functions. The specific results are presented in Figure 2.11.
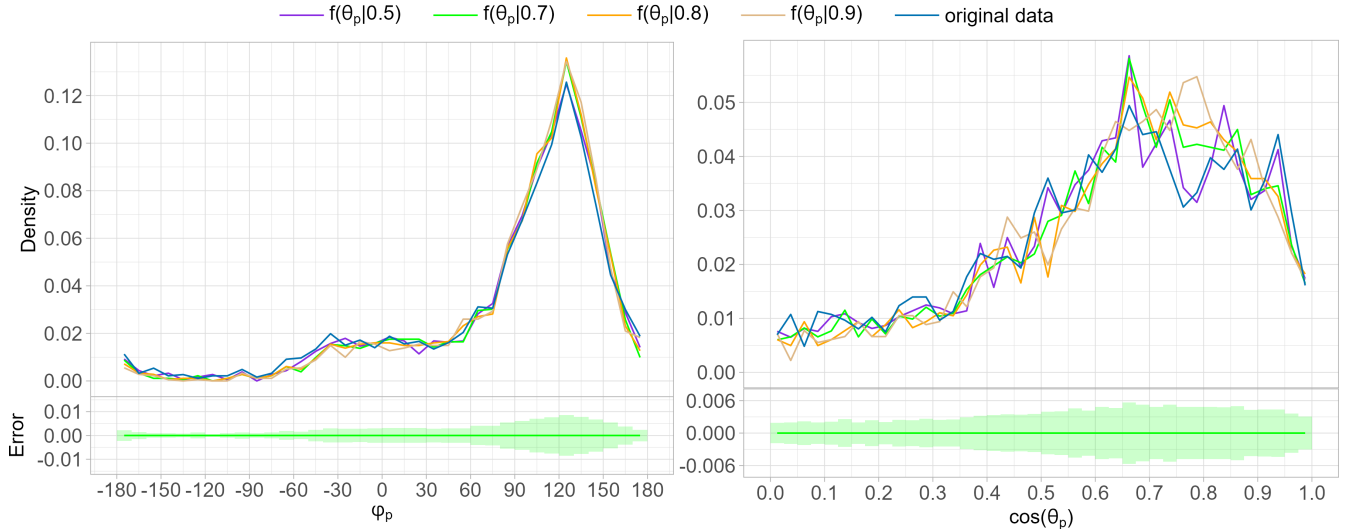
Figure 2.11: Polarization direction distributions in Region B. The left panel shows the ϕ_p distribution (binwidth = 10), and the right panel shows the $\cos(\theta_p)$ distribution (binwidth = 0.025). Different colors represent the use of various antenna functions. The green shaded areas below each graph represent the error bar for the results calculated with $f(\theta_a \mid 0.7)$ (green line).

As is clear by comparing Figure 2.4, 2.7 and Figure 2.11, the polarization direction distribution in Region B is different from that in Region A, indicating that the polarization direction distribution might be location-dependent. We then examine the effect of different antenna functions on the distribution. Similar to the case in Region A, the results for Region B are also minimally affected by the different antenna functions. However, in more detail, we observe that some changes slightly exceed the margin of error and occur at angles different from those corresponding to Region A.

For the ϕ_p distribution, the differences between the lines are mostly within the error margin. The largest discrepancy between the lines is around $\phi_p = 125°$, with a difference of approximately 0.01. In contrast, in Figure 2.4, the maximum difference between the lines is about 0.035, and significant differences can be observed in the range from $\phi_p = -100°$ to $-63°$. For the $\cos(\theta_p)$ distribution, unlike in Figure 2.7, we observe clear differences between the lines in the range $\cos(\theta_p) = 0.75$ to 0.8, which exceed the error margin. Additionally, similar to Figure 2.7, there is a distinct peak in the original data around $\cos(\theta_p) = 0.52$, where it is significantly larger than for other antenna functions. Whether this phenomenon is meaningful will require analysis of polarization distributions in more regions.

From the overall trend, we seem to observe a pattern: when the θ_p distribution is less affected by changes in the antenna function, the ϕ_p distribution is more sensitive to these changes, and vice versa—if the ϕ_p distribution is less affected by antenna function changes, then the θ_p distribution tends to show more pronounced differences. To further explore and understand the effects of the antenna function on the polarization direction distribution, we will extend the analysis to other, more distant regions.

2.3   Background Analysis in Different Locations

Figure 2.12 shows the distribution of the polarization direction of the main axis at different locations, as well as the changes after modifying the antenna function. To better highlight the differences, we compared only the original antenna function (blue) with the $f(\theta_a \mid 0.9)$ modified one (red), as $f(\theta_a \mid 0.9)$ had the most significant impact on the distribution in previous studies. In Figure 2.12, we label the different regions using the azimuth angle ϕ_a between the region and the reference antenna, along with the distance
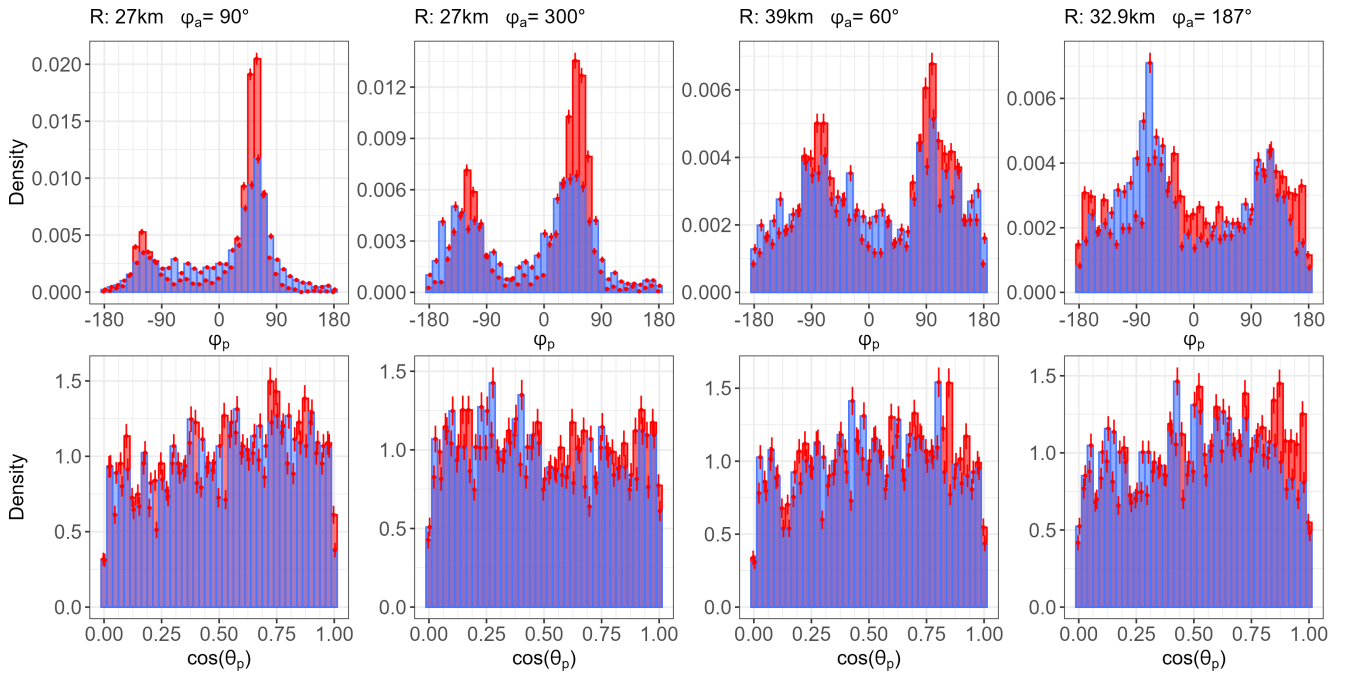
Figure 2.12: The top row shows the main axis ϕ_p distributions(binwidth=$10^{c}irc$), and the bottom row shows the main axis $\cos(\theta_p)$ distributions(binwidth=0.025). The blue bars represent the original data, and the red bars represent $f(\theta_a \mid 0.9)$. Each column corresponds to a different region, with the top and bottom figures representing the same region.

R from the region to the reference antenna. To better analyze the influence of different antenna functions on the polarization direction, we first selected locations where the $\cos(\theta_p)$ distribution was approximately uniform. This approach minimizes the impact of background sources that are concentrated within specific $\cos(\theta_p)$ intervals, allowing for a clearer evaluation of the overall effect of the antenna function on the polarization direction.

From Figure 2.12, we can conclude that the effect of different antenna functions on polarization direction is clearly location-dependent, and the way they alter the ϕ_p distribution varies significantly between regions.

- For the region at R = 27 km, $\phi_a = 90°$, the antenna function modified with $f(\theta_a \mid 0.9)$ shows many more sources with a polarization direction ϕ_p between $50°$ and $67°$ than for the unmodified case.

  changing the antenna function greatly increases the number of background sources in the $\phi_p = 50°$ to $67°$ range.

- In the region at R = 27 km, $\phi_a = 300°$, a similar pattern is observed, but the range where ϕ_p increases is broader, approximately from $\phi_p = 35°$ to $75°$.

- For the region at R = 39 km, $\phi_a = 60°$, the range of increased background sources is even wider, roughly from $\phi_p = 67°$ to $135°$, but the magnitude of increase is smaller compared to the first two regions.

A common observation for these three regions is that after changing the antenna function, the number of sources increases in the ranges where the original ϕ_p distribution (blue) already had a denser concentration of sources. Simply put, this amplifies the peaks in the original distribution. However, in the region at R = 32.9 km, $\phi_a = 187°$ (previously studied as Region A), the antenna function changes the distribution by decreasing the number of sources in the previously dense ϕ_p range while increasing it in other ranges.

17

When analyzing the $\cos(\theta_p)$ distribution for these regions, similar to the situation in Region A, the differences between antenna functions are minimal. This may suggest that in regions where the $\cos(\theta_p)$ distribution is relatively uniform, the antenna function has a smaller impact on the $\cos(\theta_p)$ distribution. However, this is a preliminary observation based on a small number of samples, and further comparisons across more regions are required to verify this hypothesis.
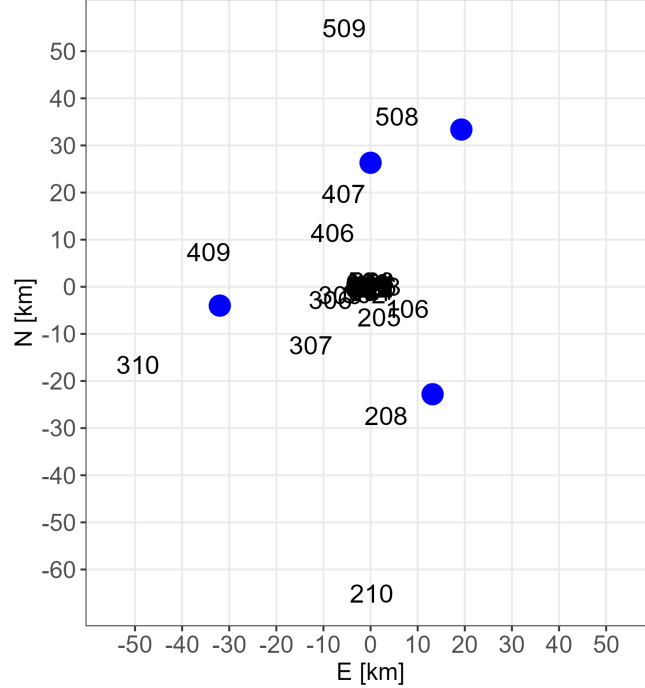

Figure 2.13: The schematic locations of LOFAR stations and the positions of the regions from Figure 2.12 (blue dots), with an "A" marking the location of Region A. The numbers correspond to the identifiers of the LOFAR stations. Due to the high density of stations near the core, the identifiers in the center overlap. However, since the analysis in this section focuses on the distribution of stations in other areas and simply acknowledges the high density at the core, no special adjustments were made.

To better understand the different effects of the antenna function on the ϕ_p distribution across regions, we plotted Figure 2.13 to compare the distribution of LOFAR stations with the positions of the regions under study. Here, the numbers represent station identifiers, and the blue dots indicate the locations of the four regions shown in Figure 2.12, with the reference antenna at the center.

Before analyzing the reasons behind the observed change in Figure 2.12, it is important to first explain how changing the antenna function affects polarization direction. The functions we selected to adjust the antenna function are designed such that they are approximately equal to 1 when the zenith angle θ_a between the region and the antenna is small, and gradually decrease as θ_a increases. In practical terms, this means that after changing the antenna function, antennas closer to the signal source, which have larger θ_a, experience minimal change in the received signal strength. In contrast, antennas that are farther away, with smaller θ_a values, experience a significant reduction in signal strength. In this scenario, the algorithm adjusts the polarization direction of the signal to fit the new signal strength differences.

Then, we discuss why in some regions the ϕ_p distribution of background sources increases only within a narrow range, while in other regions the increase occurs over a wider range but with a smaller amplitude. From Figure 2.13, it appears that the distance from the central, antenna-dense region (where the reference antenna is located) may be the cause. For R = 27 km and $\phi_a = 90°$, the zenith angle θ_a with respect to

the reference antenna is

$$\theta_a = \arctan\left(\frac{\sqrt{N^2 + E^2}}{h}\right) = 77.16°,$$

while for the region at $R = 39$ km, θ_a is 81.15°. These angles fall within the range where changes in the antenna function have the largest effect. The region at $R = 27$ km is closer to the antenna-dense area, meaning that the signals received are likely stronger than those at $R = 39$ km. After proportionally reducing the signal strength, the reconstruction process requires more adjustments to accommodate these changes.

Region A, as shown in Figure 2.13, is surrounded by more densely packed stations. Nearby antennas in stations RS409, RS307, and RS310 are distributed in completely different directions, forming a triangular pattern. In the final reconstruction, whether using the original antenna function or $f(\theta_a \mid 0.9)$, these three antennas always receive much stronger signals than the others, and the distances between Region A and these antennas are very similar. Thus, even when the antenna function is changed, the difference in signal strength between them remains minimal, resulting in only minor changes in the reconstruction. Additionally, the antennas are distributed in various directions around Region A, unlike the other three regions where most antennas are located on the same side. This distribution may explain why the angles in the concentrated region of the ϕ_p distribution change with adjustments to the antenna function. For Region A, although, as previously mentioned, the signal strength from the strongest stations—RS409, RS307, and RS310—does not vary significantly with changes in the antenna function, the signal differences across antennas in various directions may lead the program to favor different angles during reconstruction. In contrast, for the other three regions, the changes induced by the antenna function primarily affect antennas on one side, thus maintaining the same preferred angles as before. This suggests that the non-uniformity of the polarization direction distribution may also be influenced by the spatial distribution of the antennas. More detailed analysis of this will be conducted in Sections 3.

2.4 Analyzing Discrete Bright Sources

In the previous sections, we observed that the reconstructed polarization of the background is sensitive to modifications of the antenna function. Therefore, we aim to determine whether the same sensitivity applies to discrete sources. Due to time constraints, we are unable to investigate the entire lightning event. Instead, we selected a set of discrete sources located at t = 825.835 ms, N = -12.2230, E = -27.1079, and h = 7.7250. These sources were chosen because they are spatially isolated from all other sources, making them an ideal and clean test case for this analysis. This will allow us to examine the effects of antenna function modifications on discrete sources without interference from nearby sources.

The strategy for analyzing discrete sources is different with the one used for background analysis in the previous section. Here, the focus is on individual data points rather than overall distribution. By setting a threshold for intensity, data points exceeding this threshold are considered as sources. These sources are then analyzed using different antenna functions to analyse the changes in intensity, position, and polarization.

Firstly, the variations in intensity and position are plotted, as shown in Figure 2.14. By setting a threshold for Intensity, five discrete sources are identified. These sources are then labeled from 1 to 5 in descending order based on their intensity in the original data. In Figure 2.14, the size of the circles corresponds to the intensity, while the colors represent different antenna functions. It is evident that the intensity changes are consistent with those discussed in the background analysis: the antenna function multiplied by a larger changed function have higher intensities in the result. For sources 2-5, the positions calculated using different antenna functions are almost identical. The only exception is for source 1, which is also the source with the highest intensity. The positions calculated using different antenna functions are clustered
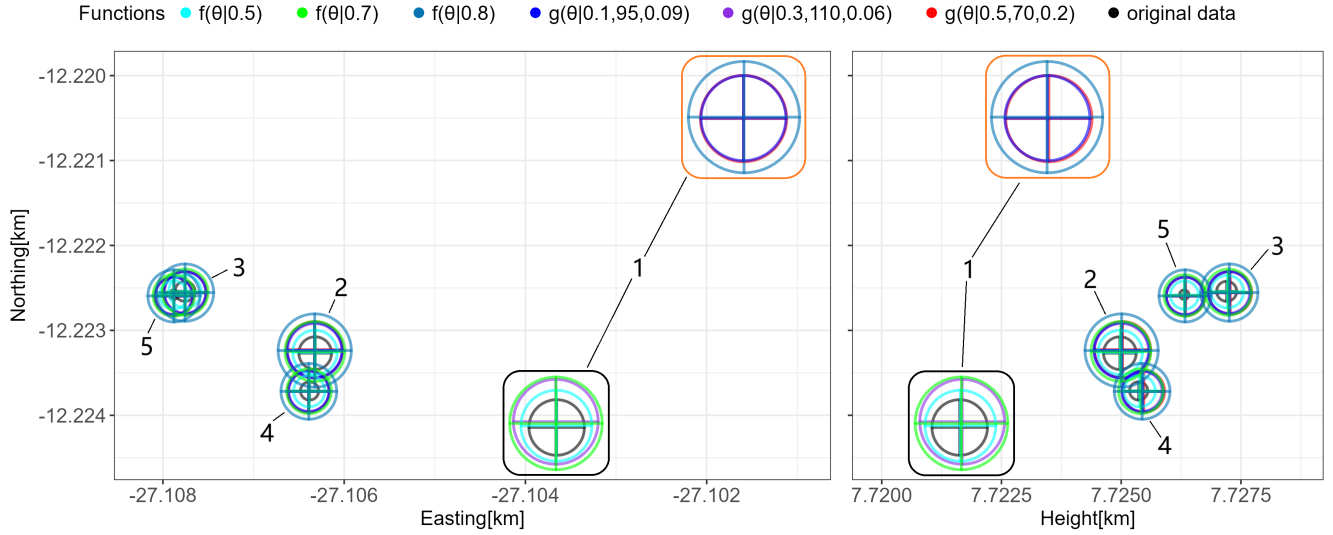
Figure 2.14: Position of 5 discrete sources. Different colors represent different antenna functions, and the numbers indicate the source identifiers. The black and orange rounded boxes highlight the source that share the same identifier.

in two different regions, marked by the black box and the orange box in Figure 2.14. The two regions differ by approximately 4 meters in horizontal distance and about 10 meters in height. While these differences may seem significant in this figure, they are minor when compared to the typical scale of a lightning leader, which spans roughly 1000 meters.

This is a surprising phenomenon, and the position change is discrete, with no positions found between the two regions. Therefore, it is considered to be an error that occurs when the antenna function meets certain conditions. However, it is difficult to understand why it only affects the source with the highest intensity. Besides it, from the position of sources 2 to 5 with different functions, we can conclude that using suitable θ_a -dependent antenna functions does not affect the position of the source.

To further understand the sudden positional shift of the highest intensity source in Figure 2.14, additional functions around the functions that caused this change were selected for further processing, as shown in Figure 2.15. In the left panel, the black line represents the antenna functions causing the positions of sources distributed within the black box in Figure 2.14, while the orange line represents those causing it to be distributed within the orange box. Different colors in the right panel represent different functions. The most significant difference between the black and orange lines is found in the red boxed area in the left panel, also shown in an enlarged view. The orange line is lower than the black line in the interval where θ_a equals 78.5° to 81.5°. Thus, a hypothesis is that: if the function value at a certain angle within this interval is less than a value between 0.538 and 0.575, it may lead to erroneous output results.

Figure 2.16 shows the impact of different antenna functions on the polarization direction of discrete sources. The figure is divided into three parts, each representing one of the three polarization axes. Different colors correspond to different antenna functions, with the x-axis representing $\cos(\theta_p)$ and the y-axis representing ϕ_p. The size of each circle represents the intensity of a source, with the factor used to scale the circle size varying between the main, secondary, and tertiary axis plots to ensure that sources in the secondary and tertiary axis plots are still visible despite having lower intensities.

First, it is clear that the intensity of the discrete sources changes with different antenna functions in the same manner as observed with the background sources. Secondly, we observe that, for most antenna functions, the polarization direction of the discrete sources changes continuously as the antenna function
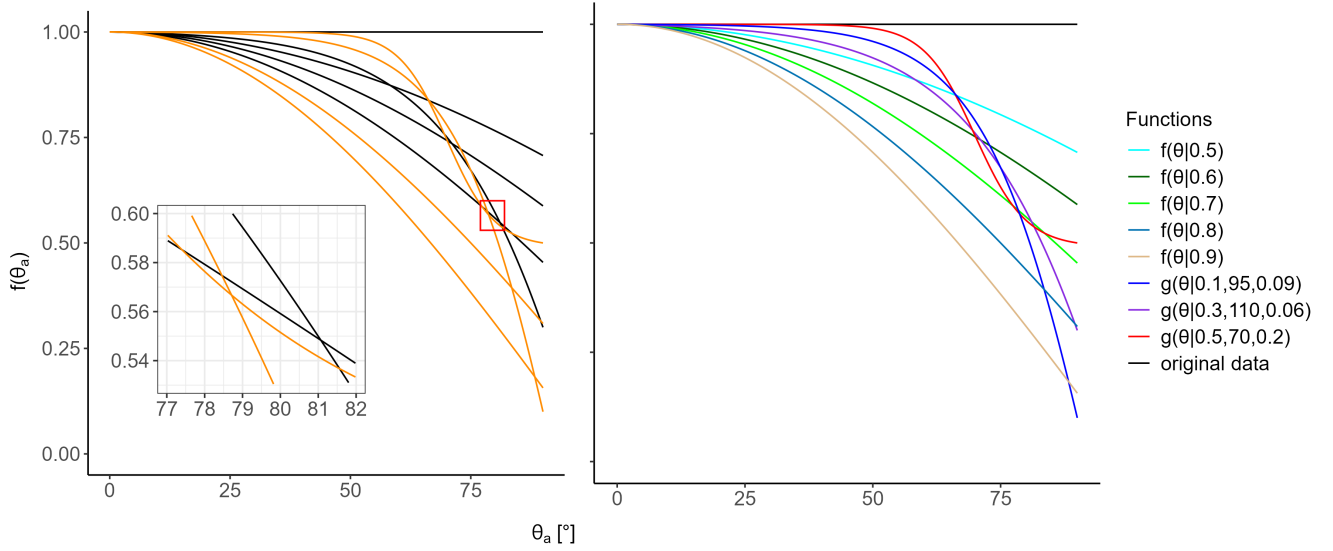
Figure 2.15: Different θ_a-dependent functions. In the left panel, the black curve corresponds to the reconstructed position of Source 1 within the black rounded box, while the orange curve corresponds to the reconstructed position of Source 1 within the orange rounded box. A zoomed-in view of the red box is shown in the bottom-left corner of this panel, where the orange curve is consistently lower than the black curve. The right panel uses different colors to indicate the various functions. This figure is identical to Figure 2.2 and is placed here for comparison purposes.
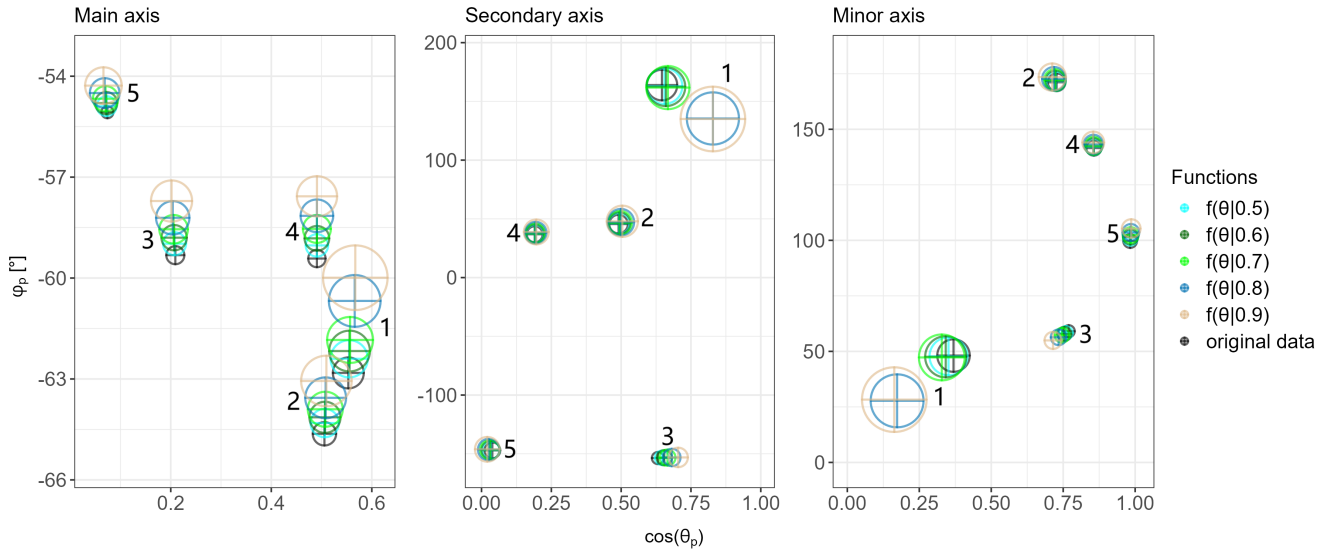

Figure 2.16: Polarization directions of 5 discrete sources, with each of the three panels corresponding to one of the three polarization axes. The numbers indicate the source identifiers. Different colors represent the use of various antenna functions, and the size of the markers is proportional to the intensity of each source. Due to the large difference in average source intensity across the different axes, the marker sizes corresponding to the intensity are scaled differently for each axis for visualization purposes.

changes. This issue, as discussed previously, may arise because the angle θ_a of the source relative to different antennas is not the same. When we multiply the antenna function by a θ_a-dependent function, the signals received by more distant antennas decrease more significantly, causing the program to adjust the polarization direction to fit the relative strength changes between antennas.

Looking at the main axis plot, the magnitude of this change is very small: $\cos(\theta_p)$ remains nearly unchanged, while ϕ_p varies by only about 3°. Another noteworthy point is the discrete shift we observed for Source 1 (the source with the highest intensity) in the position plot from Figure 2.14, which can also be seen in the polarization direction plots. This source shows a slight discontinuity between the green circle($f(\theta_a \mid 0.7)$) and blue circle($f(\theta_a \mid 0.8)$) in the main axis figure, and this change is especially noticeable in the secondary and tertiary axis plots.

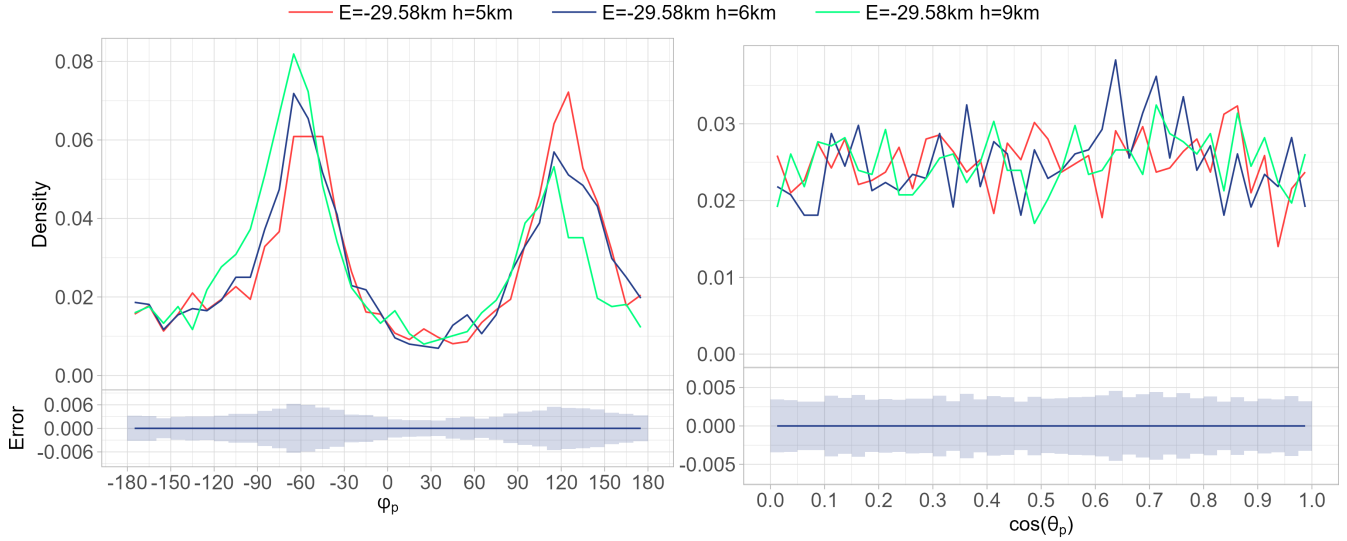# 3 Location Dependence of Background Polarization

Figure 3.1: Polarization direction distributions for the main axis. The left panel corresponds to ϕ_p(binwidth=10°), and the right panel corresponds to $\cos(\theta_p)$(binwidth=0.025). Different colors represent different regions. The same horizontal coordinates (E = -29.58 km, N = 0) are used, with varying altitudes of $h = 5$ km, 6 km, and 9 km.

In the previous section, we established that the polarization direction distribution of the background is location-dependent. To explore this in more detail, we compare the polarization direction distributions across different regions. First, Figure 3.1 shows the polarization direction distribution at E = -29.58 km, N = 0 at different heights. It can be observed that as the height of the region increases, the entire ϕ_p distribution shifts steadily to the right, while the $\cos\theta_p$ distribution remains largely unchanged.
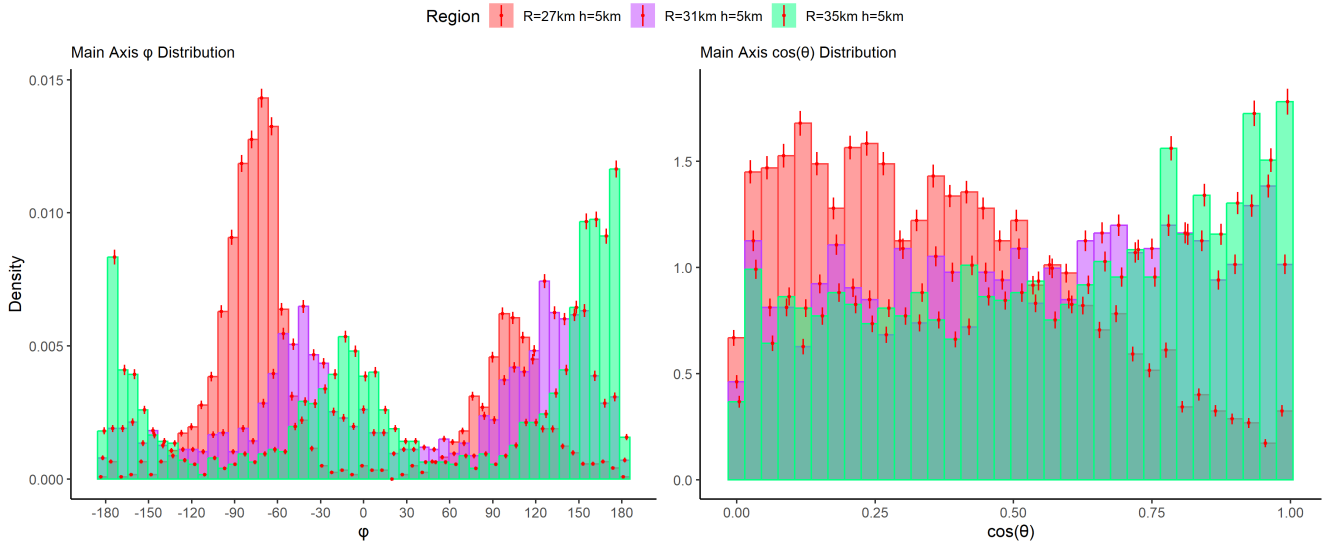

Figure 3.2: Polarization direction distributions for the main axis. The left panel corresponds to ϕ_p, and the right panel corresponds to $\cos(\theta_p)$. Different colors represent different regions. A constant altitude of $h = 5$ km is used, while the horizontal coordinates vary: $R = 27$ km (E = -26.53 km, N = 0), $R = 31$ km (E = -30.59 km, N = 0), and $R = 35$ km (E = -34.64 km, N = 0).

Figure 3.2 illustrates the changes in the polarization direction distribution at a fixed height but varying radii (R) from the reference antenna. The figure shows a clear pattern in the changes: as R increases, the ϕ_p distribution shifts to the right, with the peak on the right becoming more pronounced while the peak on the left diminishes. For the $\cos\theta_p$ distribution, as R increases, the left half decreases while the right half increases.

3.1 Background Analysis at Different Azimuth Angles

We aim to determine whether the location-dependent polarization direction distribution differences mentioned above are related to the azimuth angle ϕ_a relative to the reference antenna. To do this, we selected the reference antenna as the origin of our coordinate system and explored regions located on circles with radii R equal to 27 km and 39 km, both at an altitude of 6 km. To save time, we only explored voxels starting at $\phi_a = 0°$ and incrementing by 15 degrees.

To compare the polarization direction distribution differences for regions with the same ϕ_a but different radii, we used the two-sample Cramér-von Mises test. Similar to the previously mentioned Kolmogorov-Smirnov (K-S) test, the Cramér-von Mises test is employed to quantify differences between two samples. However, these tests are suited to different situations due to differences in their calculation methods, which will be discussed below.
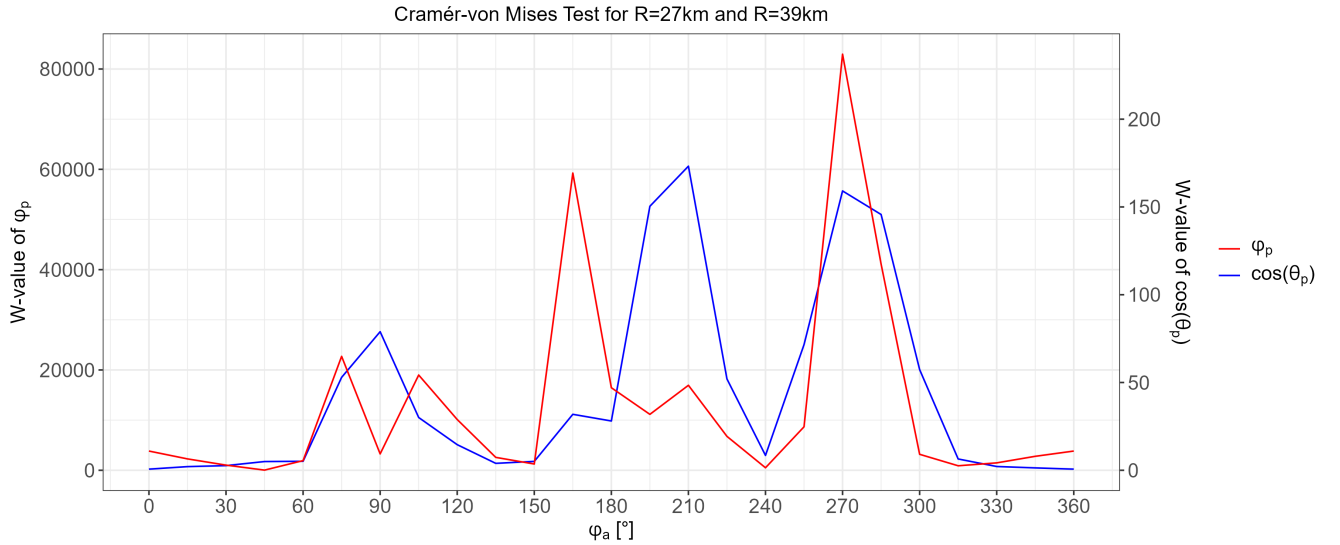

Figure 3.3: Cramér-von Mises test of the polarization data between $R = 27$ km and $R = 39$ km at different ϕ_a. The red line corresponds to the ϕ_p results (left y-axis), and the blue line corresponds to the $\cos(\theta_p)$ results (right y-axis).

Given two sets of samples $x_1, x_2, x_3, ..., x_N$ and $y_1, y_2, y_3, ..., y_M$, where $\hat{F}_N$ and $\hat{G}_M$ represent the empirical cumulative distribution functions (eCDFs) of these two sets, the Cramér-von Mises test [16] calculates the difference between them using the following formula:

$$W = \left[\frac{NM}{N+M}\right] \int_{\infty}^{\infty} \left[\hat{F}_N(x) - \hat{G}_M(y)\right]^2 d\hat{H}_{N+M}(x, y), \tag{5}$$

where $(N+M)\hat{H}_{N+M} = N\hat{F}_N(x) + M\hat{G}_M(y)$.

In summary, the Cramér-von Mises test, unlike the Kolmogorov-Smirnov Test, does not emphasize the maximum difference between empirical distribution functions. Instead, it focuses on the overall differences

24

across the entire distribution and is more sensitive in the latter part of the distribution. Given these characteristics, the Cramér-von Mises test test is better suited to this scenario.

Figure 3.3, shows the results of a two-sample Cramér-von Mises test performed for R equal to 27 km and 39 km under different azimuth angles. The blue line represents the differences in the ϕ_p distribution, while the red line corresponds to the $\cos(\theta_p)$ distribution. Regarding the y-axis scale, the left y-axis is used for ϕ_p, and the right y-axis is for $\cos(\theta_p)$. A higher W value indicates a greater difference between the distributions. It can be observed that in the azimuth ranges of ϕ_a from 0° to 60° and 315° to 360°, the distribution is less sensitive to changes in R. However, outside of these ranges, changes in R lead to significant differences in the distributions. Notably, the angles that are less sensitive to changes in R correspond to areas with a much lower density of antennas. This may suggest that the density of antennas plays a key role in how sensitive the polarization direction distributions are to variations in the radial distance.

3.2   Calibration

By observing the polarization direction distributions in Figure 3.1 and Figure 3.2, it becomes clear that changes in distribution patterns follow certain trends, which may be caused by incorrect calibration. Therefore, in this section, we will attempt to use different calibration files to examine their impact on the polarization direction distribution.

We selected three different calibration files: "Calibration 1," "Calibration 2," and "Calibration 3". These calibrations were obtained from the LOFAR Lightning group and resulted from fitting different selections of lightning sources. They differ by slight timing offsets between the various LOFAR stations, often less than 5ns. A detailed discussion of these calibration differences falls however beyond the scope of the present work. And "Calibration 3" is the calibration file we used in previous analyses and will be abbreviated as C3 in subsequent figures. We explored a circle with a radius of 39 km and a height of 6 km from the reference antenna, and, as in previous studies, we selected voxels every 15 degrees. Upon comparison, we found that the effect of different calibration files on the distribution is also location-dependent. In the following part, we will use several figures to illustrate this phenomenon and demonstrate how different calibration files affect the polarization direction distribution at various locations.

First, for ϕ_a between $-45°$ and 60°, the results obtained from different calibrations are mostly consistent. In Figure 3.4, several regions within this range were selected. The left panel shows the distribution of ϕ_p for these regions under different calibrations, while the right panel shows the distribution of $\cos(\theta_p)$ for the same regions. As we can see, although there are significant variations in the distribution of the two polarization directions depending on the location, the results obtained using different calibrations for the same region are almost completely overlapping. Even the small differences observed are within the margin of error.

In contrast, for ϕ_a between 185° and 200°, the differences between the calibrations are most pronounced. As shown in Figure 3.5, both the position and height of the peak show significant changes. For other regions, different calibration files also cause differences, but the variations are not as significant.

To more comprehensively demonstrate that the effect of calibration files on polarization direction distribution is location-dependent, we need to quantify this difference. While the Cramér-von Mises test could be used as before, this test is most applicable for comparing two samples. Since we want to display the results for three different calibration files simultaneously, performing pairwise Cramér-von Mises tests would be too cumbersome and time-consuming. Instead, standard deviation can effectively capture the characteristics of a distribution. Thus, we calculated the circular standard deviation of ϕ_p for each region using different calibration files. It is important to use circular standard deviation because ϕ_p is an angle
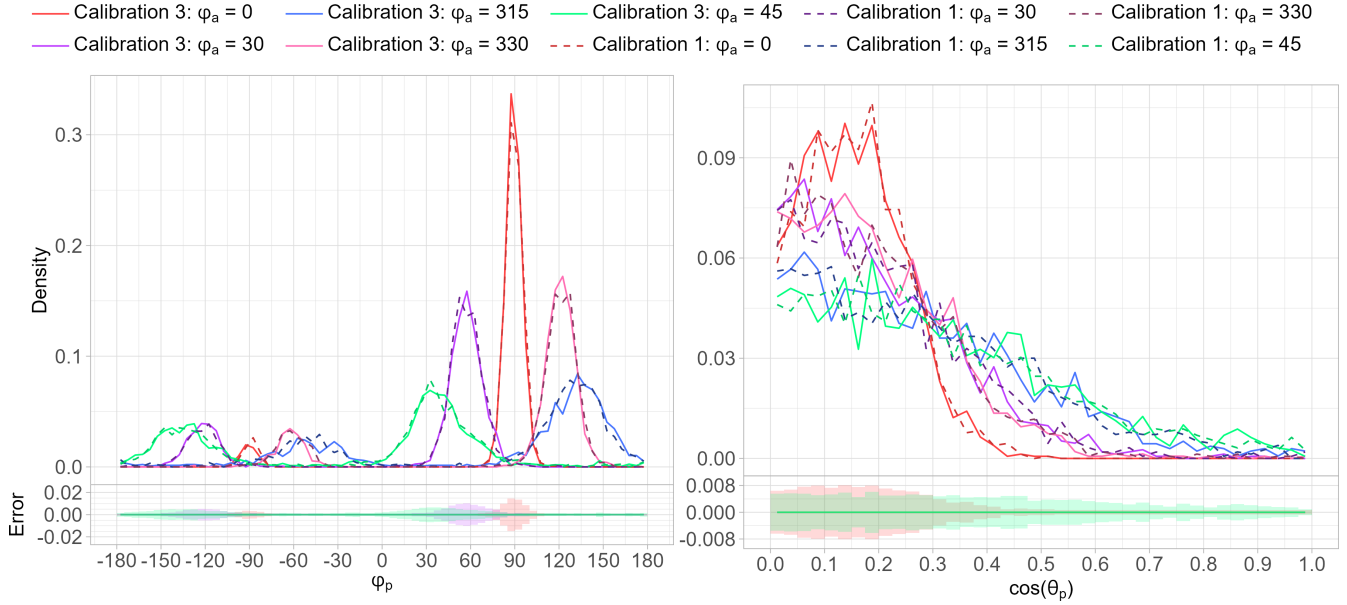
Figure 3.4: Polarization direction distributions at $R = 39$ km around $\phi_a = 0°$. The left panel corresponds to ϕ_p, and the right panel corresponds to $\cos(\theta_p)$. Different colors represent different ϕ_a, with solid lines indicating the use of Calibration 3 and dashed lines representing Calibration 1. The shaded regions below the graphs represent the error bars for the corresponding ϕ_a data, with colors matching the respective ϕ_a.
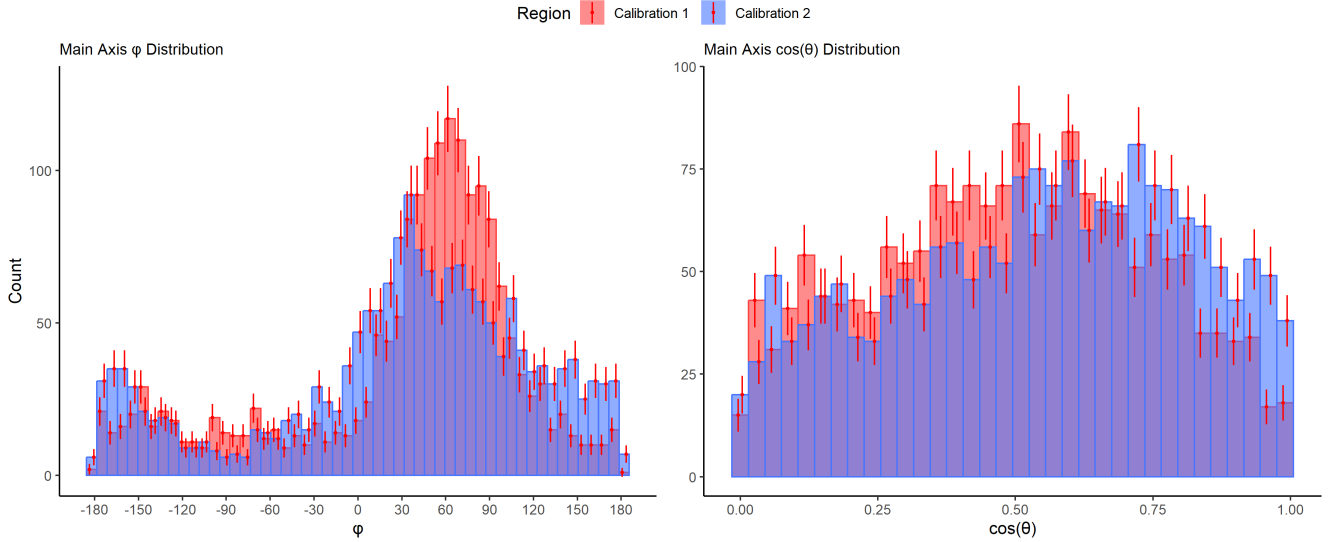

Figure 3.5: Polarization direction distributions at $R = 39$ km and $\phi_a = 195°$. The left panel corresponds to ϕ_p, and the right panel corresponds to $\cos(\theta_p)$. Red bars represent the data using Calibration 3, while blue bars represent the data using Calibration 1.

distributed in the range of $-180°$ to $180°$, where $-180°$ and $180°$ represent the same direction. Using a regular standard deviation might yield misleading results, so we opted for circular standard deviation.

Circular standard deviation is specifically used to represent the standard deviation for angular data. For a set of angles $\phi_1, \phi_2, \ldots, \phi_n$, the circular standard deviation S is calculated as [17]:

$$S = \sqrt{-2\ln(r)}, \tag{6}$$

26

where r is the mean resultant length of this set of angles:

$$r = \frac{1}{n}\sqrt{\left(\sum_{i=1}^{n}\cos(\phi_i)\right)^2 + \left(\sum_{i=1}^{n}\sin(\phi_i)\right)^2}. \tag{7}$$
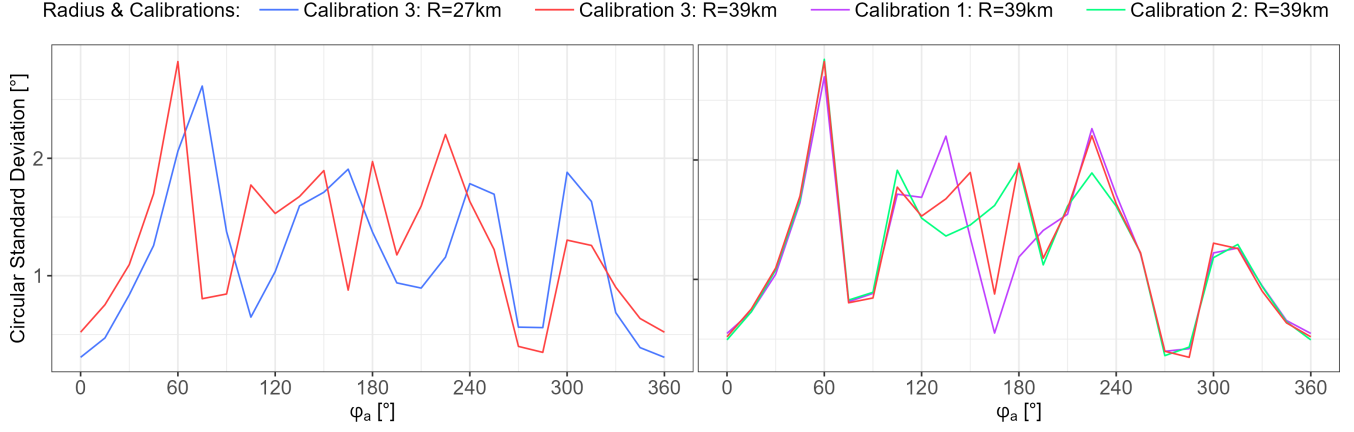

Figure 3.6: Circular Standard Deviation of ϕ_p for different ϕ_a regions. Different colors correspond to varying radii and calibrations. The left panel compares different radii ($R = 27$ km and $R = 39$ km), while the right panel compares different calibrations.

In Figure 3.6, we present the results, divided into two panels. The right panel corresponds to the data from different calibration files, where the x-axis represents different ϕ_a values (to distinguish between regions). The left panel corresponds to the data for different radial distances R, as mentioned in Section 3.1, and is included here for better comparison. First, we observe that the impact of changing the calibration file on the ϕ_p distribution is significantly smaller than the impact of changing the radial distance R. This is particularly true for regions with ϕ_a outside the range of 105° to 225°, where the calibration has virtually no effect. However, within this range, especially around 180°, there are noticeable differences in the ϕ_p distribution due to calibration changes.

Looking at the ϕ_p distribution for different locations, as seen in Figure 3.4, regions near $\phi_a = 0°$ and its vicinity exhibit much more non-uniform polarization direction distributions compared to other regions, yet the impact of calibration on these regions is minimal. Therefore, we can conclude that calibration is not the cause of the non-uniformity in the polarization direction distribution. Additionally, similar to what was discussed in Section 3.1, the regions most affected by calibration changes are those with higher antenna density. A similar conclusion can be drawn regarding calibration: in regions with denser antenna placement, the effect of calibration changes is more pronounced.

3.3 Uniformity Analysis

As previously mentioned, we divide the entire space into multiple rectangular voxels and analyze each voxel individually. In this study, the voxel used is a rectangular prism with a length and width of 0.05 km and a height of 0.3 km. As shown in Figure 3.7, after examining the position distribution of background sources within each rectangular region, we found that these sources are concentrated along the edges of the voxels. In this figure, the bin width for Easting and Northing is 0.003 km, while the bin width for height is 0.015 km. The errors are calculated as $\sqrt{N+1}$. This uneven distribution is unusual, as the theoretical expectation is that the background source positions should be uniformly distributed.
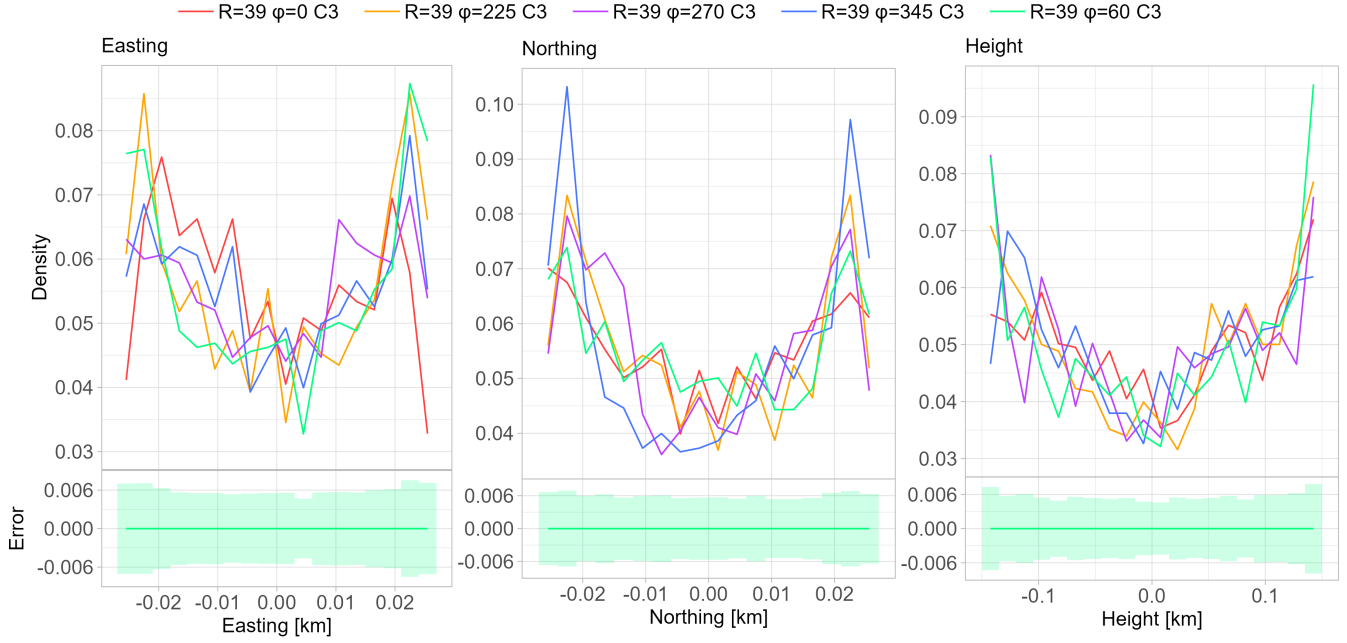
Figure 3.7: Position Distributions. The three panels represent Easting, Northing, and Height, respectively. Different colors represent different regions. The green shaded area below corresponds to the error for $R = 39$ km, $\phi_a = 60$, using Calibration 3 (C3).

Additionally, as seen in Figure 3.7, similar to the previously discussed polarization direction distribution, the degree of unevenness in the position distribution also depends on the specific location of the region being analyzed. Furthermore, the distribution pattern varies across the three dimensions (Easting(E), Northing(N), and Height(h)). Just because a region shows the most uneven distribution along the Easting dimension, it does not necessarily mean that the same region will have the most uneven distribution in the Northing or Height dimensions.

To gain further insight into this phenomenon, similar to our approach with the polarization direction, we examined how changing the antenna function impacts the position distribution. The results are shown in Figure 3.8, which explores the same region as Section 2.1, identified as Region A. As can be seen, altering the antenna function has a smaller effect on the position distribution compared to its impact on polarization direction, with changes falling within the error margins.

Since modifying the antenna function generally has a limited effect on the position distribution, and the location of the region has a much more significant influence, this section will investigate the combined effects of polarization direction distribution and position distribution of background sources in different regions, along with the relative positions of these regions to all antennas.

3.3.1 Method for Quantifying Non-uniformity

To quantify the differences in distribution across various regions, we focus on comparing the degree of non-uniformity in these distributions. The degree of non-uniformity can be represented in different ways, depending on the type of distribution. In this study, we choose to calculate the standard deviation of multiple points along the distribution curve. More specifically, as shown in Figure 3.7, each line represents the density distribution for a specific location with a specified bin width. We extract the density values corresponding to each interval and denote them as $d_1, d_2, d_3, \ldots, d_n$, where d_i represents the density of
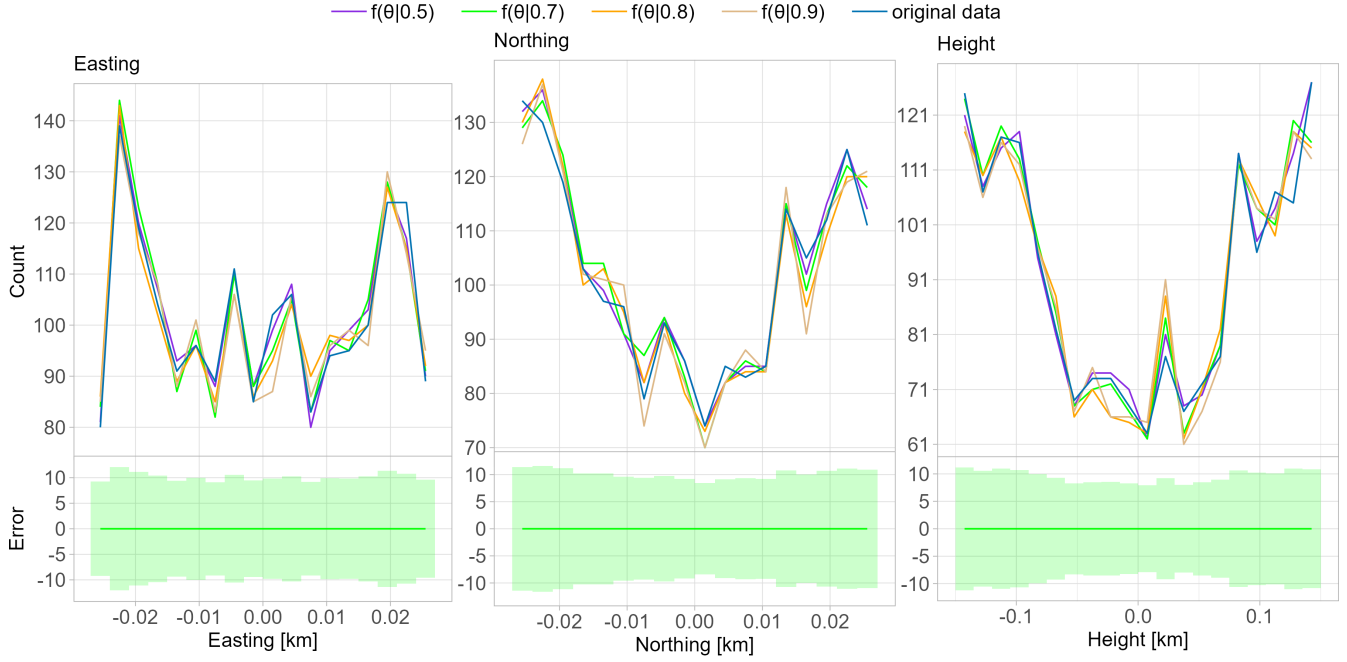
Figure 3.8: Position Distributions. The three panels represent Easting, Northing, and Height, respectively. Different colors represent the different antenna functions used. The green shaded area below corresponds to the error for the results calculated with $f(\theta_a \mid 0.9)$.

the i-th interval. The standard deviation D is then calculated as:

$$
\mathrm{D} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (d_i - \bar{d})^2}
\tag{8}
$$

where $\bar{d}$ is the mean of $d_1, d_2, \ldots, d_n$. A larger standard deviation indicates greater fluctuations in the curve, meaning a higher degree of non-uniformity. In the following sections, we calculate the standard deviation for ϕ_p, $\cos(\theta_p)$, E, N, and h at different locations using the method in eq. 8, and interpret these values as the non-uniformity of the corresponding quantity at each location. One drawback of this method is that the choice of bin width can significantly influence the distribution curve, so it is essential to select an appropriate bin width, although this can introduce larger errors.

There are, of course, other methods to represent the degree of non-uniformity. We will discuss two alternative approaches and explain why we chose not to adopt these methods, opting instead for the approach described above:

A common approach is to directly calculate the standard deviation of all the data points. That is, instead of calculating the standard deviation for $d_1, d_2, d_3, \ldots, d_n$ as described above, we calculate the standard deviation directly for values such as $\phi_{p1}, \phi_{p2}, \ldots, \phi_{pn}$ to provide measure of uniformity for ϕ_p at each location. However, we believe this method is not well-suited for the type of distributions we are studying. For position distributions, most regions exhibit a characteristic pattern where there are two concentrated areas on either side. In the most typical cases, the distribution shows two peaks of similar size on the curve, with the average value falling in the central region. Standard deviation primarily increases as both the number of points deviating from the mean and the degree of deviation rise. In the type of distribution we are studying, the standard deviation would not only increase with rising non-uniformity (i.e., as the two peaks grow) but would also increase as the peaks shift further away from the mean. This means that the standard deviation is influenced by both the concentration of values at the extremes (which reflects

non-uniformity) and the distance of those peaks from the mean, making it a less ideal metric for capturing pure non-uniformity in such cases.

This approach has other issues as well. In the typical case mentioned, a larger standard deviation would indeed suggest a more uneven distribution. However, in some regions where the distribution is concentrated on only one side, such as the purple line in the Easting plot of Figure 3.7, the concentration is primarily on the right side. For such a distribution, a smaller standard deviation could paradoxically indicate greater non-uniformity. Therefore, calculating the standard deviation of all points' values directly is not suitable for position distributions because it represents different meanings for single-peak and double-peak distributions.

For polarization direction distributions, a similar problem arises. In the azimuth polarization direction, $-180°$ and $180°$ represent the same direction, so the standard deviation cannot be calculated in the usual way. Instead, circular standard deviation should be used. Moreover, similar to the position distribution, azimuth polarization direction curves often exhibit two peaks, spaced $180°$ apart. When the two peaks are of similar size, the standard deviation will be larger than for a uniform distribution. However, when there is a significant height difference between the two peaks, the standard deviation could be smaller than for a uniform distribution, making it an unsuitable measure of non-uniformity in such cases.

For zenith polarization direction, the standard deviation is an appropriate measure of non-uniformity, as a larger standard deviation correlates with a more uniform distribution. However, for consistency in comparison across different types of distributions, it is preferable to use the same metric for "non-uniformity." Therefore, while standard deviation works for some cases, we aim to establish a consistent standard for quantifying non-uniformity across all the distributions we study.

In addition to directly calculating the standard deviation of all points, another common method is to use statistical tests like the Kolmogorov-Smirnov test (K-S test) or similar approaches. Specifically, this involves comparing each distribution to a uniform distribution using the K-S test(eq.4) or a similar test. A larger test statistic would indicate a greater difference from a uniform distribution, and this method does not rely on the choice of bin width.

However, such tests, especially those based on comparing cumulative distribution functions (CDFs), tend to be more sensitive to differences in the head of the distribution and less sensitive to differences in the tail. This limitation is particularly pronounced in the K-S test. As a result, two distributions with similar degrees of non-uniformity might produce different results simply because one is concentrated in the early part of the range while the other is concentrated later. While such results have their own significance, our goal is to focus solely on the "degree of non-uniformity" without being influenced by other factors like where the concentration occurs within the distribution. Therefore, we have decided not to adopt this method in favor of a more direct approach that isolates the non-uniformity characteristic.

3.3.2   Overview of Polarization Direction Distribution

Figure 3.9 provides a more intuitive understanding of the relationship between the non-uniformity in polarization direction distribution and the locations of the antennas. The figure is divided into two parts, on both panels the numbers indicate LOFAR station numbers and the dots indicate the locations where the background has been analyzed. The color of each point reflects the degree of non-uniformity (as discussed in Sections 3.3.1) in the $\cos\theta_p$ distribution for that region. On left panel, there is a green arrow extending from each point, whose direction represents the angle corresponding to the highest peak in the density distribution of the polarization direction ϕ_p, and the arrow's length is proportional to the density at that angle. To help understand this, Figure 3.10 shows the density distribution of the polarization directions for several regions, with the green dot marking the peak position in the ϕ_p section on the left.
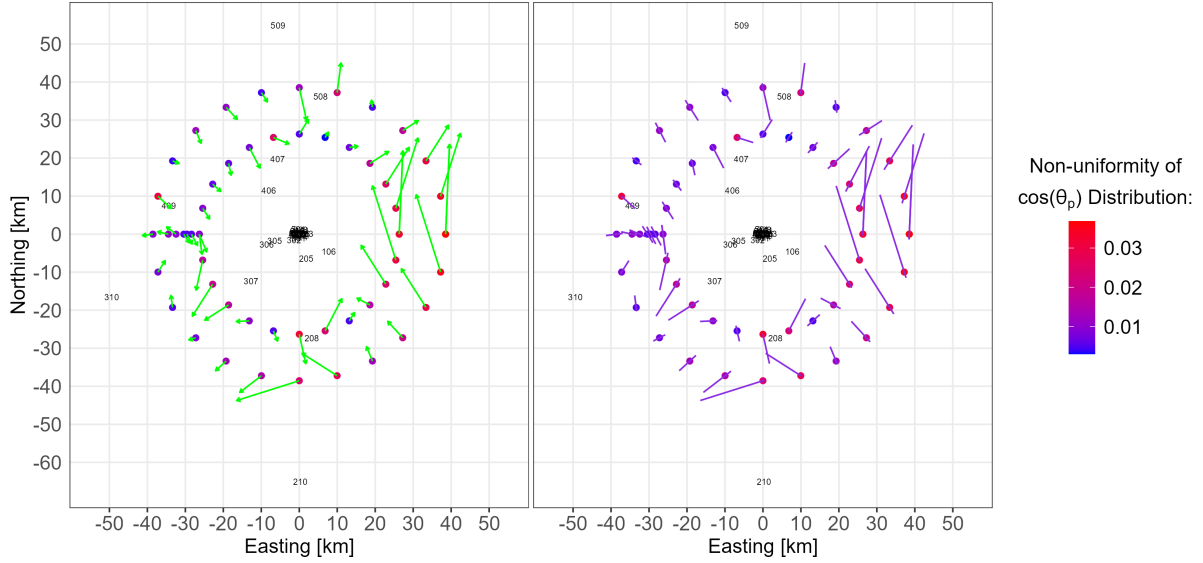
Figure 3.9: The numbers represent LOFAR station numbers, and the dots represent different regions. The color of each dot indicates the non-uniformity of the $\cos(\theta_p)$ distribution. Green lines (left panel) and purple lines (right panel) extend from each dot, representing the polarization direction of concentration and non-uniformity of the ϕ_p distribution under different algorithms (see text for details).
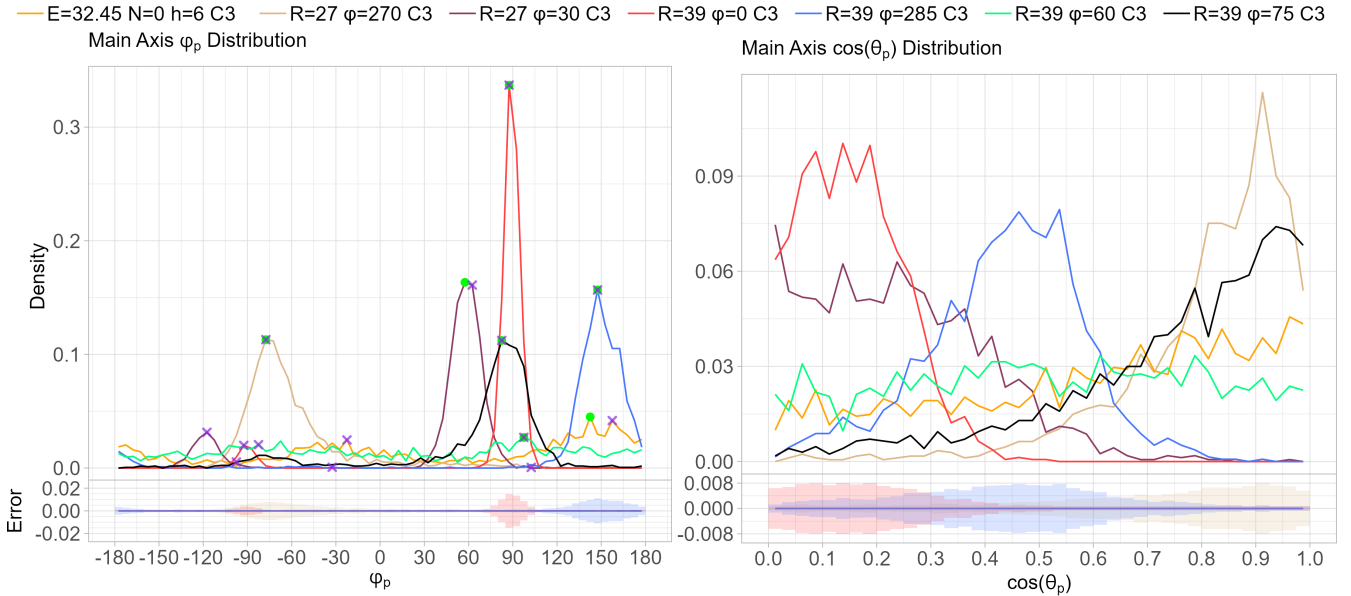

Figure 3.10: Polarization direction distributions. The left panel corresponds to ϕ_p, and the right panel corresponds to $\cos(\theta_p)$. Different colors represent different locations. The green dot indicates the direction pointed by the green arrow in Figure 3.9, and the purple cross corresponds to the direction indicated by the purple line in Figure 3.9. The shaded regions below the graphs represent the error bars for the corresponding locations, with colors matching the respective data.

In the right panel of Figure 3.9, purple lines are used instead of green arrows. This is because the ϕ_p distribution of polarization direction typically has two peaks, spaced 180° apart. The purple line represents the direction of the angle corresponding to the maximum value obtained by summing the density at one angle and its opposite (i.e., the angle ±180°). The length of the purple line corresponds to the combined density of both angles. In the left panel of Figure 3.10, the purple cross marks this calculated position. In

31

most regions, the green arrow direction roughly aligns with one of the directions of the purple line, but in regions where the two peaks are almost equal in size, such as in the case of the orange line, they may not coincide. In the right panel of Figure 3.10, clear differences in the θ_p uniformity across various regions can be observed. For regions with a non-uniform distribution of $\cos(\theta_p)$, the θ_p distribution exhibits a single, distinct peak. The subsequent sections will delve into the relationships between this uniformity and the antenna positions, as well as between the θ_p value at the peak and the antenna configurations.

3.3.3   Uniformity of Polarization Direction Distribution

In this section, we will focus on analyzing the non-uniformity of the polarization direction distribution. The preference in the background source polarization distribution (i.e., the angle or cosine value corresponding to the peak in the polarization distribution shown in Figure 3.10) will be discussed in detail in Section 3.3.4. In Figure 3.9, we observe that regions where stations are distributed only on one side tend to have higher non-uniformity in both of phi_p and $theta_p$ distribution. Additionally, we can see that regions very close to the stations exhibit higher non-uniformity in the $\cos(\theta_p)$ distribution (indicated by the red color). Therefore, in exploring the relationship between the uniformity of the polarization direction distribution and the position of the antennas, we define two quantities to characterize the antenna distribution:

- **Distance to Nearest Antenna**: This is the distance from a given region to the closest antenna.

- **Maximum Angle Between Adjacent Antennas (M-value)**: This is calculated by first determining the angle of all antennas relative to the region, then finding the maximum difference between consecutive angles when sorted in ascending order. A larger value indicates that the antennas are distributed within a narrower angular range.
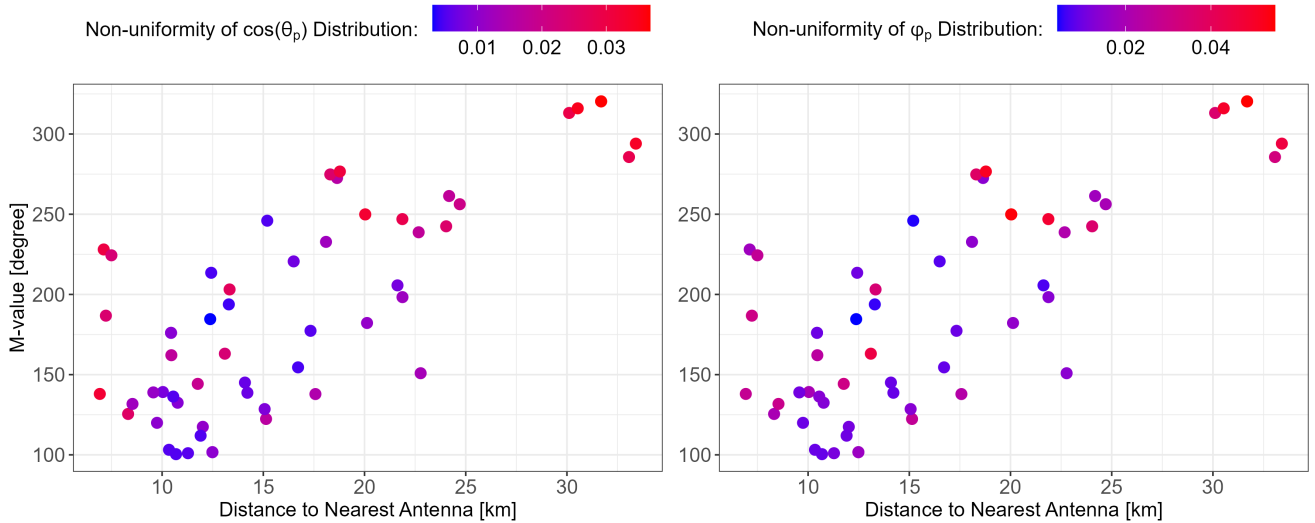

Figure 3.11: Relations between the non-uniformity of polarization direction distributions and antenna positions. Darker red indicates greater non-uniformity. The left panel shows $\cos(\theta_p)$, and the right panel shows ϕ_p.

Figure 3.11 shows the relationship between the Distance to Nearest Antenna and M-value for all regions, with the non-uniformity of polarization directions distribution represented by different colors in the left and right panels for each polarization direction. First, we observe that in most regions, a greater Distance

to Nearest Antenna corresponds to a larger M-value, but there are some exceptions, which justifies analyzing these two quantities separately. Second, the color indicates that regions with a very small Distance to Nearest Antenna exhibit significant non-uniformity in their polarization directions distribution. Moreover, the non-uniformity increases as both the M-value and Distance to Nearest Antenna grow. However, beyond this trend, it is difficult to identify more detailed patterns from the colors alone, especially for regions with similar non-uniformity values.
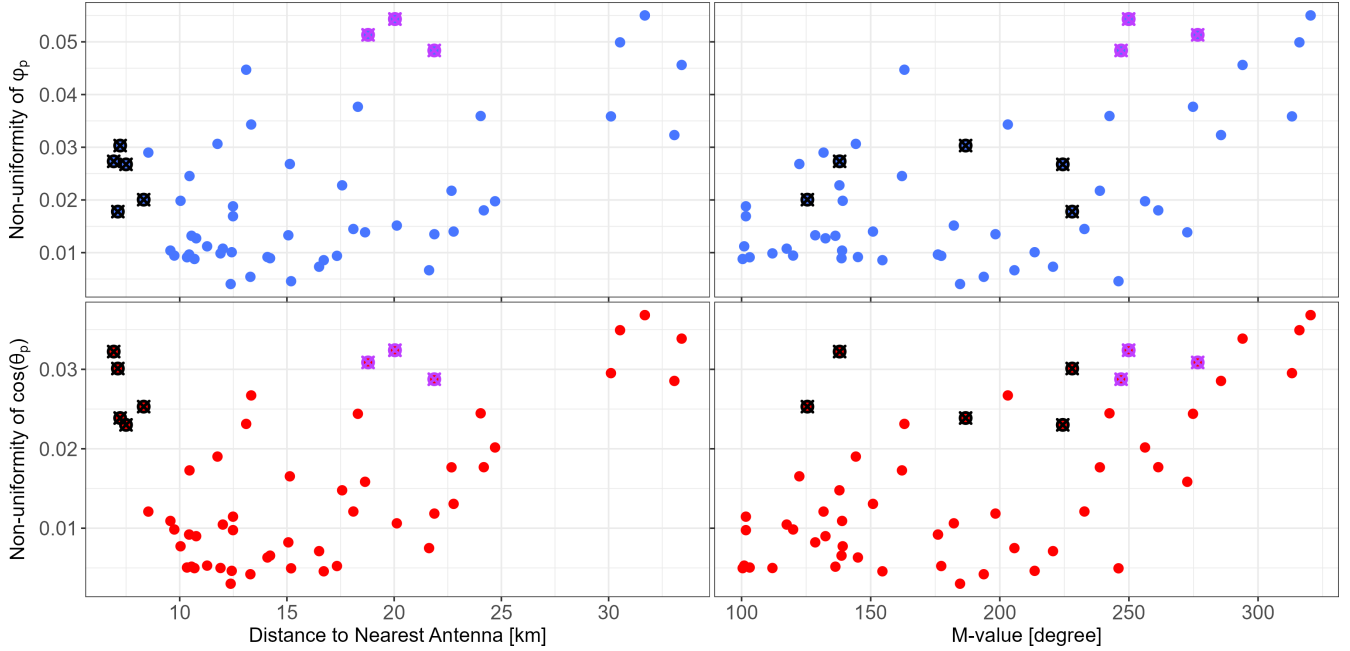

Figure 3.12: The two left panels show the relationship between the distance to the nearest antenna and the non-uniformity of the two polarization direction distributions, while the two right panels show the relationship between the M-value and the non-uniformity of the polarization direction distributions. Blue dots represent the ϕ_p data, and red dots represent the $\cos(\theta_p)$ data. Black squares highlight points very close to the antennas, while purple squares mark points that deviate from the overall trend in the distance to the nearest antenna plots.

To facilitate a more intuitive analysis, Figure 3.12 provides a more detailed view of these relationships. There are four panels, with blue points representing ϕ_p data and red points representing $\cos(\theta_p)$ data. The two left-hand panels plot the Distance to Nearest Antenna against non-uniformity for both polarization directions, while the two right-hand panels show the relationship between the M-value and non-uniformity. It's important to note that in calculating the M-value, antennas farther than 47 km were excluded, as distant antennas contribute less to the signal strength. The justification for this threshold will be discussed in detail later.

Looking first at the $\cos(\theta_p)$ data (red points), we see that in regions that are neither too close to the antennas nor have an overly narrow M-value, the non-uniformity distribution is scattered. Although there is a weak trend suggesting that non-uniformity increases with both quantities, it is not very pronounced, and there are some outliers.

To investigate further, black squares in Figure 3.12 mark regions that are extremely close to antennas (with a distance to the nearest antenna less than 8.5 km), while purple squares highlight outliers in the Distance to Nearest Antenna plot. The selection criteria for purple squares are not strict; they simply indicate points that deviate from the overall trend for the purpose of further analysis. The black squares reveal that many of the outliers in the M-value plot originate from regions near antennas. When these

points are excluded, a clearer trend of increasing non-uniformity with higher M-values emerges. As for the purple squares, all of them exhibit large M-values, suggesting that this may explain why these points deviate from the trend in the Distance to Nearest Antenna plot. Once the black and purple points are removed, the overall upward trend becomes more pronounced.

Now considering the ϕ_p data (blue points), we find similar features to the $\cos(\theta_p)$ data, with one key difference: regions very close to antennas (black squares) show much less non-uniformity in ϕ_p compared to $\cos(\theta_p)$. This suggests that the distribution of $\cos(\theta_p)$ is more strongly affected by both proximity to antennas and the angular distribution of antennas, while ϕ_p is primarily influenced by the angular distribution of antennas. This pattern is also evident in Figure 3.11, where most points have similar colors, but those closest to antennas are distinctly bluer in the ϕ_p plot than in the $\cos(\theta_p)$ plot.
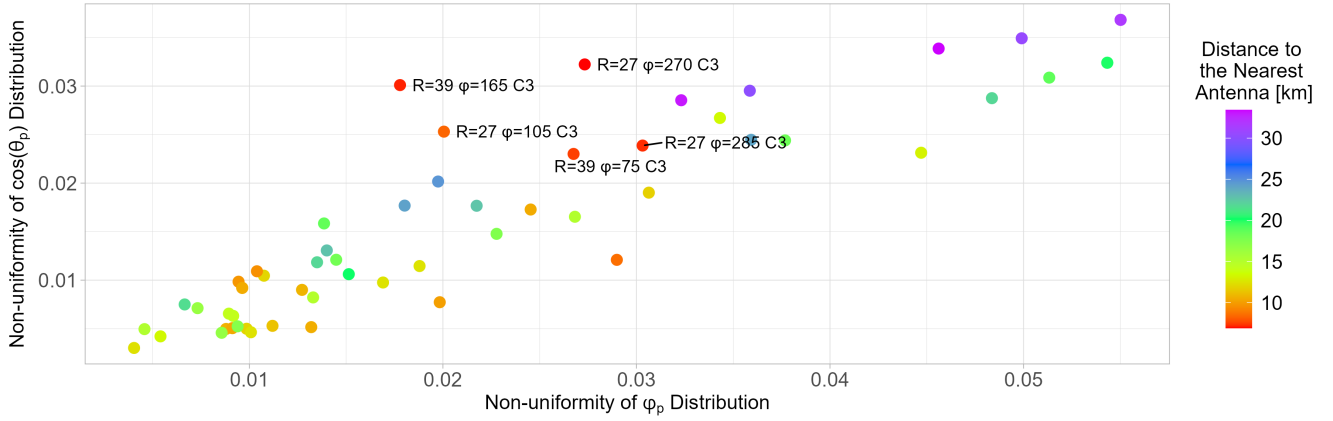

Figure 3.13: Relations between the non-uniformity of the two polarization direction distributions, with different colors representing the distance to the nearest antenna. Points where the distance to the nearest antenna is less than 8.5 km are labeled.

In Figure 3.13, the relationship between the non-uniformity of the two polarization directions is plotted, with color indicating distance to the nearest antenna, and labels marking regions where Distance to Nearest Antenna is less than 8.5 km. Excluding these regions near antenna, a strong positive correlation is evident: the higher the non-uniformity in ϕ_p, the higher the non-uniformity in $\cos(\theta_p)$.

As previously mentioned, signals from distant antennas are generally weaker and may have less influence on the polarization direction and position distribution. To account for this in our study, we considered adding distance-dependent weighting functions when calculating the M-value. However, since we were unsure how to optimally define such a weighting function and for simplicity, we instead applied a threshold distance, excluding antennas beyond this threshold when calculating the M-value.

Figure 3.14 shows the effect of excluding antennas beyond different distances on the "M-value vs. Non-uniformity" plot. The number in the top left corner of each panel represents the distance (in kilometers) beyond which antennas were excluded, and "All" indicates no threshold was applied.

In the panels for 58 km and All, we see that without significant filtering, the previously discussed trend is not evident, and points are scattered irregularly in the range of $100° - 250°$. Similarly, when smaller thresholds, such as 35 km, are used, the points are still distributed randomly, spanning the range of $100° - 360°$ without any clear pattern. However, when a threshold between 41 km and 51 km is applied, the trend becomes much more visible, with the overall pattern closely resembling that in Figure 3.12. This suggests that filtering out antennas beyond this range is effective in revealing the trend we had previously identified in the relationship between M-value and non-uniformity.
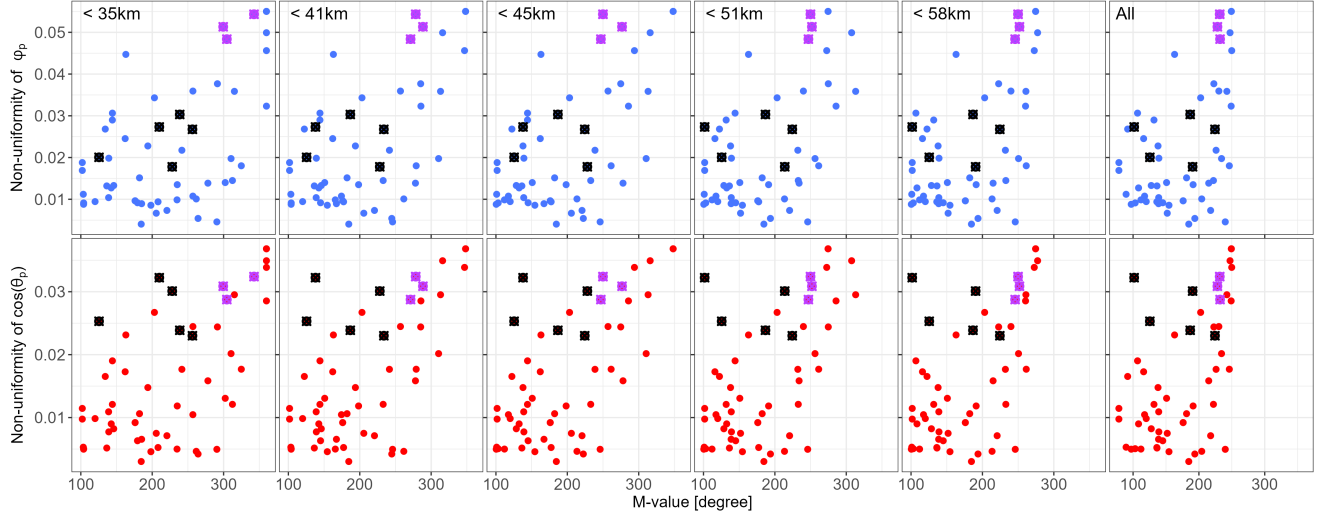
Figure 3.14: Relations between the non-uniformity of the two polarization directions and the M-value. Blue dots represent the ϕ_p data, and red dots represent the $\cos(\theta_p)$ data. Each column represents the results using different thresholds (see text), with the selected threshold indicated in the upper-left corner of each column.

3.3.4 Angular Preferences in Polarization Direction Distribution

In this section, we will examine how antenna positions affect the angular preferences in the polarization direction distribution, which refers to the angle or cosine value corresponding to the peak in the distribution. For simplicity, we will refer to these as the peak of ϕ_p and the peak of θ_p throughout the discussion. We will begin by focusing on the peak of ϕ_p.

From Figure 3.9, we observe that the green arrows tend to align parallel to the distribution of most of the antennas or point perpendicular to the direction with the highest number of nearby antennas. To verify this, we calculated the average angle of all the antennas relative to a given region, represented by black arrows on the left side of Figure 3.15. It is clear that most of the arrows point towards the central area. Since signals from farther antennas are weaker, we applied some weighted functions, shown on the right side of Figure 3.15, to reduce the influence of distant antennas and then recalculated the weighted average angle. Despite this adjustment, the direction of the black arrows remained largely unchanged.

It is evident from the left side of Figure 3.15 that no distinct relationship between the direction of the green arrows and the black arrows is observable. Therefore, further investigation is needed to explore the angle corresponding to the peak of ϕ_p. The reason that the average angle of the antennas in each region tends to point towards the center is likely due to the large number of central stations. We speculate that for the central region, which contains numerous antennas, their proximity may limit their overall influence, meaning that multiple nearby antennas may not exert significantly more impact than a single antenna. Therefore, in future studies, it may be more reasonable to use weighting functions that account for both the density of antennas within a unit area and their distance, effectively reducing the weight assigned to regions with higher antenna concentrations.

Now, let's discuss the relationship between the antenna distribution and the peak position of the $\cos(\theta_p)$ distribution. Figure 3.16 illustrates the relationship between the M-value and the peak position (left) and the Distance to Nearest Antenna and the peak position (right).

As seen in Figure 3.10, for regions with low non-uniformity, the fluctuations in the $\cos(\theta_p)$ distribution are mostly within the error range, making the identification of a "peak" less meaningful. It is difficult to judge
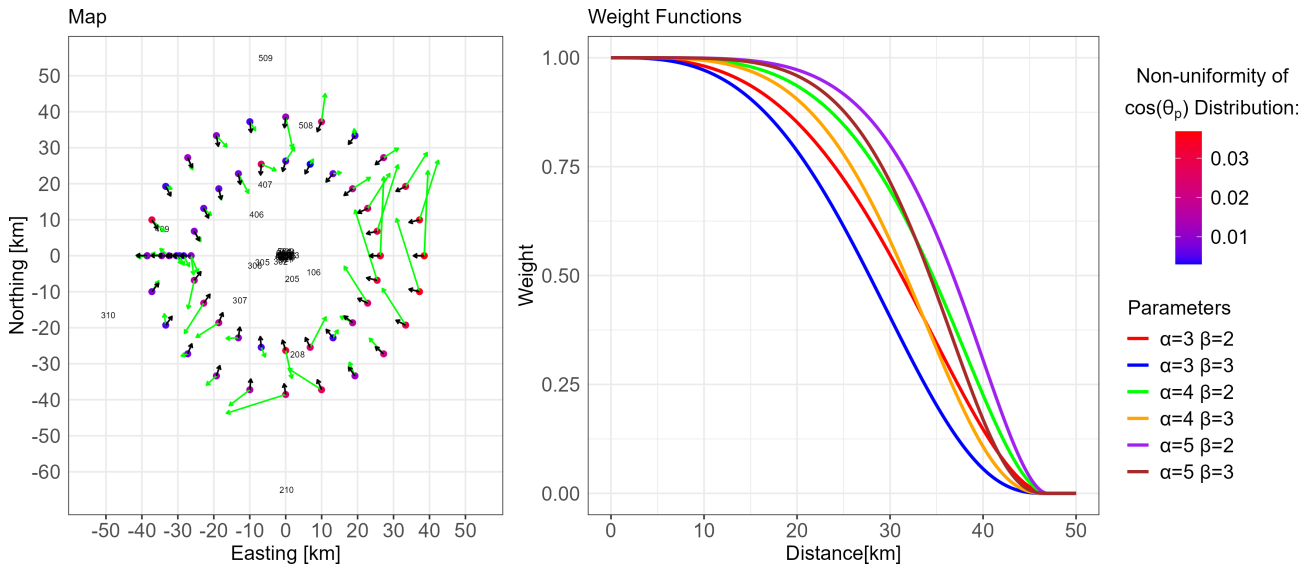
Figure 3.15: The left panel is identical to Figure 3.9 with the addition of black arrows, representing the average direction of stations at each location. The right panel shows the weighting function used for calculating the weighted average.
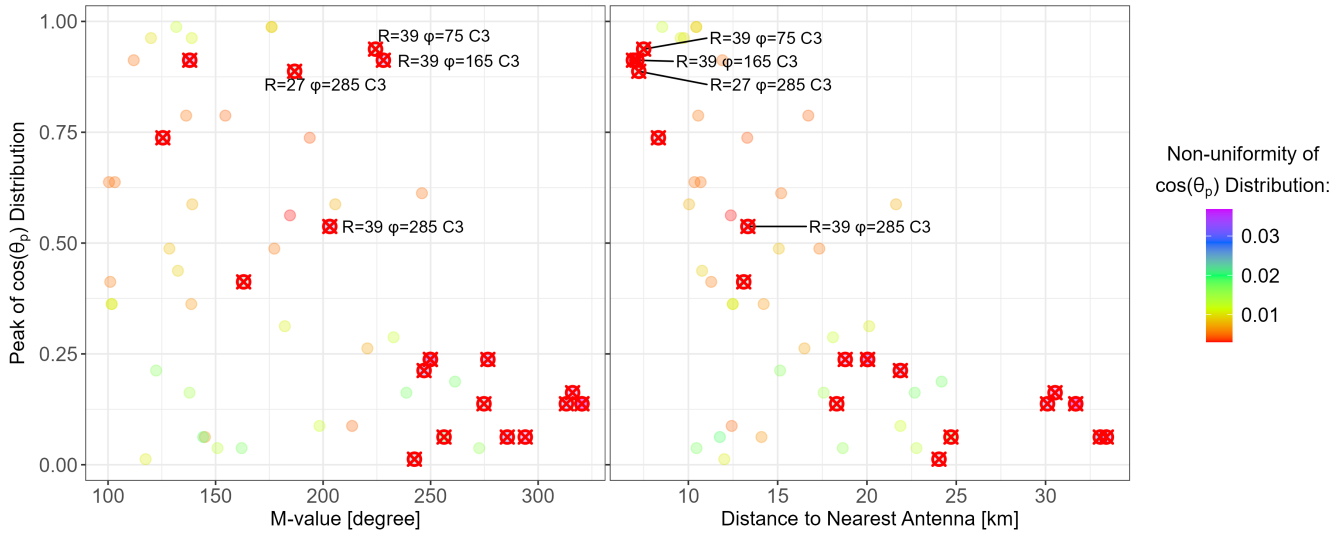

Figure 3.16: The left panel shows the relations between the peaks of the $\cos(\theta_p)$ distribution and the M-value, while the right panel shows the relations between the peaks of the $\cos(\theta_p)$ distribution and the distance to the nearest antenna. Red squares highlight points where the non-uniformity of the $\cos(\theta_p)$ distribution exceeds 0.02.

whether the observed peak is due to actual distribution patterns or simply a result of noise. Therefore, in Figure 3.16, we set a threshold for non-uniformity at 0.02, and regions where the non-uniformity of the $\cos(\theta_p)$ distribution exceeds this threshold are highlighted with red squares.

In the Distance to Nearest Antenna plot (right), there is a clear pattern: regions farther from the antennas tend to have $\cos(\theta_p)$ values closer to 0, meaning θ_p approaches 90°, while regions closer to the antennas tend to have $\cos(\theta_p)$ values closer to 1, meaning θ_p approaches 0°. This suggests a strong relationship between proximity to antennas and the value at which the peak of $\cos(\theta_p)$ occurs. In the M-value plot (left), the pattern is less pronounced, and several points deviate from the overall trend. By marking these

36

irregular points, we see that they are all from regions located very close to antennas. When excluding these points, which are likely influenced by their proximity to the antennas, a clearer relationship between the M-value and the peak of the $\cos(\theta_p)$ distribution becomes apparent.

In summary, in this section, we observed a clear relationship between the angular preferences in the $\cos(\theta_p)$ distribution and the position of the antennas, particularly the distance to the antennas. However, the underlying reasons for this relationship require further investigation.

3.3.5   Position Distribution and Antenna Position Relationship
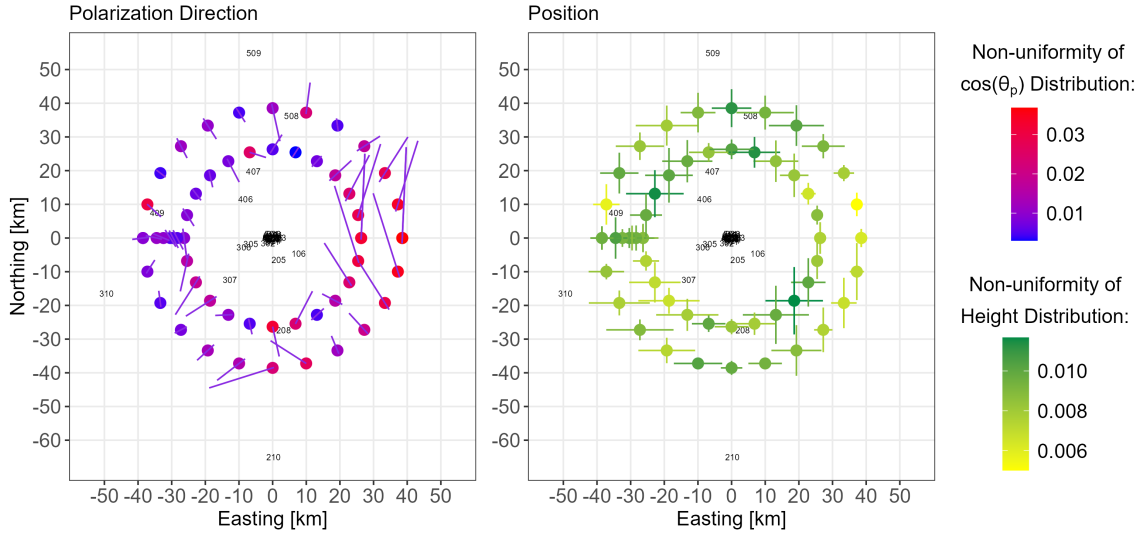

Figure 3.17: The left panel replicates Figure 3.9 for comparison. In the right panel, the numbers represent LOFAR station numbers, and the dots represent different regions. The color of each dot indicates the non-uniformity of the height distribution. Each dot extends with horizontal and vertical lines, where the horizontal line length represents the non-uniformity in the Easting distribution, and the vertical line length represents the non-uniformity in the Northing distribution.

This section will first analyze the non-uniformity of the background source position distribution. Similar to previous figures, Figure 3.17 is divided into two panels: the right panel shows the non-uniformity of the position distribution(as discussed in Sections 3.3.1), while the left panel depicts the non-uniformity of the polarization direction distribution. The left panel was previously discussed but is included here for comparison with the position distribution non-uniformity. In the right panel, black numbers indicate the locations of various LOFAR stations, and each point represents the region under investigation. From each point, lines extend to represent the non-uniformity across three dimensions. The horizontal line length represents the non-uniformity in the Easting distribution, the vertical line length represents the non-uniformity in the Northing distribution, and the color represents the non-uniformity in the Height distribution.

As shown in Figure 3.17, regions aligned with the direction of nearby antenna distributions tend to exhibit higher positional non-uniformity. This is especially evident in the eastern regions. For $\phi_a \in [-30°, 30°]$, the antennas are primarily distributed north-south on the western side relative to this region, and accordingly, the Northing distribution shows greater non-uniformity compared to the Easting distribution. In contrast, for regions with $\phi_a \in (180°, 270°]$, we observe higher Easting distribution non-uniformity compared to Northing. Although the antenna distribution relative to these regions is not strictly aligned east-west, antennas distributed north-south are located farther from these regions. Conversely, for regions with $\phi_a \in$

[60°, 180°), the Easting and Northing distribution non-uniformities are nearly identical, as the antenna distribution around these regions includes both east-west and north-south orientations, all at relatively close distances. Regions where antennas are distributed in all directions exhibit significantly lower non-uniformity in both Easting and Northing distributions than other regions. However, these regions display higher non-uniformity in height distribution. The above observations are general distribution patterns inferred from the figure and require further analysis to verify their accuracy.

When comparing the left and right panels, we observe some notable differences. Regions that previously exhibited relatively uniform polarization direction distributions, such as those where ϕ_a ranges from 60° to 210°, now show greater non-uniformity in the position distribution. Conversely, regions with highly non-uniform polarization direction distributions, such as those in the east, tend to exhibit more uniform position distributions. This inverse relationship between the uniformity of polarization direction and position distribution will be discussed in more details later.

To verify whether the patterns discussed above actually exist, we need to compare the non-uniformity across different distributions in greater detail. Although the data for different physical quantities were computed using the same algorithm to quantify non-uniformity, the final scales of the non-uniformity values differ due to differences in the distribution range and the bin width chosen for each physical quantity. Therefore, to facilitate better comparison, we will now explain the method used to normalize the data for different physical quantities to the same scale.

First, we normalize the non-uniformity values for all physical quantities by dividing them by their respective maximum values. This brings all the values to a common scale. Since we are interested in comparing the overall non-uniformity of the polarization direction with the overall non-uniformity of the position distribution, we compute the average of the normalized non-uniformity values for both $\cos(\theta_p)$ and ϕ_p for each region, representing the overall non-uniformity of the polarization direction. Similarly, we compute the average of the normalized non-uniformity values for Easting, Northing, and Height to represent the overall non-uniformity of the position distribution in each region.

After this step, we noticed that the variation in non-uniformity for the position distribution is still much smaller than that for the polarization direction distribution. The polarization direction data have non-uniformity values ranging from 0 to 1, while the position data have non-uniformity values ranging from 0.6 to 1. This is reasonable, as seen in Figure 3.7 and 3.10, where the lowest values of position distribution non-uniformity are not close to zero, whereas the non-uniformity values for the polarization direction distribution in many regions are close to zero.

To achieve a more comparable scale for analysis, we want both sets of non-uniformity values to vary between 0 and 1. We apply the following transformation:

$$data_{new} = \frac{data - data_{min}}{data_{max} - data_{min}}$$

This transformation scales the non-uniformity values so that both sets range from 0 to 1, allowing for a more direct comparison of the variation in non-uniformity between the position and polarization direction distributions.

Figure 3.18 presents the data obtained using the method described above, with the figure divided into two parts: the left panel shows the relationship between M-value and the non-uniformity, while the right panel depicts the relationship between Distance to Nearest Antenna and the non-uniformity. Focusing first on the upper part of each panel, the red points represent the position distribution non-uniformity, while the blue points represent the polarization direction non-uniformity. The purple dashed lines connect data points corresponding to the same region, illustrating the inverse trend between the two types of non-uniformity. As we hypothesized earlier, the non-uniformity of position and polarization direction tend to exhibit opposite behaviors.
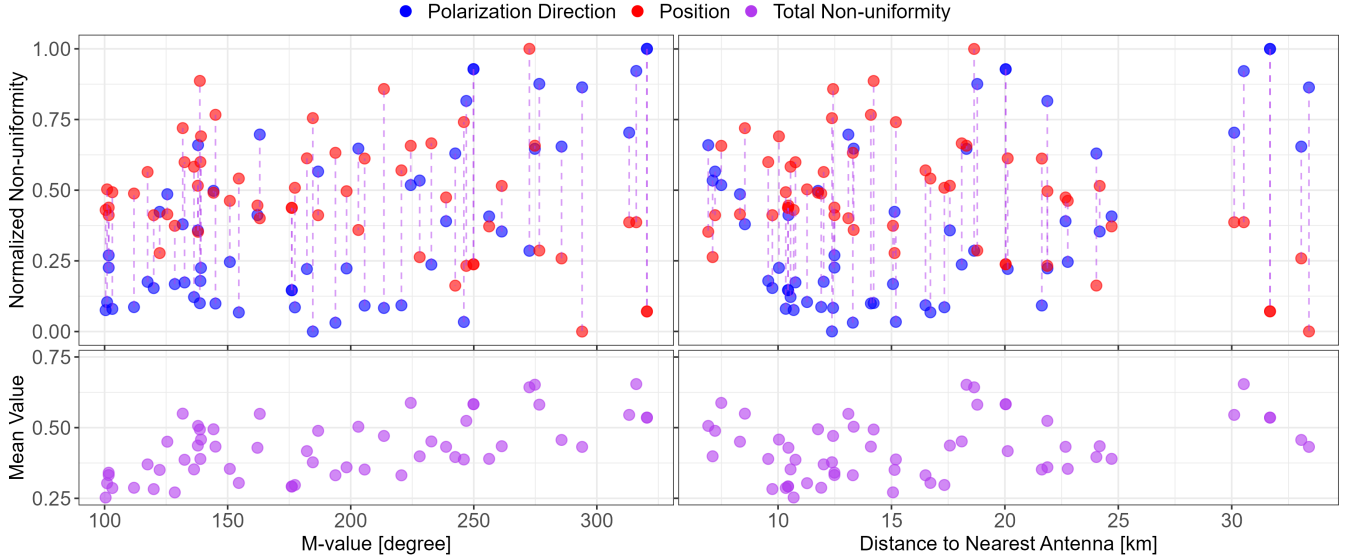
Figure 3.18: Relation between the non-uniformity of polarization direction distribution and position distribution. Blue dots represent polarization direction non-uniformity, and red dots represent position non-uniformity, with purple dashed lines connecting the two points in each data pair. The purple dot represents the mean non-uniformity of polarization direction and its corresponding non-uniformity of position for each pair. The left panel shows the distribution of these data under different M-values, while the right panel shows the distribution under different distances to the nearest antenna.

In the lower part of each panel, we plotted the average of the position and polarization direction non-uniformity for each region, indicated by purple points, which we refer to as Total Non-uniformity. Since both position and polarization direction non-uniformity are distributed within the range of 0 to 1, the Total Non-uniformity, as the average of these two values, should also lie within this range. However, as shown in Figure 3.18, the Total Non-uniformity values are concentrated within a relatively narrow range compared to the full scale of 1.

Looking closely at the Figure 3.18, we observe that Total Non-uniformity is constrained to a range of 0.25 to 0.63. Additionally, for regions with similar Distance to Nearest Antenna or M-value, the Total Non-uniformity varies only within a narrow margin of approximately 0.2. For example, for regions where the M-value is between 135° and 145°, the polarization direction non-uniformity spans from 0.05 to 0.7, while the position non-uniformity ranges from approximately 0.3 to 0.9. Despite this, the Total Non-uniformity for these regions is confined within a smaller range of 0.35 to 0.52. This observation suggests that, even though the individual non-uniformity for position and polarization direction may vary significantly, the Total Non-uniformity remains relatively stable, indicating a balancing effect between the two types of non-uniformity.

Regarding the trend of Total Non-uniformity with M-value or Distance to Nearest Antenna, while a positive correlation seems visible in the M-value plot, we do not believe this trend is meaningful for a few reasons. First, the overall variation is too small, which suggests that the trend could be due to measurement error. Second, we used the same weight when calculating the overall non-uniformity for polarization direction and position distribution, which could introduce bias. For ϕ_p and $\cos(\theta_p)$, their contributions to non-uniformity are likely different, and a similar argument can be made for the three dimensions of position (Easting, Northing, Height). Since each voxel is a rectangular box with a square horizontal cross-section but a height much greater than the horizontal dimensions, the contribution of Height to non-uniformity might differ significantly from Northing or Easting. In this case, the assumption that all components contribute equally to the overall non-uniformity may be flawed.

Given these considerations, I believe Figure 3.18 provides strong evidence for the conclusion that "the non-uniformity of position distribution and polarization direction distribution are inversely related." However, identifying other meaningful patterns will require further exploration into how the non-uniformity contributions of different physical quantities should be weighted. This refinement is necessary before drawing conclusions about trends involving Total Non-uniformity.

3.4   Summary and Discussion

In this study, we first identified non-uniformity in the polarization direction distribution of background sources. To investigate the cause, we initially experimented with different antenna functions. Although we observed small changes in the distribution when altering the antenna functions, there was no direct correlation between these changes and the uniformity of the polarization direction distribution. Notably, this limited impact of changing antenna functions was not only observed in the background source analysis but also in the exploration of discrete sources, where the changes in antenna functions had very limited influence on both the final position and polarization direction.

Exploring the polarization direction distribution in different regions, we discovered that the non-uniformity is location-dependent, and by examining adjacent regions, we observed consistent patterns. In this context, we explored how the polarization distribution changes with azimuth angle ϕ_a at different radial distances. Additionally, we investigated the effect of calibration changes on the polarization direction distribution. Ultimately, we found that both radial distance and calibration changes affect the polarization direction distribution in a location-dependent manner, with a clear relationship to the density of antennas. Regions with higher antenna density are more sensitive to small variations in position and calibration.

In later Section 3.3, we focused on the relationship between antenna positions and both polarization direction and position distribution, particularly examining non-uniformity. We found that Distance to Nearest Antenna and M-value significantly influence both the polarization direction and position distribution. Additionally, we discovered that the uniformity of the polarization direction distribution is inversely related to the uniformity of the position distribution. This phenomenon may suggest that the cause of the non-uniformity in the polarization direction and position distribution of background sources is not related to the antenna distribution. Although the polarization direction distribution of background sources tends to be more uniform under a more evenly distributed antenna array (low M-value), this also leads to a more uneven position distribution. This could be due to an issue within the algorithm used in the reconstruction process, where the program may need to make adjustments—deviating from the correct position or polarization direction—to fit the received signals.

4 Appendix

In this study, we explored the polarization direction distributions of the main axis for all regions. Shown are $\cos\theta_p$ (top scale, blue, binwidth = 0.039) and ϕ_p (bottom scale, red, binwidth = 7). The Easting (E), Northing (N), and height (h) values are indicated in the titles. Figures 4.1 to 4.3 display data using calibration file 3, Figure 4.4 uses data from calibration file 1, and Figure 4.5 shows data from calibration file 2.
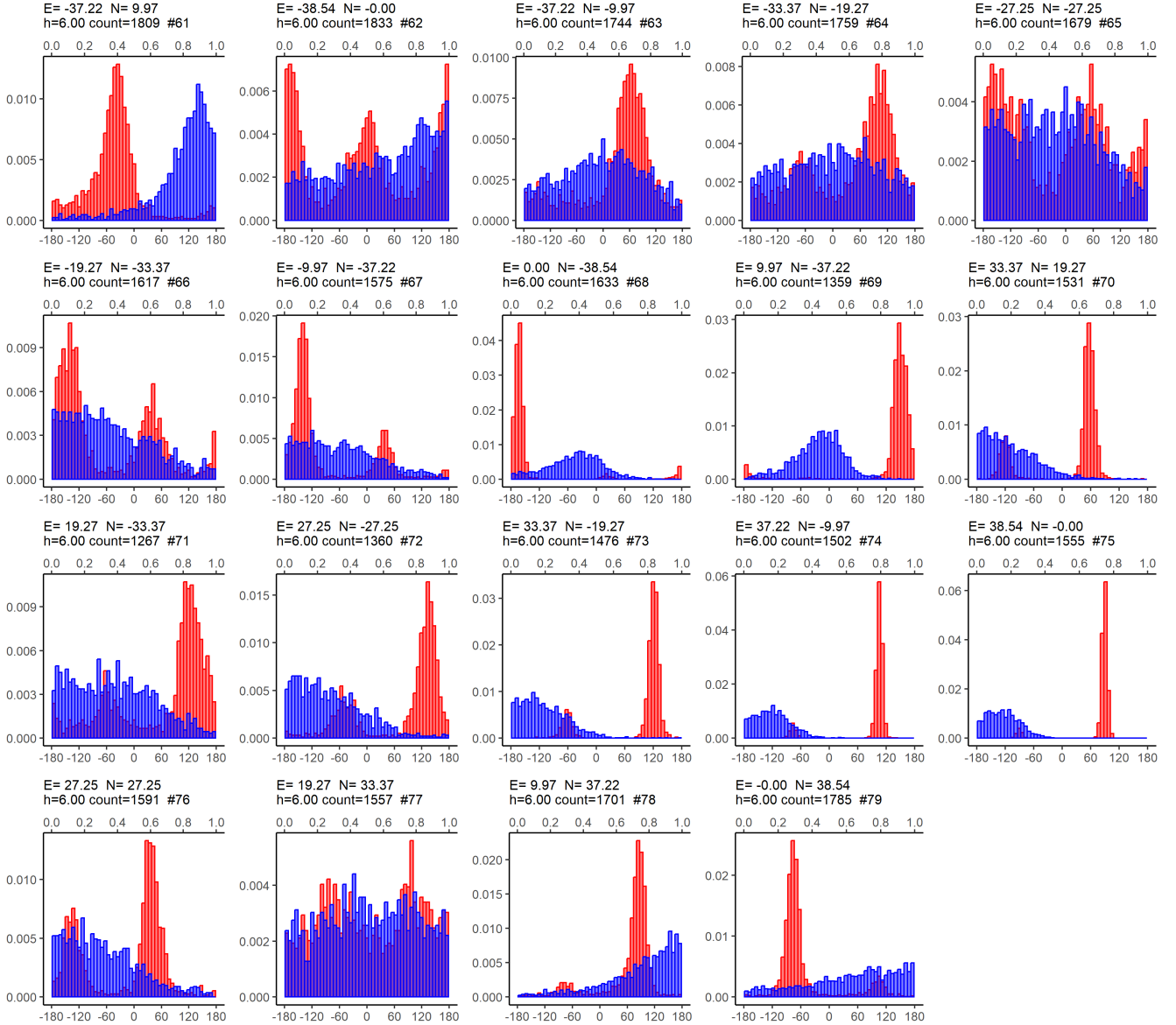

Figure 4.1: The Main Axis Distribution with calibration file 3
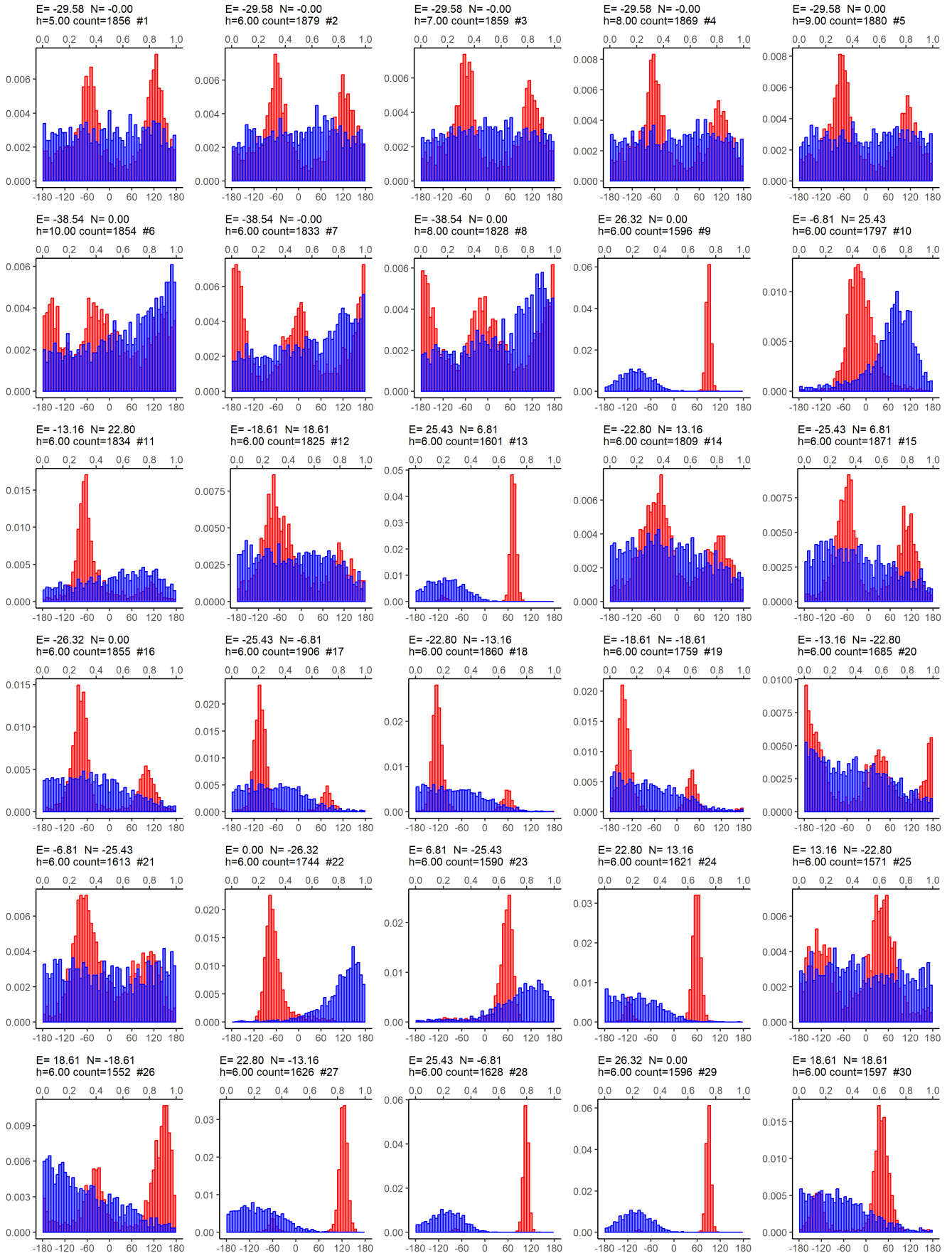
Figure 4.2: The Main Axis Distribution with calibration file 3
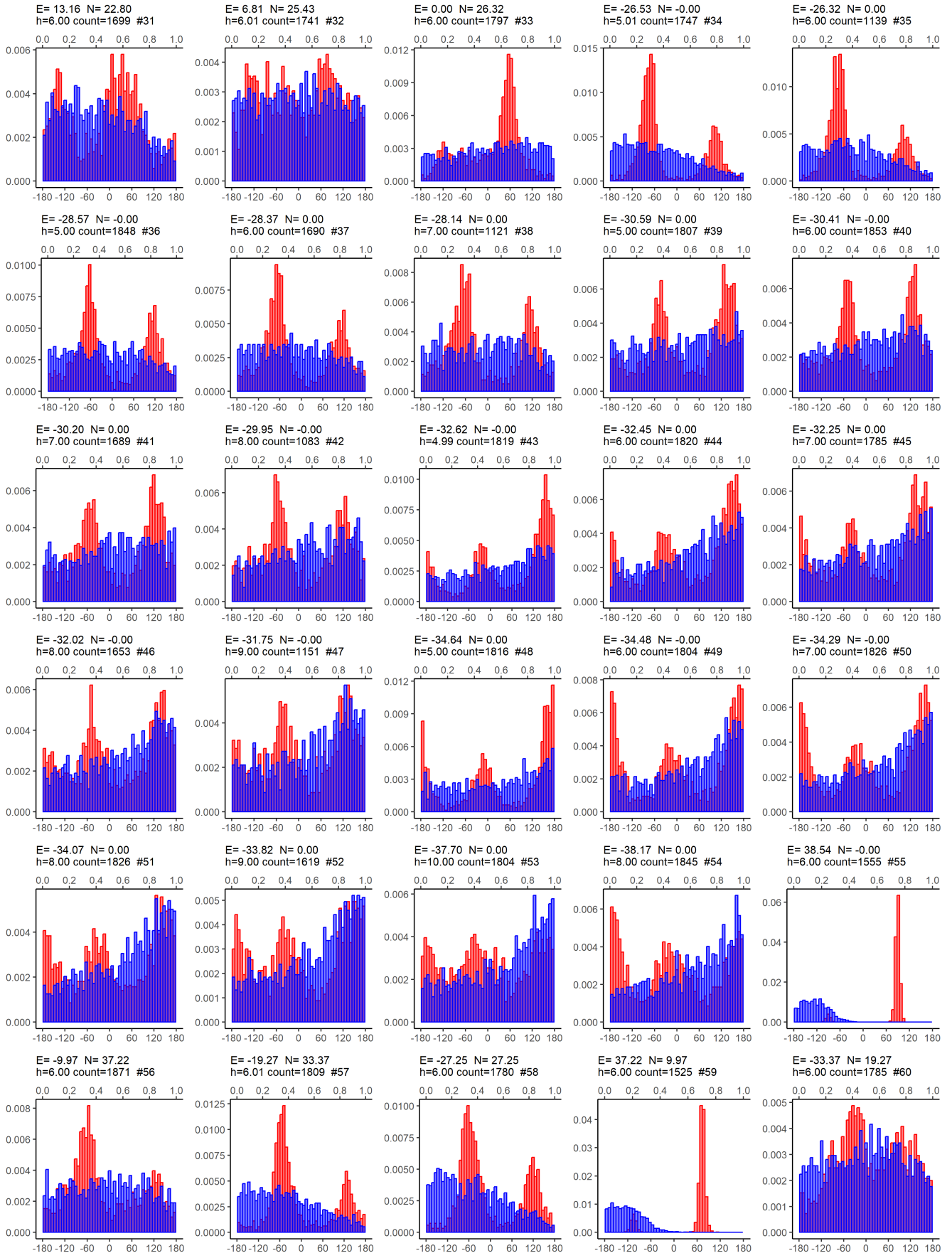
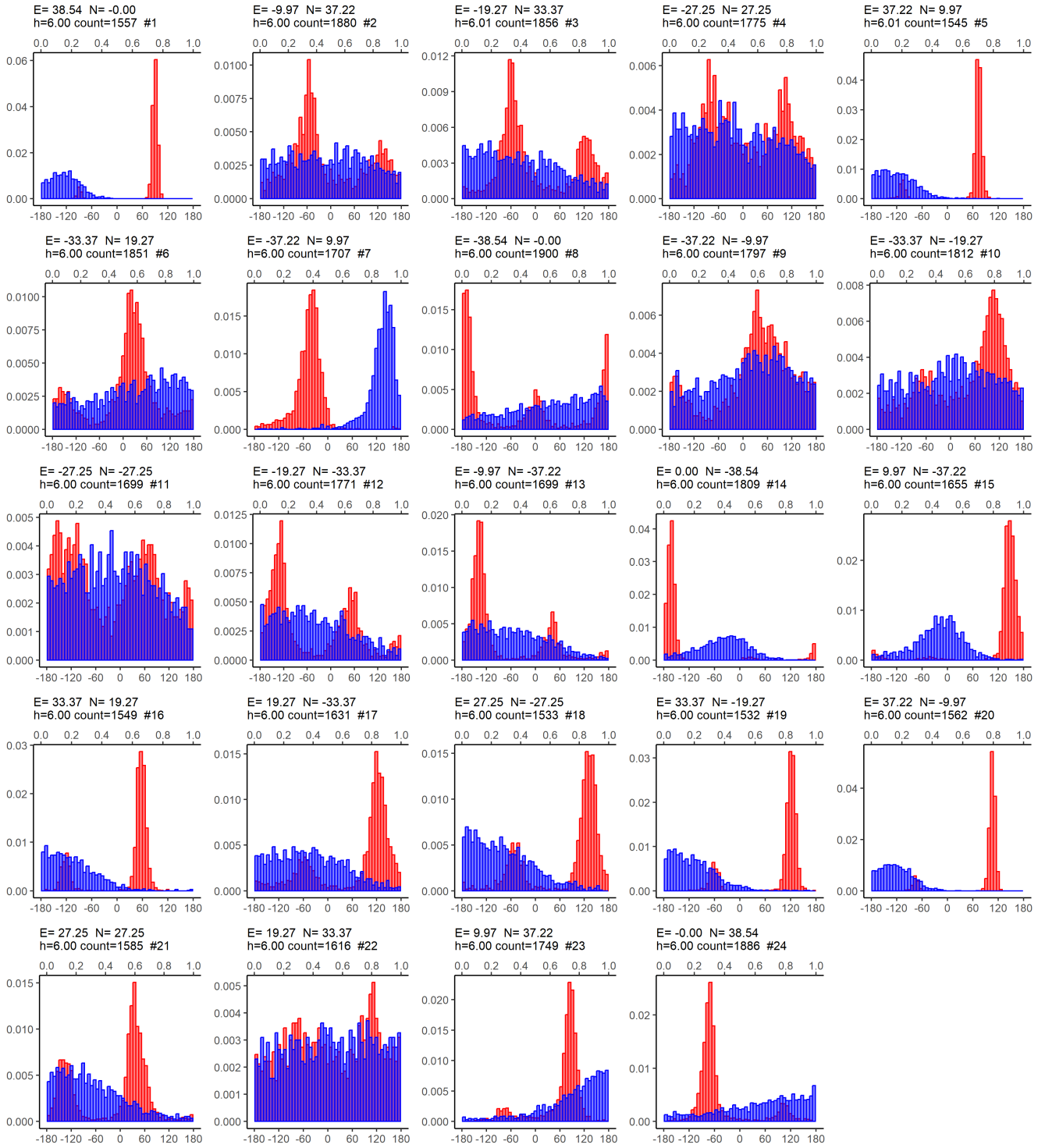Figure 4.3: The Main Axis Distribution with calibration file 3

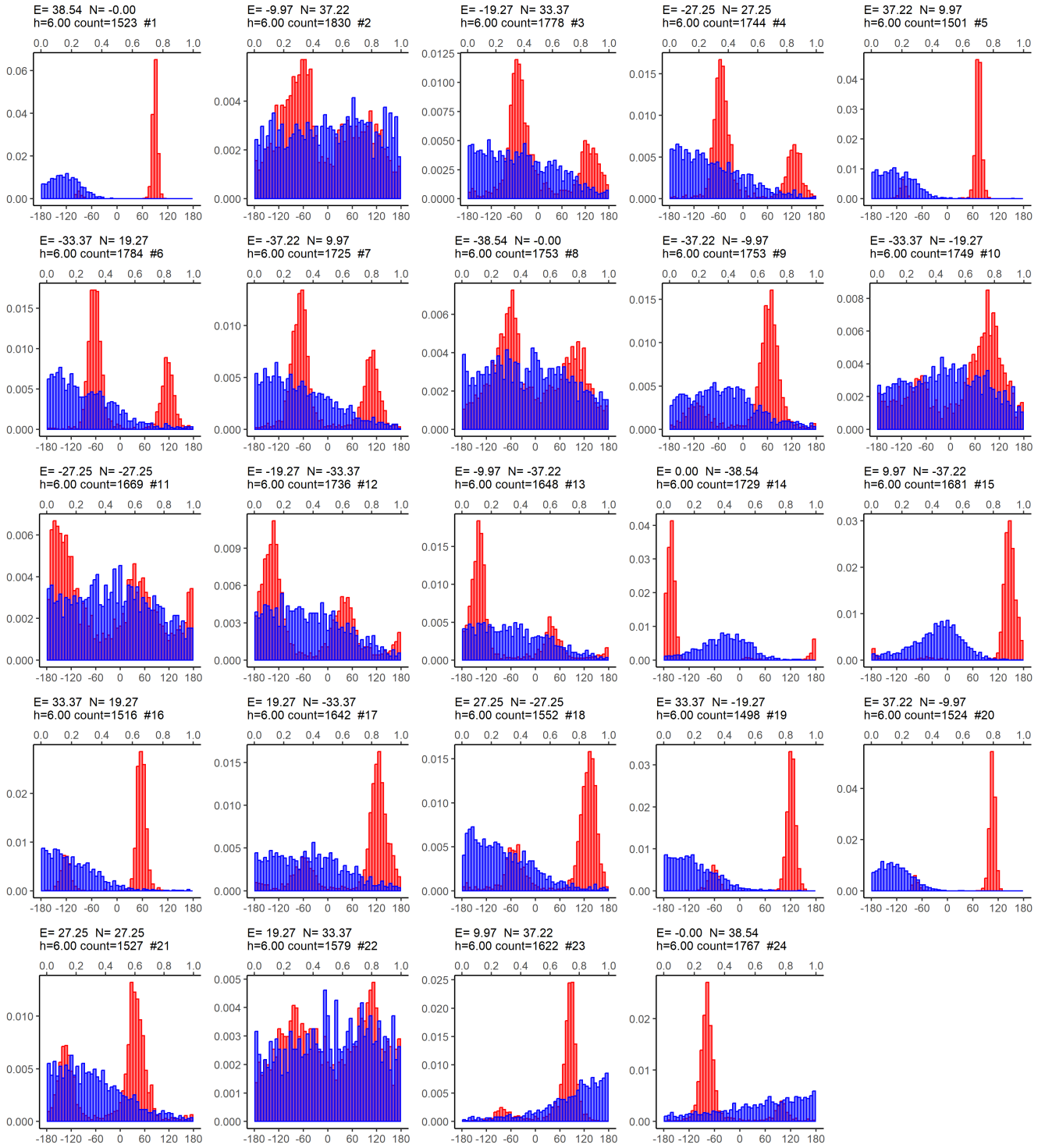Figure 4.4: The Main Axis Distribution with calibration file 1

Figure 4.5: The Main Axis Distribution with calibration file 2

References

[1] J. R. Dwyer and M. A. Uman. The physics of lightning. *Physics Reports*, 534(4):147 – 241, 2014. The Physics of Lightning.

[2] B. M. Hare, O. Scholten, J. Dwyer, T. N. G. Trinh, S. Buitink, S. ter Veen, A. Bonardi, A. Corstanje, H. Falcke, J. R. Hörandel, T. Huege, P. Mitra, K. Mulrey, A. Nelles, J. P. Rachen, L. Rossetto, P. Schellart, T. Winchen, J. Anderson, I. M. Avruch, M. J. Bentum, R. Blaauw, J. W. Broderick, W. N. Brouw, M. Brüggen, H. R. Butcher, B. Ciardi, R. A. Fallows, E. de Geus, S. Duscha, J. Eislöffel, M. A. Garrett, J. M. Grießmeier, A. W. Gunst, M. P. van Haarlem, J. W. T. Hessels, M. Hoeft, A. J. van der Horst, M. Iacobelli, L. V. E. Koopmans, A. Krankowski, P. Maat, M. J. Norden, H. Paas, M. Pandey-Pommier, V. N. Pandey, R. Pekal, R. Pizzo, W. Reich, H. Rothkaehl, H. J. A. Röttgering, A. Rowlinson, D. J. Schwarz, A. Shulevski, J. Sluman, O. Smirnov, M. Soida, M. Tagger, M. C. Toribio, A. van Ardenne, R. A. M. J. Wijers, R. J. van Weeren, O. Wucknitz, P. Zarka, and P. Zucca. Needle-like structures discovered on positively charged lightning branches. *Nature*, 568:360–363, 2019.

[3] Brian M. Hare, O. Scholten, J. Dwyer, U. Ebert, S. Nijdam, A. Bonardi, S. Buitink, A. Corstanje, H. Falcke, T. Huege, J. R. Hörandel, G. K. Krampah, P. Mitra, K. Mulrey, B. Neijzen, A. Nelles, H. Pandya, J. P. Rachen, L. Rossetto, T. N. G. Trinh, S. ter Veen, and T. Winchen. Radio emission reveals inner meter-scale structure of negative lightning leader steps. *Phys. Rev. Lett.*, 124:105101, Mar 2020.

[4] I. Kolmašová, O. Scholten, O. Santolík, B. M. Hare, P. Zacharov, R. Lán, N. Liu, and J. R. Dwyer. A strong pulsing nature of negative intracloud dart leaders accompanied by regular trains of microsecond-scale pulses. *Geophysical Research Letters*, 50(10):e2023GL103864, 2023. e2023GL103864 2023GL103864.

[5] O. Scholten, B.M. Hare, J. Dwyer, N. Liu, C. Sterpka, S. Buitink, A. Corstanje, H. Falcke, T. Huege, J.R. Hörandel, G.K. Krampah, P. Mitra, K. Mulrey, A. Nelles, H. Pandya, J.P. Rachen, T.N.G. Trinh, S. ter Veen, S. Thoudam, and T. Winchen. Distinguishing features of high altitude negative leaders as observed with LOFAR. *Atmospheric Research*, 260:105688, 2021.

[6] O. Scholten, B. M. Hare, J. Dwyer, N. Liu, C. Sterpka, S. Buitink, T. Huege, A. Nelles, and S. ter Veen. Time resolved 3D interferometric imaging of a section of a negative leader with LOFAR. *Phys. Rev. D*, 104:063022, Sep 2021.

[7] B. M. Hare, O. Scholten, S. Buitink, J. R. Dwyer, N. Liu, C. Sterpka, and S. ter Veen. Characteristics of recoil leaders as observed by lofar. *Phys. Rev. D*, 107:023025, Jan 2023.

[8] O. Scholten, B. M. Hare, J. Dwyer, N. Liu, C. Sterpka, K. Mulrey, and S. Ter Veen. Searching for intra-cloud positive leaders in vhf. *Scientific Reports*, 13(1):14485, 2023.

[9] B. M. Hare, O. Scholten, J. Dwyer, C. Strepka, S. Buitink, A. Corstanje, H. Falcke, J. R. Hörandel, T. Huege, G. K. Krampah, P. Mitra, K. Mulrey, A. Nelles, H. Pandya, J. P. Rachen, S. Thoudam, T. N. G. Trinh, S. ter Veen, and T. Winchen. Needle propagation and twinkling characteristics. *Journal of Geophysical Research: Atmospheres*, 126(6):e2020JD034252, 2021.

[10] Stan Heckman. *Why does a Lightning Flash have Multiple Strokes?*. PhD thesis, Massachusetts Institute of Technology, January 1992.

[11] Daniel P. Jensen, Richard G. Sonnenfeld, Mark A. Stanley, Harald E. Edens, Caitano L. da Silva, and Paul R. Krehbiel. Dart-leader and k-leader velocity from initiation site to termination time-resolved

with 3d interferometry. *Journal of Geophysical Research: Atmospheres*, 126(9):e2020JD034309, 2021. e2020JD034309 2020JD034309.

[12] Michiel A. Brentjens and Jan David Mol. *LOFAR Overview*, pages 19–36. Springer International Publishing, Cham, 2018.

[13] O. Scholten, B. M. Hare, J. Dwyer, C. Sterpka, I. Kolmasova, O. Santolik, R. Lan, L. Uhlir, S. Buitink, A. Corstanje, H. Falcke, T. Huege, J. R. Hörandel, G. K. Krampah, P. Mitra, K. Mulrey, A. Nelles, H. Pandya, A. Pel, J. P. Rachen, T. N. G. Trinh, S. ter Veen, S. Thoudam, and T. Winchen. The initial stage of cloud lightning imaged in high-resolution. *Journal of Geophysical Research: Atmospheres*, 126(4):e2020JD033126, 2021. e2020JD033126 2020JD033126.

[14] O. Scholten, B. M. Hare, J. Dwyer, N. Liu, C. Sterpka, I. Kolmašová, O. Santolík, R. Lán, L. Uhlíř, S. Buitink, T. Huege, A. Nelles, and S. ter Veen. Interferometric imaging of intensely radiating negative leaders. *Phys. Rev. D*, 105:062007, Mar 2022.

[15] Gourab Ghatak, Hardhik Mohanty, and Aniq Ur Rahman. Kolmogorov–smirnov test-based actively-adaptive thompson sampling for non-stationary bandits. *IEEE Transactions on Artificial Intelligence*, 3(1):11–19, 2022.

[16] T. W. Anderson. On the Distribution of the Two-Sample Cramer-von Mises Criterion. *The Annals of Mathematical Statistics*, 33(3):1148 – 1159, 1962.

[17] K. V. Mardia. Statistics of directional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(3):349–393, 1975.