RHODES UNIVERSITY

Where leaders learn

RHODES UNIVERSITY

MASTERS THESIS

SolarKAT: A Solar Imaging Pipeline for MeerKAT

Author:
Victória da Graça Gilberto Samboco

Supervisors:
Prof. Oleg Smirnov (RATT)
Prof. Ian Heywood (U. Oxford)

A thesis submitted in fulfilment of the requirements for the degree of Master of Science

in the

Centre for Radio Astronomy Techniques & Technologies
Department of Physics and Electronics

December 2023

Dedicated to myself, my parents Gilberto Samuel Samboco and Sandra José Samboco and my siblings Inácio Samboco, Stela Mindzo, Felizardo Samboco, and the twins Sandra and Olga Samboco.

RHODES UNIVERSITY

Abstract

Department of Physics and Electronics

Master of Science

SolarKAT: A Solar Imaging Pipeline for MeerKAT

by Victória da Graça Gilberto Samboco

Solar interference poses a significant challenge in radio interferometric observations, particularly with the increasing sensitivity of modern new-generation telescopes. This thesis presents the SolarKAT pipeline, a novel approach designed to mitigate solar interference in MeerKAT observations. The pipeline incorporates a series of steps, including self-calibration (second generation calibration or 2GC), precise determination of the Sun's position, phase centre adjustments, creation of region-based masks, deconvolution, prediction, solar model subtraction, and peeling.

We applied the SolarKAT pipeline to three datasets that feature the Sun in different conditions (frequency band and angular distance from the Sun to the telescope pointing position). These observations were obtained from three MeerKAT telescope surveys: ThunderKAT, MIGHTEE and LADUMA. We compared the visual images, peak fluxes, flux density, RMS and pixel distribution to evaluate the pipeline. Our results showed a notable reduction in solar interference. This is evidenced by the improved image quality, reduction in RMS and pixel distribution values, and consistent peak flux measurements after applying the pipeline. SolarKAT has not only improved the data quality but also demonstrated to be a valuable tool in producing high-quality solar images, which can be a helpful resource for solar physics and space weather forecasts.

This study showcases the potential of the SolarKAT pipeline in enabling high-quality radio interferometric observations, even in the presence of solar interference. Unlike conventional methods that often discard corrupted visibilities (e.g. flagging), our approach focuses on recovering them. Additionally, the SolarKAT pipeline naturally delivers detailed images of the Sun. Our findings contribute to advancing the field of radio interferometry, providing a valuable tool for researchers seeking to enhance the accuracy of their observations and conduct studies in solar physics and space weather.

Key words: Radio interferometry, solar interference mitigation, RFI, solar-imaging pipeline, Data methods, MeerKAT telescope.

Acknowledgements

I want to express my deepest gratitude to my supervisors, Oleg Smirnov and Ian Heywood, for all the guidance and support during this Master. In addition, I am thankful for the opportunity and believe I could make it to this stage. Thanks for being the best supervisors I could ever have in this process. I am also grateful to Dr. Landman Bester for contributing to this work.

I am also incredibly grateful to RATT members who were keen to help and answer my questions when needed, especially Robert, Lexy, Eric, Javas, Portia, and Ulrich. To Eric and Vuyo, thank you for making my first experience in Grahamstown smooth.

On a personal note, I would like to thank my family for their unconditional support and encouragement throughout my studies. Special thanks to my friends, especially Laurinda Macarringue, who gave me emotional and psychological support when needed; even in Mozambique, she was always present.

My sincere appreciation goes to my Landlords, Peter black and Jaco, for making me feel home; because of you, I have a second home in South Africa. Guys, thank you so much for being there for me, for all the support, and especially for taking care of me when I had medical issues; many, many thanks. Michael and Don, thank you. Also, thanks to Patience for being such a fantastic housemate in my last year of this degree.

Finally, I would like to acknowledge the NRF-SARAO for the financial support.

Declaration of Authorship

I, Victória da Graça Gilberto Samboco, declare that this thesis titled, '**SolarKAT: A Solar Imaging Pipeline for MeerKAT**', and the work presented is my own. I confirm that this work submitted for assessment is my own and is expressed in my own words. Any uses made within it of the works of other authors in any form (e.g., ideas, equations, figures, text, tables, programs) are appropriately acknowledged at any point in their use. A list of the references employed is included.

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is entirely my work.
- I have acknowledged all main sources of help.
- Where the thesis is based on my work jointly with others, I have made clear exactly what

Signed: _____

Date: _____

List of Conference Presentations

- **2023/11** - SARAO Postgraduate and Postdoctoral Conference, Cape Town, SA: Talk
- **2023/10**- Galaxy evolution group meeting, Astrophysics department, Oxford, UK: Talk
- **2023/09** - ThunderKAT Meeting, Astrophysics Department, Oxford, UK: Talk and poster exhibition
- **2023/09** - Rhodes Postgraduate Annual Conference, SA: Talk
- **2023/06** - BIUST summer school, Botswana: Flash talk
- **2023/04** - Young Scientists' Conference Conference, Ukraine: Talk
- **2022/11** - SARAO Postgraduate and Postdoctoral Conference, Johannesburg, SA: Poster exhibition

An abstract was submitted for the 29th Young Scientists' Conference on Astronomy and Space Physics (28 Apr 2023, 13:40) and made publicly available at https://indico.cern.ch/event/1279127/contributions/5373911/ and a video of the presentation here https://www.youtube.com/watch?v=uQLfSAEk6Dk&list=PL4vLHOXxSfCm1jw2YwftHJhD1JKpOwAMp&index=4.

Contents

List of Figures

List of Tables

Abbreviations

1GC	First Generation Calibration
2GC	Second Generation Calibration
3GC	Third Generation Calibration
BMAJ	Beam Major Axes
BMIN	Beam Minor Axis
BPA	Beam Position Angle
CASA	Common Astronomy Software Applications
CARTA	Cube Analysis and Rendering Tool for Astronomy
DD	Direction Dependent
DI	Direction Independent
DDEs	Direction Dependent Effects
DIEs	Direction Independent Effects
DR	Dynamic Range
DEC	Declination
RA	Right Ascension
DMS	Degree Minute Seconds
DME	Distance Measurement Equipment
E−CDFS	Extended Chandra Deep Field South
FoV	Field of View
FITS	Flexible Image Transport System
FT	Fourier Transform
GMS	Global System for Mobile Communications
GPS	Global Positioning System Satellites
HMS	Hours Minutes Seconds
LOFAR	Low-Frequency Array

LADUMA	Looking At the Distant Universe with the MeerKAT Array
MWA	Murchison Widefield Array
MIGHTEE	MeerKAT International GHz Tiered Extragalactic Exploration
MS	Measurement Set
PSF	Point Spread Function
PyBDSF	Python Blob Detector and Source Finder
RMS	Root Mean Square
RFI	Radio Frequency Interference
RIME	Radio Interference Measurement Equation
SDO	Solar Dynamic Observatory
SNR	Signal-to-Noise Ratio
VLA	Very Long Array
SKA	Square Kilometre Array
SKA-MID	Square Kilometre Mid
TOPCAT	Tool for Operations on Catalogues And Tables
ThunderKAT	The Hunt for Transients in MeerKAT

Chapter 1

Introduction

Radio astronomy is a powerful tool for studying and understanding the origin and evolution of the Universe. Radio telescopes observe many astronomical objects, including stars and galaxies. However, radio observations are subject to unwanted natural and artificial interference, known as Radio Frequency Interference (RFI). RFI is a growing issue for ground-based radio telescopes with the most sensitive receivers (Ford & Buch, 2014; Mesarcik et al., 2022).

One of the most significant challenges radio astronomers can face when conducting research with MeerKAT observations is RFI caused by the Sun. The Sun is a strong source of radio emissions, which can vary significantly by multiple orders of magnitude in periods of intense activity (Kansabanik et al., 2022). This massive source emits strong radio signals (Pawsey, 1946), and its radiation can be considered a nuisance by radio astronomers trying to observe other radio sources in the sky. This emission can interfere with radio interferometric observations of faint sources, reducing the sensitivity of the radio telescope and introducing artefacts that make the data calibration process and analysis challenging. Furthermore, the presence of the Sun in the sky during the observation can make it challenging to observe other astronomical sources, particularly faint sources (Ford & Buch, 2014).

The Sun is such a strong source of radio emission (Hey & Hey, 1973) that solar fringes can readily be seen on individual baselines of the MeerKAT telescope. The solar disk was successfully imaged with the telescope pointed almost 90 degrees away from the Sun (Samboco, 2023). In conventional terms, when we observe in optical, the telescope is pointed directly at the object we want to see. However, in radio astronomy, external sources can influence observations of

objects in the telescope's main field of view (FoV). This is due to how radio waves interact with the Earth's atmosphere and the telescopes.

As stated in the abstract from Samboco (2023), aside from observations where the Sun is very close to the telescope's optical axis, the observing conditions that lead to the most prominent solar interference still need to be investigated. This lack of understanding motivates efforts to develop techniques that will help us characterise and effectively mitigate solar interference, optimising the scientific potential of MeerKAT.

The MeerKAT radio telescope is a precursor radio interferometer for the Square Kilometre Array mid-frequency (SKA-MID) telescope, which is located in the Karoo region of the Northern Cape Province, in South Africa (Jonas & MeerKAT Team, 2016). Before the SKA-MID, MeerKAT is considered the world's most sensitive decimetre-wavelength radio interferometer array. The MeerKAT telescope comprises 64 antennas with a 13.5 m diameter and a maximum baseline of 8 km. MeerKAT operates in the UHF band (580-1015 MHz) and L-band (856-1712 MHz), with S-band (1.75 - 3.5 GHz) currently being commissioned. According to Jonas & MeerKAT Team (2016), solar interference is likely stronger in the UHF band. MeerKAT is a powerful telescope for studying celestial objects. However, its sensitivity and wide FoV confront a unique challenge—solar interference that can corrupt observations even when the Sun is not the primary target. Conventional solutions, such as using strong attenuators during daytime observations, prove to be impractical for MeerKAT, restricting its ability to capture weaker calibrators. In response, our research goal was to develop a robust, automated pipeline, SolarKAT, to mitigate solar interference, recover the visibilities and increase the scientific value of MeerKAT observations.

Mitigating solar interference has significant implications for the scientific product of radio telescopes, including MeerKAT and the SKA, the latter of which will be multiple times more sensitive than MeerKAT. Certain observations are done at night or during sunrise and sunset periods to avoid solar interference. This schedule can limit the discovery of sources that can only be seen during the day. Effectively subtracting the Sun from observations will allow MeerKAT users to observe at any time of the day without fearing the Sun's interference in the data. It will also provide scientists with more quality and reliable data for analysis. Additionally, a natural by-product of this project will be many images and movies that resolve the solar disk and its complexes of sunspots.

This takes us to the core question this thesis aims to address: How can solar interference be effectively mitigated in MeerKAT observations while preserving data integrity and enhancing the scientific potential of the telescope? To answer this question, the following sections explore the development of a novel solar imaging pipeline designed to address this challenge and unveil its potential impact on enhancing the quality of radio astronomy data and research.

The pipeline was developed primarily using the software development efforts of the Rhodes Centre for Radio Astronomy Techniques and Technology (RATT) group. The RATT group has a mission to conduct research into novel radio astronomy calibration, imaging, data analysis algorithms and technologies required to maximise the science of telescopes such as MeerKAT and the next-generation telescope, SKA. The pipeline is designed to be automated so that it can be used to process large datasets of observations. The pipeline works by first determining the location of the Sun concerning the antenna orientations and the celestial coordinates of the targets. Once the location of the Sun is known, the pipeline will then characterise the strength of the solar contamination. This information can further be used to influence the scheduling strategies of MeerKAT and SKA-MID.

Efforts have been undertaken to detect and mitigate RFI in radio astronomy. However, these have primarily targeted RFI caused by artificial sources such as the Global Positioning System (GPS) satellites, flight Distance Measurement Equipment (DME) and the Global System for Mobile Communications (GSM) (Sihlangu, 2019); this can be handled during the calibration process, by flagging contaminated antennas, channels, or baselines. These artificial RFI sources can easily distort or corrupt the weak signals from astronomical sources. Mitigating RFI caused by the Sun presents a unique challenge due to the variability and unpredictability of solar radio emissions. Due to its strong radio emission, the Sun can contaminate a wide range of frequencies in radio observations from lower to higher frequencies and flagging may not be sufficient to mitigate its interference effectively. Additionally, the Sun is an extended and resolved source, which poses an additional problem. In this case, additional techniques are required to address the issue of solar interference. This thesis presents an innovative approach that acknowledges the complexity of solar interference and proposes a fully automated and comprehensive method for its mitigation by subtracting the Sun's and its influence from MeerKAT data. This approach brings a very neat outcome by recovering the solar-contaminated data rather than discarding it.

1.1 Thesis outline

This work presents the results of the implementation of a pipeline we developed for solar interference mitigation in MeerKAT. This thesis is organised as follows. Chapter 2 presents an overview of the radio astronomy fundamentals, including the Radio Interferometer Measurement Equation (RIME) discussion of calibration, imaging and deconvolution. This Chapter also discusses solar radio interference, its sources, mitigation techniques, and observation strategies for the Sun and the motivation behind this work. In Chapter 3, we describe the SolarKAT pipeline, including the pipeline architecture and tools, the pipeline overview, a description of each step of the workflow and the evaluation metrics used to assess the effectiveness of the pipeline. Chapter 4 gives details of the step-by-step process of the data analysis carried out on the data. In Chapter 5, we present and discuss the pipeline implementation results and evaluate the pipeline's effectiveness in mitigating solar interference while preserving the characteristics of the sources of interest. Finally, Chapter 6 provides the conclusions and future work.

Chapter 2

Radio Astronomy Fundamentals

This Chapter aims to comprehensively understand the key concepts and techniques of radio astronomy and solar interference. According to Condon & Ransom (2016), radio astronomy is the area of astrophysics that studies radio emissions from natural sources in the sky. Compared to visible light, radio emission has longer wavelengths and tends to be weaker. To detect these signals, radio telescopes are used, then the signal is recorded and processed, and finally, an image is produced for analysis. This Chapter begins by discussing a single-dish telescope, followed by an introduction to radio interferometry, specifically focusing on a two-element interferometer. Subsequently, the Radio Interferometer Measurement Equation (RIME) and the crucial steps of calibration and imaging in interferometry are presented. Finally, it introduces radio interference focused on solar interference, its significance, and its impact on radio astronomy observations.

2.1 A single dish telescope

Radio telescopes are instruments that collect radiation from celestial sources. Generally, a radio telescope must fulfil two basic requirements: sensitivity and angular resolution (Ishak, 2019; Burke et al., 2019). The sensitivity depends on the size of the collecting area while minimising the contribution of unwanted signals simultaneously. The total dimensions of a telescope determine the angular resolution. This section considers a single aperture telescope with a diameter D, consisting of a single parabolic-shaped antenna that collects radiation from distant astronomical sources.

According to the Rayleigh criteria, the angular resolution of a single telescope is given by the Equation:

$$\theta \approx \frac{\lambda}{D} \tag{2.1}$$

Where λ is the observing wavelength, and D is the diameter of the dish. Smaller values of θ mean higher resolution, allowing radio astronomers to have more details of the observed source; this can be achieved by increasing the value of D. Therefore, huge dishes are prohibitively expensive (Ishak, 2019). Hence, a solution is to use an aperture synthesis array telescope, which combines various single-dish telescopes that work together as a single telescope. This is possible by using a technique called radio interferometry.

2.2 Radio interferometry

Radio interferometry is a technique used in radio astronomy to increase the resolution of a telescope by using multiple antennas simultaneously to observe the same source. This technique is known as *aperture synthesis*. In this process, signals from each antenna are combined to create an interference pattern containing information about the source. While for an aperture dish or a single dish, the angular resolution is given by the Equation 2.1, in an *aperture synthesis* system, the angular resolution is defined by the Equation:

$$\theta \approx \frac{\lambda}{b_{max}} \tag{2.2}$$

In Equation 2.2, b_{max} is the maximum distance between two single-dish radio telescopes in an interferometric array. Compared with the single-dish telescope, an interferometer resolution can be significantly higher since, in terms of baseline, the larger the baseline, the greater the angular resolution and more details can be resolved in an observed source.

After discussing the principles of radio interferometry, this section proceeds to demonstrate a specific implementation of a two-element interferometer. A two-element interferometer operates with two antennas. In Subsection 2.2.1, we discuss a simple implementation of a radio interferometer, describing how it collects signals and how we can obtain images from its measurements. In addition, we discuss how an interferometer can bring disadvantages to the observed signal.

2.2.1 A two-element interferometer

The deductions presented in this section are based on a summary made from Ishak (2019) and Thompson (2017).

Consideration will be given to a two-element interferometer, as illustrated in Figure 2.1. It consists of two antennas, A_1 and A_2, separated by the distance b, known as a baseline. Both antennas observe the same source, which is far field from the interferometer, meaning that the wavefront is effectively planar. A planar wavefront means that the wave is flat or two-dimensional. The radiation from the observed source is received in the form of electromagnetic waves in the direction **s**, with a specific frequency described as shown in Equation 2.3:

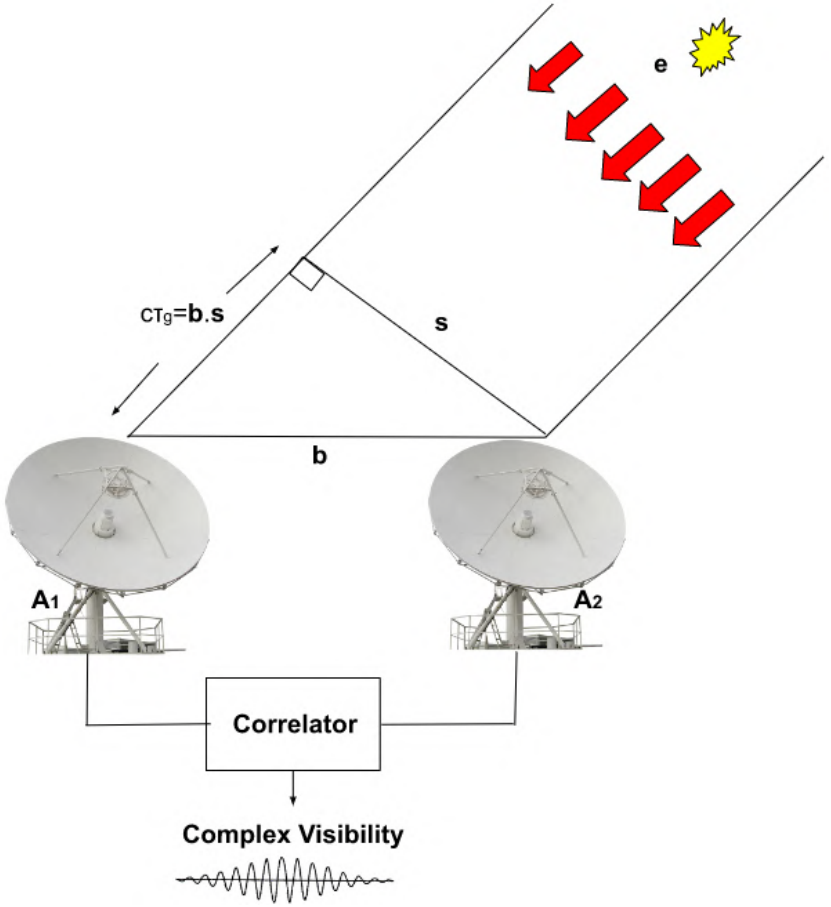


FIGURE 2.1: A Basic schematic diagram of a two-element interferometer. Image adapted from Thompson (1999, 2017) and Ishak (2019)

$$E(t) = Acos(2\pi\nu t + \Phi) \tag{2.3}$$

Where $E(t)$ represents the amplitude of the electromagnetic wave at a given time t, A is the amplitude of the wave, which is the maximum value of the wave's amplitude, ν is the frequency of the wave, t is the time at which the amplitude is measured, and Φ is the phase of the wave, which represents the initial angle of the wave at time t = 0.

Knowing that the angular frequency of a wavefront (ω) can be given by Equation 2.4:

$$\omega = \frac{2\pi}{T} = 2\pi\nu \tag{2.4}$$

Where T is the period of the wavefront and ν is the frequency of the wavefront.

Replacing 2.4 into 2.3 we have the final electromagnetic wave being described as:

$$E(t) = Acos(\omega t + \Phi) \tag{2.5}$$

But, because our two antennas are separated from a certain distance b, and the antenna A_1 has an extra path length compared with the antenna A_2, the radiation coming from $\mathbf{e}$ will reach the antenna A_2 before reaching antenna A_1, causing what we call geometric delay τ_g, represented in Equation 2.6, which is the time difference between the plane wave arrival at the antennas A_1, and A_2. We want to ensure we receive the same wavefront $E(t)$ in both antennas by determining the geometric delay τ_g given by:

$$\tau_g = \frac{b.s}{c} \tag{2.6}$$

Where c represents the speed of light in the medium, s represents the distance from the observed point to the observer along the direction of propagation. Now that we have obtained τ_g, we can represent the received signals (in $A1$ and $A2$) by the quasi-monochromatic Fourier components of a frequency ν as:

$$\begin{aligned} A_1(t) &= |A_1|cos\omega(t + \tau_g) \\ A_2(t) &= |A_2|cos(\omega t) \end{aligned} \tag{2.7}$$

These two waveforms of the signals (from 2.7) are then used as input for the correlator, which results in

$$\langle A_1(t) * A_2(t) \rangle = \frac{1}{T} \int_0^T A_1 * A_2 \, dt = |A_1 * A_2| cos(\omega \tau_g) \tag{2.8}$$

Where $\omega\tau_g$ represents the angular frequency times geometric delay, $|A_1 * A_2|$ represents the amplitude proportional to the intensity of the received power, and the angle brackets represent the average. The oscillatory term of the Equation 2.8 describes the periodic variation of the observed signal over dt as the Earth rotates (Thompson, 1999).

To produce an interferometric image, it is necessary to invert the Fourier Transform of the sky brightness. So far, we have only been considering a radio point source. Now, we express the interferometer output regarding radio brightness integrated over the sky. The sky brightness can be expressed as the intensity $I(s)$. It is measured in $Wm^{-2}Hz^{-1}sr^{-1}$, which represents the amount of flux per frequency ν at a collecting area $A(s)$ through a solid angle $d\Omega$ in the direction s. The integration of the intensity over the solid angle gives us the flux density, which is measured in Jansky ($1\ Jy = 10^{-26}Wm^{-2}Hz^{-1}$). Note that the signal power received in bandwidth $\Delta\nu$ from the source element $d\Omega$ is $A(s)I(s)\Delta\nu\ d\Omega$, where $A(s)$ is the effective collecting area in direction s, which is assumed to be the same for each antenna. The resultant correlator output is proportional to the received power and the cosine term. Setting s to 1, we can represent the correlator output from the solid angle of the signal as:

$$dr = A(s)I(s)\Delta\nu d\Omega cos(\omega\tau_g) \tag{2.9}$$

which from Equation 2.6 can be written in terms of baseline and source position vectors as:

$$r = \Delta\nu \int_s A(s)I(s)cos(\frac{2\pi\nu b.s}{c})d\Omega \tag{2.10}$$

The Equation 2.10 represents τ_g in terms of quantities that we can easily measure (baseline and source position). From the Equation 2.10, we assume that

- The bandwidth $\Delta\nu$ is sufficiently small that the dependence of A, I on ν can be ignored
- The source must be in the far field of the interferometer so that the incoming wavefront can be considered planar.

- The responses from different points in the source can be added independently.

When we observe a radio source to make an interferometric image, we specify the phase reference position or phase tracking centre. This position can be represented as $s = s_o + \sigma$, where σ is the distance between s_o and s. By substituting s in the Equation 2.10, we have the correlator output given by

$$r = \Delta\nu cos(\frac{2\pi\nu bs_0}{c}) \int_s A(\sigma)I(\sigma)cos(\frac{2\pi\nu b\sigma}{c})d\Omega - \Delta\nu sin(\frac{2\pi\nu bs_0}{c}) \int_s A(\sigma)I(\sigma)sin(\frac{2\pi\nu b\sigma}{c})d\Omega \tag{2.11}$$

At this point, we can introduce the visibility, defined as the measure of the spatial coherence of the electric field of a source emitting an EM wave. Visibility is a complex quantity given by the equation:

$$V = |V|e^{i\Phi\nu} = \int_s A(\sigma)I(\sigma)e^{(\frac{-2\pi\nu b\sigma}{c})}d\Omega \tag{2.12}$$

Where $A(\sigma) = \frac{A(\sigma)}{A_0}$ is the normalised antenna reception pattern, A_0 is the response of the beam centre, $\Phi\nu$ is the phase factor of the complex visibility, $I(\sigma)$ is the intensity function, ν is the frequency, b, c and σ are the baseline length, the speed of light and the source function.

We consider the case in which the antennas trace the source, and the system responds to the variation in the brightness distribution $A(\sigma)I(\sigma)$. Separating the real and imaginary parts of the visibility Equation (2.12), we have:

$$\begin{aligned} A_0|V|cos\Phi v &= \int_s A(\sigma)I(\sigma)cos(\frac{2\pi\nu b\sigma}{c})d\Omega \\ A_0|V|sin\Phi v &= \int_s A(\sigma)I(\sigma)sin(\frac{2\pi\nu b\sigma}{c})d\Omega \end{aligned} \tag{2.13}$$

Substituting 2.13 in 2.11 we have:

$$r = A_0\Delta\nu|V|cos(\frac{2\pi\nu bS_0}{c} - \Phi v) \tag{2.14}$$

The usual procedure of interferometry is to measure the phase and amplitude of the fringe pattern as represented by the cosine term of the Equation 2.14 and then correct the amplitude and phase of the visibilities by the calibration process. A source's brightness distribution is

obtained from the visibility data by inverting the transformation in the visibility Equation (2.12).

The practical application of the visibility equation requires introducing the coordinate system (u,v and ω). Where ω corresponds to the direction of interest (S_0) in relation to the celestial sphere, and u and v correspond to the sky plane. u, v and w are measured in wavelengths at the directions East and North, which are directions l, m and n measured in relation to u and v axis. Applying these new coordinates, we have:

$$\begin{aligned} \frac{\nu bs}{c} &= ul + vm + \omega n, \\ \frac{\nu bs_o}{c} &= \omega, \\ d\Omega = \frac{dldm}{n} &= \frac{dldm}{\sqrt{1-l^2-m^2}} \end{aligned} \tag{2.15}$$

Replacing the equations in 2.15 into the visibility Equation 2.12 we obtain the Equation 2.16

$$V(u,v,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(l,m)I(l,m)e^{-2\pi i[(ul+vm+\omega(\sqrt{1-l^2-m^2}-1)]}\frac{dldm}{\sqrt{1-l^2-m^2}} \tag{2.16}$$

This equation represents the complex visibilities as a function of u, v and ω. It relates the measured visibility $V(u,v,\omega)$ to the sky brightness distribution $I(l,m)$.

After obtaining the visibility equation, further processing steps, such as calibration, may be applied to ensure the quality of the imaging process. Calibration is a critical step which uses the known Radio Interferometer Measurement Equation (RIME) as a framework to model and correct for instrumental and propagation effects in the measured visibilities. In the next Section 2.3, we will discuss the RIME and how it works.

2.3 The Radio Interferometer Measurement Equation (RIME)

The Radio Interferometer Measurement Equation (RIME) is a mathematical framework used in radio interferometry to provide the basics for developing calibration methods and techniques. It was first formulated by Hamaker et al. (1996) to describe the Direction-Independent Effects (DIEs) in radio interferometric data. In addition to this, **?** presents an extended version of RIME, which incorporates the Direction-Dependent Effects (DDEs), describing complex effects that arise in radio interferometric data.

Section 2.2.1 shows that a single measurement from a two-element interferometer is a 2-D Fourier Transform of the sky brightness distribution. However, it is essential to note that this relationship only considers the geometric delay. In practice, additional factors must be considered to model the interferometric measurements accurately. These factors include instrumental effects, atmospheric conditions, antenna responses and other calibration-related parameters. The full RIME incorporates these factors to comprehend better the relationship between the measured visibilities and the sky brightness distribution, thereby reconstructing the Van Cittert-Zernike theorem. The Van Cittert-Zernike theorem is the result of the cross-correlation of the signals received in separated antennas to form an image of the intensity distribution of a source through a Fourier Transform (Thompson et al., 2017).

In Figure 2.1, we consider a single source *vece* where its electromagnetic (EM) field at a specific time and space can be described using a complex vector in an orthonormal x, y, z coordinate system (Andati, 2020). In this coordinate, system z represents the signal's propagation direction. Assuming the signal propagates as a plane wave, it does not have a component in the z-direction. This complex vector can be represented as:

$$\vec{e} = \begin{pmatrix} e_x \\ e_y \end{pmatrix} \tag{2.17}$$

Assuming that all corruptions encountered by the signal along its propagation path can be represented as linear operations concerning the signal itself, we represent it as:

$$\vec{e'} = \vec{J}\vec{e} \tag{2.18}$$

Due to the multiple effects the signal encounters along its path, each effect can be represented by a 2×2 Jones matrix. Consequently, all the corruptions affecting the signal can be represented by successive matrix multiplications of these matrices, following the order in which they occur. This sequence of matrix multiplications is commonly referred to as a Jones chain. Thus, the Equation 2.18 can be modified to incorporate the Jones chain as follows:

$$\vec{e'} = \vec{J_n}\vec{J_{n-1}} \ldots \vec{J_1}\vec{e} = \vec{J}\vec{e} \tag{2.19}$$

The order of terms in a Jones chain corresponds to the physical sequence in which the effects occur along the signal path. This is because matrix multiplication generally does not commute,

meaning that the order of multiplication matters. It is crucial to preserve this order when formulating equations involving Jones matrices.

Now, let us consider the scenario where the signal reaches the antenna and is converted into complex voltages by the antenna feeds. Suppose the existence of two feeds, denoted as x and y, which can be linear dipoles or left/right circular feeds. We assume that the voltages v_x and v_y are linear for the electric field $\mathbf{e}$. We can mathematically express the relationship between the two voltages as a matrix multiplication operation:

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix} = \vec{J}\vec{e} \tag{2.20}$$

Equation 2.20 can be interpreted as the fundamental linear relationship between the voltage vector v measured by the antenna feeds and the "original" signal vector $\mathbf{e}$ at some distant point. In this equation, the matrix J represents the total product of all propagation effects along the signal path, including electronic effects in the antenna/feed system.

Considering the two antennas p and q represented as A_1 and A_2 in image 2.1, v_p and v_q represent the signal received by each antenna. These voltage vectors are then fed into a correlator that produces four correlations pairwise between the voltage components of the two antennas. These correlations are given by:

$$\langle v_{px} v_{qx}^* \rangle, \langle v_{px} v_{py}^* \rangle, \langle v_{py} v_{qx}^* \rangle, \langle v_{py} v_{qy}^* \rangle \tag{2.21}$$

Where (*) is the complex conjugate operator. We can write the correlator products in Equation 2.21 in a matrix form to get:

$$V_{pq} = 2 \begin{pmatrix} \langle v_{px} v_{qx}^* \rangle & \langle v_{px} v_{qy}^* \rangle \\ \langle v_{py} v_{qx}^* \rangle & \langle v_{py} v_{qy}^* \rangle \end{pmatrix} \tag{2.22}$$

Which can be reduced to:

$$V_{pq} = 2 \left[\begin{pmatrix} v_{px} \\ v_{py}^* \end{pmatrix} \quad \begin{pmatrix} v_{qx}^*, v_{qy} \end{pmatrix} \right] = 2 \langle v_p v_q^H \rangle \tag{2.23}$$

Where H is the Hermitian transpose and factor 2 is introduced for conversion to obtain unit correlations for an intensity (I) of 1 Jansky (Jy). Substituting Equation 2.20 in Equation 2.23 we obtain:

$$V_{pq} = 2\langle \vec{J}_p \vec{e} \left(\vec{J}_q \vec{e} \right)^H \rangle = 2\langle \vec{J}_p \left(\vec{e} \vec{e}^H \right) \vec{J}_q^H \rangle \tag{2.24}$$

Assuming that the Jones matrices $\vec{J}_p$ and $\vec{J}_q$ are constant over the averaging interval, it is possible to move them outside the averaging operator. This can be expressed as

$$V_{pq} = 2\vec{J}_p \langle \vec{e} \vec{e}^H \rangle \vec{J}_q^H = 2\vec{J}_p \begin{pmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{pmatrix} \vec{J}_q^H \tag{2.25}$$

The terms in brackets are related to the Stokes parameters so that:

$$2 \begin{pmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{pmatrix} = \begin{pmatrix} I + Q & U + iV \\ U - iV & I - Q \end{pmatrix} = B \tag{2.26}$$

B represents the brightness matrix. The Stokes parameters I, Q, U, and V represent the total intensity linear and circular polarisation terms, respectively.

Moving the Jones matrices outside the averaging operator can separate the correlation matrix from the averaging of the Jones matrices. By separating the averaging of the correlation matrix and the Jones matrices, we can analyse and process them independently. This can be advantageous in calibration and data reduction procedures in interferometric measurements.

From B, we can formulate the RIME, which describes the relationship between the observed visibilities and the brightness distribution in the sky. For a single corrupted source, the RIME can be represented as:

$$V_{pq} = \vec{J}_p \vec{B} \vec{J}_q^H \tag{2.27}$$

Expanding $J_p J_q$ into the Jones chains, we have the RIME in the form

$$V_{pq} = \vec{J_{pn}} \left(... \left(\vec{J_{p2}} \left(\vec{J_{p1}} \vec{B} \vec{J_{q1}^H} \right) \vec{J_{q2}^H} \right) ... \right) \vec{J_{qm}^H} \tag{2.28}$$

Since the signal paths from the same source to different antennas p and q are not necessarily the same, the number of propagation effects (n and m) experienced by each signal path can differ. These differences cause phase delay, for which the derivation can be found in Smirnov (2011). The phase delay is described by a K-Jones matrix:

$$K_p = e^{-ikp} = e^{-2\pi i\left(u_{pl}+v_{pm}+w_p(n-1)\right)}, \tag{2.29}$$

For a single uncorrupted source, the RIME turns to:

$$V_{pq} = \vec{K}_p \vec{B} \vec{K}_q^H = \vec{X_{pq}}, \tag{2.30}$$

This visibility is measured by an ideal interferometer devoid of corruptions corresponding to the true source coherency X_{pq}. However, in practical scenarios, corruption along the signal path introduces additional effects, represented using the G-Jones matrix. The RIME, including the G-Jones matrix, is represented as:

$$V_{pq} = \vec{G_p} \vec{X_{pq}} \vec{G_q^H} \tag{2.31}$$

By applying the G-Jones matrix, the corruptions are accounted for, leading to a more accurate estimation of the source coherency and enhancing the reliability of the visibility measurements.

Extending the RIME formulation for discrete sources can be represented as:

$$V_{pq} = \sum_s \vec{J}_{sp} \vec{B}_s \vec{J}_{sq}^H \tag{2.32}$$

Where J_{sp} is the total Jones matrix in the direction **s** for antenna p. Considering the DIE effect (G_p) term and Direction-Dependent (DD) term $andE_{sp}$ we can write Equation 2.32 as:

$$V_{pq} = \vec{G_p} \left(\sum_s \vec{E}_{sp} \vec{X_{spq}} \vec{E}_{sq}^H \right) \vec{G_q^H} \tag{2.33}$$

These deductions are considered for discrete sources, whereas the sky has a continuous brightness distribution in real life. To get the total visibility measured by an interferometer, we must

consider brightness from all directions. According to Smirnov (2011), when we project the celestial sphere onto a tangential plane (l, m) centred on a particular field and consider the effects on direction (toward the field) in the uvw coordinate system as DD effects, we can derive the full sky RIME as:

$$V_{pq} = \vec{G_p} \left(\int \int_{lm} B_{pq} e^{-2\pi i(u_{pq}l + v_{pq}m)} dl dm \right) \vec{G_q^H} \tag{2.34}$$

The full sky RIME considers the DD effects that arise when observing extended or complex sources across the sky. It considers variations in the instrumental and propagation effects as a function of direction.

V_{pq} can be referred to as the observed visibility measured by the interferometer for a baseline pq. The observed visibility is a 2-D Fourier transformation of the apparent sky brightness seen by the baseline pq.

By performing the appropriate transformation and calculations, the RIME allows us to extract information about the brightness distribution of the sky from the observed visibilities, providing a framework for calibration (which we will discuss in the following section) and interpreting the measured data.

2.4 Calibration

In the previous sections, we discussed the RIME and how it relates the observed visibilities to the sky brightness distribution. The RIME provides a mathematical framework incorporating the instrumental and propagation effects to model the observed visibility. Now, we will turn our attention to the calibration process. The observed visibilities are affected by several factors that introduce errors in the observed data, including gains, instrumental errors, RFI, and DDEs. Correcting these issues in the observed data by solving each Jones term in the RIME in Equation 2.34 to reveal the intrinsic properties of the observed target is called calibration. By correcting these imperfections in the observed data, we improve the data quality, enabling high-quality images and facilitating scientific analysis of the radio sky. The RIME guides the development of novel calibration algorithms and methods that iteratively solve for the best-fit calibration solutions, minimising the difference between observed and true visibilities. In astronomy, the

calibration process begins in the instrumentation stage and ends in the final data visualisation (imaging).

In interferometry, calibration involves three techniques: Direction-Independent (DI) calibration, which includes first-generation calibration (1GC), second-generation calibration (2GC) or selfcal, and DD calibration or third-generation Calibration (3GC) according to Noordam & Smirnov (2010). In this work, we discuss two calibration techniques, 2GC and 3GC, as the data was obtained with 1GC carried out. Below, we describe each calibration technique, focusing on what is needed to understand the pipeline data processing in Chapter 3.

2.4.1 First Generation Calibration

First-Generation Calibration (1GC) is usually the first step implemented in radio astronomical data processing. This technique was developed during the early stages of interferometric observation before 1980 and is still crucial nowadays (Noordam & Smirnov, 2010). 1GC aims to estimate and correct instrumental and propagation errors using calibrator sources. These sources have well-known parameters such as flux density, shape and spectrum. In 1GC, primary and secondary calibrators are compared to the models themselves, and absolute flux, delay, and bandpass corrections are derived for the observed data and applied to the target data.

Absolute flux calibration

Consists of determining the actual flux of the sources in the field. This process requires a very bright and ideally invariable calibrator source.

Delay calibration

Delay errors in the signal path arise due to atmospheric delay, electronic delay, geometric delay due to the separation of antennas in a baseline, and errors in delay tracking in the correlator due to inaccurate models for the geometric delay. Delay calibration measures and corrects the remaining errors from the geometric delay (Fomalont & Perley, 1999). The delay can be represented by a K-Jones scalar diagonal matrix:

$$\vec{K} = \begin{pmatrix} e^{i\Phi} & 0 \\ 0 & e^{i\Phi} \end{pmatrix}, \tag{2.35}$$

where $e^{i\Phi}$ is the phase delay which is defined in Equation 2.29.

Bandpass calibration

Delay calibration is used to adjust for the propagation difference of the individual signal paths at the input of the correlator. Its primary purpose is to remove the significant phase gradient across the observed frequency band that would otherwise occur. To handle changes in antenna gain with frequency in radio interferometry, one approach is to consider the baseline-based complex gain as a function of frequency ($B_{pq}(\nu)$) (Fomalont & Perley, 1999). This frequency-dependent gain correction process is called bandpass calibration (Fomalont & Perley, 1999). The bandpass calibration requires a bright, invariable source with a known spectrum and is represented by the B-Jones matrix:

$$\vec{B} = \begin{pmatrix} b_x(\nu) & 0 \\ 0 & b_y(\nu) \end{pmatrix}, \tag{2.36}$$

Where $b(\nu)$ is the complex frequency variable gain, the indices x and y represent the antenna feeds.

In summary, to perform 1GC, we first i) observe the calibrator source to measure the gains (amplitude and phase) as a function of time; ii) observe a bright source with a known flux density and spectrum to measure the absolute flux calibration, bandpass and residual delays. This process involves alternating observations between the target and the calibrator source. 1GC or primary calibration involve flagging the data based on the RFI, spectral windows, channels, scans and antennas if necessary. After calibrating the calibrators, we apply their solutions to the target source and image it. Generally, images generated from the 1GC data have low Dynamic Ranges (DRs), which can be improved by performing 2GC techniques.

2.4.2 Second Generation Calibration (2GC)

After transferring the solutions from the calibrator to the target source, usually, the data continues with residual errors. This occurs because the calibrators are observed at times and positions in the sky that differ from the target. Self-calibration, most known as selfcal (Cornwell & Wilkinson, 1981) or 2GC allows solving the antenna gains using the model of the target data itself to refine the calibration solutions and reduce imaging artefacts. Selfcal assumes that the effects of the primary beam over time are constant across all the antennas in a homogeneous array (Smirnov, 2011).

The Self-calibration techniques include:

1. Making an image of the target after applying the calibrator solutions;
2. Using the model generated during the previous imaging process to calibrate the data over a specific solution interval (Selfcal);
3. Making an image of the target after the process of Selfcal;
4. During the imaging process, which includes deconvolution, the model data will be updated, and we will use this model over the same solution interval used in point 2. This processing cycle can be repeated for imaging and calibration until we are satisfied with the image quality.

There are different software for 2GC (CASA (Bean et al., 2022), CubiCal (Kenyon et al., 2018), etc), but we are interested in QuartiCal (Kenyon et al., 2023), a successor of CubiCal. QuartiCal is a Python-based calibration package that extensively uses Dask and Numba, which aims to be an all-in-one solution to calibration at scale.

1GC and 2GC techniques focus on solving for DIEs, while the 3GC technique is reserved for DDEs. In the following subsection, we will discuss the 3GC technique.

2.4.3 Third Generation Calibration: Peeling

The wide FoV of radio telescopes such as the Low-Frequency Array (LOFAR, Van Haarlem et al. (2013)), MeerKAT (Jonas & MeerKAT Team, 2016) and the Square Kilometre Array (SKA) give rise to the DDEs, which are showing to be a significant issue. They can no longer be ignored (Sirothia, 2017). These DDEs bring a new era of challenges to the calibration process. Third-generation calibration techniques focus on correcting DDEs in the observed data.

Traditional self-calibration methods assume that DDEs are insignificant and can be addressed in the image plane after calibration and deconvolution. However, as the sensitivity, field of view (FoV), and fractional bandwidth of modern interferometric arrays like LOFAR (Van Haarlem et al. (2013)), MWA (Lonsdale et al. (2009), MeerKAT (Jonas & MeerKAT Team, 2016) increases, the impact of significant DDEs becomes apparent. Unmodeled DDEs have the potential to corrupt the observations, leading to image artefacts that make detecting fainter sources difficult. Correcting DDEs becomes crucial, especially when detecting fainter sources (**?**).

In this context, mitigation strategies involve subtracting non-target sources influenced by DDEs from the data. Radio astronomy boasts various DD calibration techniques, including DDFacet (Tasse et al., 2023) and *peeling* (Noordam, 2004). The choice between these techniques depends on specific data characteristics. DDFacet, suitable when data exhibits numerous artefacts, uses KillMS (Tasse, 2023) to solve the RIME; however, it is at the cost of increased time consumption. In this investigation, our emphasis lies on the *peeling* technique.

In radio interferometry, *peeling* refers to a calibration process which aims to eliminate the contribution of prominent sources from observed data (Kenyon et al., 2018). This becomes particularly relevant when dealing with bright sources like the Sun, which can introduce unwanted signals in the observed field. Peeling effectively isolates and subtracts sources with robust radio emissions from the data, thus facilitating a more precise analysis of desired sources within the FoV. The main objective is to increase the overall observation's sensitivity. This process involves generating a model of the bright source and subsequently subtracting this model from the visibilities.

Traditionally, this procedure involves modelling sources in the primary field. However, in this study, we introduce an innovation: peeling a complex source, the Sun, situated beyond the FoV. This necessitates meticulous consideration of the source's position, morphology, and spectral characteristics. Succinctly, the peeling process encompasses the subsequent steps:

1. Source identification: Select the brightest source. The peeling begins by identifying and selecting a strong source of interest.
2. Model the source: Create a model by determining the self-calibration solutions of the selected source.
3. Source subtraction/peeling: Subtract/peel the selected source from the visibilities, multiplied by the relevant gains.
4. Return to step one and repeat the process using the updated model.

In summary, the goal of the calibration process is to remove as much as possible the effects of instrumental and atmospheric errors in the measured visibilities and other problems encountered by RFI sources, which will be discussed in the next Chapter. The correction of these effects must be applied to the visibilities before the imaging process to ensure high-quality images for analysis.

2.5 Imaging

Imaging refers to reconstructing images of the observed sources using the measured visibilities. In the case of interferometric data, this process is called *radio synthesis imaging*, which involves inverting the Equation 2.16 to get a synthesised image defined as:

$$\frac{A'(l,m)I'(l,m)}{\sqrt{1-l^2-m^2}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} V(u,v,\omega)e^{-2\pi i[(ul+vm+\omega(\sqrt{1-l^2-m^2}-1)]}dudvd\omega \tag{2.37}$$

$A'(l,m)$ represents the FT of the aperture illumination function (how the antenna or the array of antennas is illuminated). $I'(l,m)$ is a function that describes the sky brightness distribution, which we are trying to measure. We can simplify the Equation 2.16 by reducing it to a two-dimensional Fourier Transform following two conditions:

- Considering the $\omega - axis$ to be in a celestial pole and equal to zero;
- Considering $|l|$ and $|m|$ small enough that we image a small FoV in the sky and $(\sqrt{1-l^2-m^2}-1)\omega \approx -\frac{1}{2}(l^2+m^2)\omega \approx 0$

It also assumes that $A(l,m) = 1$ to simplify each condition. Taking into consideration these conditions, we obtain the Van Cittert-Zernike theorem:

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(l,m)e^{-2\pi i(ul+vm)}dldm \tag{2.38}$$

Which is the Fourier Transformation of the sky brightness (l,m). Having the Equation 2.38, we can perform an inverse FT of the visibility equation resulting in the following:

$$I(l,m) = F^{-1}\{V(u,v)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} V(u,v)e^{2\pi i(ul+vm)}dudv \tag{2.39}$$

where $F^{-1}\{V(u,v)\}$ symbolises an inverse Fourier Transform of the visibilities. The inverse Fourier Transform of the visibility Equation (2.38) given by Equation 2.40 reconstructs the image by integrating the complex visibilities over all spatial frequencies and calculating the complex amplitudes at each position in the image plane. These amplitudes represent the brightness or intensity of the sources in the sky.

The visibility is a 2-D function representing the observed source's brightness for a two-element antenna separated by a certain baseline b. In the case of infinite antennas and extended baselines, we use the sampling function in terms of the 2-D Dirac delta function or $\delta-$distribution (Briggs et al., 1999) in Equation 2.40 to interpolate the visibility function for all antennas in the interferometer array. This allows the production of a single image from the data collected by the antennas in the radio telescope array.

$$S(u,v) = \sum_{k=1}^{M} \delta(u-u_k, v-v_k) \tag{2.40}$$

As it derives from the visibility function, the sampling visibility function is also a 2-D function that describes how the radio interferometer samples the visibility function. The sampling process involves multiplying the Fourier Transform of the visibility function by the sampling function. Applying the sampling function (in Equation 2.40) to the visibility function, we obtain:

$$V^S(u,v) = \sum_{k=1}^{M} S(u,v)V'(u_k,v_k) \tag{2.41}$$

Where $V^S = SV'$, V^S is the sampled visibility and $V'(u_k, v_k)$ or V' is the observed visibility. The prime in the observed visibility means that the visibility is corrupted by "noise". Applying the Fourier Transform of the sampled visibility (in Equation 2.40), we obtain:

$$I^D(l,m) = F\{V^S\} = F\{SV'\} = F\{S(u,v)V'(u_k,v_k\} \tag{2.42}$$

According to the convolution theorem from Almeida (1997), the FT of a product of functions is the convolution of their FTs. On this note, Equation 2.42 can be expressed as:

$$I^D(l,m) = F\{S(u,v)\} * F\{V'(u_k,v_k)\} \tag{2.43}$$

Where * denotes convolution, I^D is the reconstructed sky image, $F\{V'(u_k, v_k)\}$ results in the true sky image, which we can denote as $I_{true}(l,m)$, $F\{S(u,v)\}$ results in the Point Spread Function (PSF), which represents the interferometer response to a $1Jy$ point source. According to these assumptions the function represented in Equation 2.43 can be written as:

$$I^D(l,m) = PSF * I_{true}(l,m) \tag{2.44}$$

The convolution of the PSF and the true sky image results in the reconstructed sky image I^D called the *dirty image*. To obtain a clean image of the observed visibilities, we apply the cleaning process (to the dirty image) called deconvolution, which will be discussed in the following section as a solution for the incomplete sampling of the visibility function.

2.6 Deconvolution

Deconvolution is a powerful tool in radio astronomy data for improving the image quality obtained from interferometric data from instruments such as the MeerKAT telescope. The reconstructed image accuracy can be limited by two primary issues in the observed visibilities: the limited spatial frequency distribution (u, v) and errors in the measured visibilities. These issues can be improved through the deconvolution process. A standard deconvolution technique in radio interferometry is the CLEAN algorithm designed by Högbom (1974). This is an essential tool for improving the image quality of radio astronomy aperture synthesis observations (Clark, 1980). Deconvolution separates the intensity distribution of the visibility data into point sources. Then, it replaces each one with the corresponding response to a beam free of sidelobe artefacts.

In a noise-less system, the deconvolution functions are derived from the equation 2.44 and can be represented as:

$$F\{I_{true}(l,m)\} = \frac{F\{I^D(l,m)\}}{F\{PSF\}}; \tag{2.45}$$

In radio interferometry, when we measure the sky visibility, we are convolving the sky with the array response (PSF) and some instrumental noise n, turning the Equation 2.44 to:

$$I^D(l,m) = PSF * I_{true}(l,m) + n \tag{2.46}$$

Equation 2.42 considers $n = 0$, turning it into a simple deconvolution process called *inverse filtering*. Due to several instrumental effects, n is not equal to zero in interferometry. If an inverse filter is applied to a noise signal, it results in a poor estimation of the original signal. To address this issue, we need two aspects: the existence of the PSF response of the array and the knowledge that every source in the sky is convolved (blurred) with the PSF. Considering this aspect, it is first necessary to identify sources in the sky with high signal-to-noise (SNR).

Then, we subtract the PSF multiplied by a factor from the corresponding position in the image. This factor considers the different brightness levels of the sources, and then the position and the adjusted brightness are added to the sky model.

Since each pixel in the image contains a combination of noise and potential flux from another source, not all flux from a single pixel is removed at once. Instead, an iterative deconvolution approach is used. This approach uses the CLEAN algorithm, which involves gradually removing a portion of the flux from a pixel until a certain noise threshold is reached, improving the clarity and quality of the image by iteratively (by setting a certain number of iterations) removing the PSF effects.

2.6.1 The CLEAN algorithm

CLEAN is an iterative algorithm for non-linearly deconvolving instrumental PSF from the observed data (Starck et al., 2002). This subsection is mainly based on (Högbom, 1974). The CLEAN algorithm consists of the following steps:

1. Compute the dirty image (I^D) and the response to a point source by a Fourier transformation of both the visibility and the weighted transfer function.

2. Find the maximum intensity point in the dirty image (I^D) and the position of the maximum pixel in the dirty beam and subtract it, including the full sidelobe pattern at the beam centre. The peak amplitude of the subtracted PSF is equal to the loop gain, which by analogy has negative feedback in electrical systems.

3. Record the position and the amplitude of the point source subtracted into a model that will become the cleaned image.

4. Repeat step 2, replacing the dirty image with the dirty image produced in the previous iteration. The stopping criteria can include the major loop gain, the number of iterations, the cleaning threshold, or finding a negative component (Offringa & Smirnov, 2017). The final result of this process will be the so-called *restored image*. Note that the restored image is the final residual, plus the model convolved with the *restoring beam*. The restoring beam can be determined using the point source observation by observing a point source-like object where the PSF's size and shape will reveal the restoring beam's characteristics.

5. Determine the restored beam by using a calibrator source. By observing the calibration source, we can measure the telescope response to different frequencies, and this information can be used to determine the restoring beam. The restoring beam is crucial for the deconvolution process. Other products of deconvolution are residual images and sky model images.

The clean algorithm was initially developed assuming that the sky represents a random distribution of point sources (Cornwell, 2008). However, traditional CLEAN methods encounter limitations when dealing with complex sources with diffuse emissions and a wide range of spatial scales like the Sun. To address these challenges, Cornwell (2008) presented the MULTISCALE CLEAN algorithm as an extension of the original CLEAN algorithm. This algorithm is helpful for extended sources by incorporating a multiscale approach to the cleaning process. The following subsection summarises the multiscale clean based on Offringa & Smirnov (2017); Cornwell (2008); Thompson (2017).

2.6.2 MULTISCALE CLEAN deconvolution

MULTISCALE CLEAN (Cornwell, 2008) is based on the traditional CLEAN algorithm but extends its capabilities by considering multiple scales in the deconvolution process. While the standard CLEAN algorithm considers a single-scale PSF to model the blurring effects, the MULTISCALE CLEAN algorithm recognises that different structures within the image may have distinct scales that require different treatment. The MULTISCALE CLEAN algorithm is widely used to improve the quality and resolution of images by addressing the presence of structures at different scales.

The main idea behind the MULTISCALE CLEAN algorithm is to decompose the image into multiple scales or levels, representing different frequency components or structures. This decomposition enables a more accurate characterisation of the object's complexity. A suitable PSF is utilised at each scale to iteratively remove the blurring effects and enhance the fine details in the deconvolved image. This process preserves the features and structures present in the observed image, including both small-scale details and larger-scale structures. This makes it particularly valuable for studying objects like the Sun, where various spatial scales are involved.

Building upon what was discussed on interferometry, the RIME, calibration, imaging, and deconvolution techniques, we now shift our focus to another critical aspect that can impact the

accuracy and reliability of radio observations: RFI. In addition to the challenges posed by DIEs and DDEs, the presence of RFI, including solar interference, introduces further complexities to the calibration and post-image analysis process.

2.7 Solar Radio Interference

In this section, we explore the concept of RFI, with a specific emphasis on solar radio interference. RFI can be defined as any unwanted signal that interferes with the reception of a desired observed signal, reducing the sensitivity and data quality of a radio telescope by producing artefacts in the observed data (Ford & Buch, 2014). Natural phenomena or human-made devices can cause these interference signals. With the effort of developing radio telescopes with increased sensitivity, such as LOFAR, VLA and the new generation MeerKAT and SKA-MID, the ability to detect and capture unwanted signals is also increasing. A study by Sihlangu (2019) identifies GPS, GSM and DMEs as the primary sources of RFI in the data from telescopes such as MeerKAT.

While previous studies have focused on identifying, characterising and mitigating radio interference caused by human-made sources in MeerKAT observations, we need more techniques and technologies for solar interference mitigation. Our attention now turns to solar interference, a specific type of radio interference resulting from the Sun's activity and emission. Solar interference can present unique challenges that significantly impact radio astronomy observations.

This section delves into the principal aspects of solar radio interference, including sources and existing mitigation techniques in radio astronomy data. We also highlight the need for specialised mitigation techniques to address its distinct characteristics. By understanding the context of RFI and its manifestation as solar radio interference, we can better understand the motivations and urgency behind developing effective mitigation strategies such as the one we propose in Chapter 3.

2.7.1 Sources of Solar Radio Interference

The Sun is a powerful radio emission source with significant angular size. Its properties make it a significant source of RFI in astronomy ground-based telescopes. The RFI caused by radio solar emission can be called solar RFI.

The Sun, a source with the highest flux density in the sky (Kansabanik et al. (2022, 2023)), due to its proximity to Earth, emits a wide range of electromagnetic waves, including radio waves, due to various phenomena that occur on its surface. These phenomena include solar flares, coronal mass ejections, and solar radio bursts. These events emit intense radio frequency signals that can overlap with astronomical signals of interest. The amount of energy released during these events, covering different frequencies, makes the Sun the source with the highest apparent brightness in the sky.

The intensity of the Sun's radio emission and its activity are not the only properties that make it a strong source of interference; its significant angular size also makes it a strong source of interference. The physical dimensions of the Sun result in extended radio emissions over the sky. Consequently, even when radio telescopes are not directly pointing at the Sun, its emission can still impact the observations by introducing artefacts and fringes, and the solar disk can still be captured in the primary beam. This widespread influence requires careful consideration and mitigation strategies to minimise the interference caused by solar radio emissions.

Solar radio emissions can contaminate radio telescope data, introducing high noise levels leading to decreased sensitivity and reduced image quality. Understanding the nature of solar radio interference is crucial for ensuring accurate and reliable observations in radio astronomy.

2.7.2 Mitigation Techniques for Solar Interference

Solar interference is an issue that can represent a significant challenge to address when we compare it with other sources of radio interference in radio observations. This is due to its extended radio emission across a wide range of frequencies and angular size. Human-made sources of radio emissions, such as those mentioned in Section 2.7, usually emit a narrow range of frequencies, making its mitigation possible with techniques such as flagging.

The wide range of frequencies emitted by the Sun makes traditional mitigation techniques such as flagging inappropriate for solar interference mitigation, as it can imply flagging all or almost all the data, meaning that we would not have enough data to process and analyse. In this case, a different treatment must be used to remove the Sun's contribution and ensure data preservation and fidelity by recovering the visibilities rather than discarding them.

Mitigating solar interference in radio interferometric spectrometry is an area of research that has seen few studies. This gap can be attributed to the historical context, where observations

were typically conducted using telescopes with lower sensitivity and smaller fields of view. In the past, the Sun's radio emissions were not a significant concern, as the reduced sensitivity and limited FoV of earlier telescopes mitigated the impact of solar interference.

However, with the advancements in telescope technology, sensitive instruments like MeerKAT have emerged, offering unprecedented sensitivity and broader fields of view. These technological advancements have brought the issue of solar interference to the forefront, as the Sun's intense radio emissions can now have a more pronounced impact on observations conducted during the day. The increased sensitivity and broader FoV of telescopes like MeerKAT have highlighted the need to address solar interference and develop effective mitigation techniques. Solar interference was considered a challenge even before sensitive radio telescopes such as MeerKAT existed. A paper published by Sault & Noordam (1995) discusses the problematic issue of solar interference in radio interferometry. While the number of studies addressing this issue is limited, notable contributions have been made to developing effective mitigation techniques. One pioneering study by Sault & Noordam (1995) focused on utilising a continuum subtraction method to eliminate the Sun's contribution to interferometric observations. Their work demonstrated the potential of continuum subtraction in reducing the interference caused by the Sun and other strong sources in the distant sidelobes of a radio telescope. Another recent effort for solar interference mitigation in interferometric data to note is the Sunblocker[1] software, which is also based on discarding (flagging) the affected visibilities. Hence, our approach focuses on recovering them.

In the context of MeerKAT observations, addressing solar interference and developing dedicated mitigation techniques becomes even more crucial. The unique capabilities of MeerKAT, such as its high sensitivity and wide FoV, make it particularly susceptible to solar interference. Therefore, developing a comprehensive approach tailored explicitly for MeerKAT observations is essential.

The study by Sault & Noordam (1995) provided valuable insights into solar interference mitigation, mainly through continuum subtraction. However, the field of solar mitigation in radio interferometry has seen limited exploration, and further advancements are necessary to address the challenges posed by solar interference.

Building upon the foundation laid by earlier research, this thesis aims to contribute to the field by developing the SolarKAT pipeline, a novel and comprehensive approach designed explicitly

[1] https://github.com/gigjozsa/sunblocker

for MeerKAT observations but, in principle, generalisable to other telescopes. The SolarKAT pipeline incorporates advanced techniques such as self-calibration, region-based analysis, deconvolution and peeling to overcome the limitations of previous approaches and enhance the mitigation of solar interference in MeerKAT data.

By acknowledging the earlier work of Sault & Noordam (1995), this research positions itself within the existing literature on solar interference mitigation. The SolarKAT pipeline represents an innovative and tailored solution for addressing solar interference in MeerKAT observations, filling the gap in current mitigation strategies.

2.8 Observation Strategies

The presence of the Sun during observations can pose challenges when observing faint sources. This is caused by the fact that the Sun can emit a wide range of frequencies, which can corrupt the faint signals from the observed sources, making its study challenging. The MeerKAT telescope is not designed to observe strong sources (Kansabanik et al., 2023). To minimise the impact of the Sun during observations, astronomers opt for strategies such as scheduling, where the observations are planned to happen during periods when the Sun is below the horizon, such as in the nighttime or during sunrise or sunset (Kansabanik et al., 2022), frequency selection, where radio astronomers focus on specific frequency bands less affected by solar emissions, and off-pointing (used in Kansabanik et al. (2023)), where the observation is done while the telescope is pointed away from the Sun to avoid its direct emission. In the case of highly sensitive telescopes such as MeerKAT, all these strategies can still introduce interference in the data. As we prove in this work, the telescope can capture the Sun even when pointing 80 degrees away.

2.9 Thesis Motivation

MeerKAT, probably the most sensitive radio interferometer in the world, offers various capabilities for conducting cutting-edge research in various areas of astrophysics. MeerKAT provided many successful daytime observations. However, under certain conditions, daytime observations may be severely affected by the presence of the Sun, which is a strong source of radio interference. The solar interference is so problematic that solar fringes can be visible on individual

baselines of the MeerKAT. These conditions are not thoroughly understood. Hence, this motivates the development of a robust approach to address this problem and improve the quality of MeerKAT observations in the presence of the Sun.

Solar interference can pose a significant obstacle to detecting and study accurately other celestial sources of interest, limiting the sensitivity and fidelity of the observations. Developing robust techniques for mitigating solar interference in MeerKAT observations is essential to ensure the integrity and scientific value of the acquired data. To address this issue, we present the SolarKAT pipeline. A fully automated pipeline that aims to reduce solar RFI in MeerKAT observations. This is a novel approach that uses the existence of RATT software and techniques to subtract the Sun from MeerKAT data. We aim to overcome this challenge and improve the quality of the observed data by developing a dedicated pipeline for solar interference mitigation.

The SolarKAT pipeline includes techniques such as self-calibration, imaging, region-based analysis, sun tracking, deconvolution, sun subtraction, and peeling to mitigate solar interference from the observations effectively.

The successful development and implementation of the SolarKAT pipeline will bring significant implications for radio astronomy. SolarKAT will be a valuable resource for radio astronomers in the data processing stage. By applying advanced techniques and algorithms, SolarKAT will effectively remove solar interference from the observed data, allowing astronomers to obtain high-quality data free from the contamination and effects of solar interference. Using processed and interference-free data will enable astronomers to conduct in-depth studies of celestial objects and phenomena, including galaxies, pulsars, and transient events. SolarKAT will serve not only for solar interference mitigation but also as a powerful tool for routinely high-quality solar imaging. Some scientific products that can be obtained from the pipeline include:

- Characterising solar interference
- Determining the level of solar interference
- Producing high-quality solar images that will be valuable for solar physics and space weather studies
- Understanding the conditions in which solar interference is more prominent

Moreover, the pipeline will contribute to the optimisation of observing strategies and schedules for MeerKAT and future telescopes like the Square Kilometre Array-Mid (SKA-MID). The

insights gained from mitigating solar interference in MeerKAT observations can help inform and guide the operation of these large-scale radio interferometric facilities, maximising their scientific output.

Removing the Sun from MeerKAT observations requires a High Dynamic Range and fidelity spectroscopic imaging of the Sun. The presence of a strong source, such as the Sun, in the far sidelobes of a radio telescope can limit the achievement of high image quality with these observations. Traditional methods like flagging, attenuation and excision manage to subtract RFI from human-made sources (Sault & Noordam, 1995). However, they are ineffective for the Sun because of their complexity and variability (Sault & Noordam, 1995; Kansabanik et al., 2023).

The MeerKAT sensitivity is sufficient to successfully image the Sun even in the sidelobes of the primary beam (Kansabanik et al., 2023) or in the far lobes. A good model of the Sun is all we need to subtract it from our observations effectively. Moving to the next chapter, we present our contribution to addressing the challenges of solar interference discussed previously, the SolarKAT pipeline.

Chapter 3

SolarKAT

This Chapter introduces SolarKAT, a dedicated pipeline that mitigates solar interference in MeerKAT observations. A pipeline is a sequence of processing steps carried out on data, passing the data from one step to another. The SolarKAT pipeline comprises a series of steps that aim to improve the reliability and accuracy of radio astronomy observations conducted in the presence of the Sun. This Chapter presents the details of the SolarKAT pipeline, including its architecture and tools, the pipeline overview and workflow, and the evaluation metrics used to assess the pipeline's effectiveness.

3.1 Pipeline architecture and tools

The SolarKAT pipeline consists of three main components: a Python file, a Cab definition file and a recipe YAML file. The pipeline is based on Stimela (Makhathini, 2017, 2018, Smirnov and Makhathini, inprep), a workflow management framework that facilitates the integration of various tools and libraries for smooth data processing and analysis.

- Python file: serves as the core component of the pipeline, where all the necessary functions, modules and libraries are defined. It contains libraries such as Astropy, Numpy, and Casacore and modules such as re, Glob, sys, os and subprocess to handle astronomical data and perform coordinate transformations and calculations.
- Cab definition file: is a Stimela YAML file that defines the execution environment for the pipeline, including the required software, its dependencies, and configurations. It also

establishes a connection with the Python file by defining each Python function as a cab and defining its inputs and outputs.

- Recipe YAML file: This document describes the pipeline's workflow. It defines the sequence in which each step will be executed in the pipeline and contains inputs and outputs described by the recipe's schema (more information on this can be found in [1]).

The definitions of a Cab and a recipe are based on Makhathini (Smirnov and Makhathini, inprep).

3.1.1 QuartiCal

QuartiCal (Kenyon et al., 2023) is a calibration software that brings a much faster implementation of the calibration process in radio astronomy data. It is a successor to the CubiCal (Kenyon et al., 2018) software. It is used in the SolarKAT pipeline to perform self-calibration on the data. This is also used to peel the Sun to compare its efficiency in peeling extended sources. To the best of our knowledge, this is the first time that QuartiCal has been implemented to such an extended source outside the main field. In Chapter 5, we compare simple subtraction of the Sun from the visibilities using TaQl update with the peeled image from QuartiCal, highlighting the power of QuartiCal in removing the out-of-field Sun and improving the image quality.

3.1.2 WSClean

WSClean (Offringa et al., 2014a) (w-stacking clean), a fast generic imager, is used for several essential steps in the workflow, including imaging, deconvolution and model Sun prediction. WSClean implements several gridding algorithms and offers fully automated multi-scale, multi-frequency deconvolution. WSclean also includes the chgcentre [2] tool. The chgcentre tool recalculates the uvw-values and phase rotates the visibilities to a new given phase centre, and this was done twice in our pipeline: first to change the phase centre to the Sun and the second time to restore the phase centre (from the Sun to the original field).

[1] https://stimela.readthedocs.io/en/latest/fundamentals/basics.html
[2] https://wsclean.readthedocs.io/en/latest/chgcentre.html

3.1.3 Breizorro

Breizorro (Ramaila et al., 2023) is a RATT developed software for mask generation. The pipeline uses it to create masks based on the Sun's region files generated by its Right ascension (RA) and declination(DEC). In the L-band, the mask size is 39.6 arcmin radius and 35.4 arcmin in the UHF-band. These masks isolate the Sun's emissions for each scan and facilitate a more reliable subsequent deconvolution process.

3.1.4 CASA

CASA (Bean et al., 2022) provides powerful functionalities for handling radio astronomy data and performing various data processing tasks. In SolarKAT, it is used for data splitting. It splits the MSs into individual scans using the *split* casa task.

3.2 SolarKAT Pipeline Overview

SolarKAT is a fully automated pipeline designed for solar interference mitigation in MeerKAT observations. This was mainly developed for interference mitigation. However, it is also a valuable tool for high-quality solar imaging, becoming a dual-purpose pipeline. This approach comprises various steps that compose a workflow that is sequentially processed. In summary, the pipeline steps can be outlined as follows:

1. Data Preparation
    - Load the primary (1GC) calibrated MeerKAT Measurement Set (MS) data affected by the Sun.
    - Perform initial data processing (imaging, inspection and masking).
2. Calibration
    - Perform self-calibration (2GC) to improve the calibration solutions of the primary calibrated data and enhance the data quality.
3. Solar Interference Mitigation.
    - Extract the scan numbers from the original MS.

- Split the MS into individual scans based on the extracted scan numbers and name them accordingly.
- Determine the position of the Sun (RA and DEC) relative to the telescope pointing centre per each scan.
- Save the main field RA and DEC.
- Change the phase centre to the Sun's position for each MS.
- Generate the dirty image of the Sun.
- Create regions over the Sun in the dirty images.
- Use the regions to make image masks centred on the Sun.
- Perform deconvolution using the masks centred on the Sun and Sun for each scan.
- Predict the visibility model of the Sun.
- Rephase back each scan's phase centre to the main field.
- Add columns `MODEL_DATA_SUN` and `CORRECTED_DATA_SUN` into the full MS (original MS)
- Copy the model of the Sun (`MODEL_DATA`) from each scan into the `MODEL_DATA_SUN` column in the original MS.
- Subtract the model of the Sun from the visibilities (`CORRECTED_DATA`−`MODEL_DATA`). The output of this operation is stored in the `CORRECTED_DATA_SUN` column created previously.
- Deconvolve the main field (with the Sun subtracted) to generate the improved model.
- Use the improved model visibilities and the model of the Sun to peel the Sun from the original visibilities in the full MS using QuartiCal.
- Deconvolve the final peeled data (`CORRECTED_RESIDUAL`).

These steps are explained in detail in the following section.

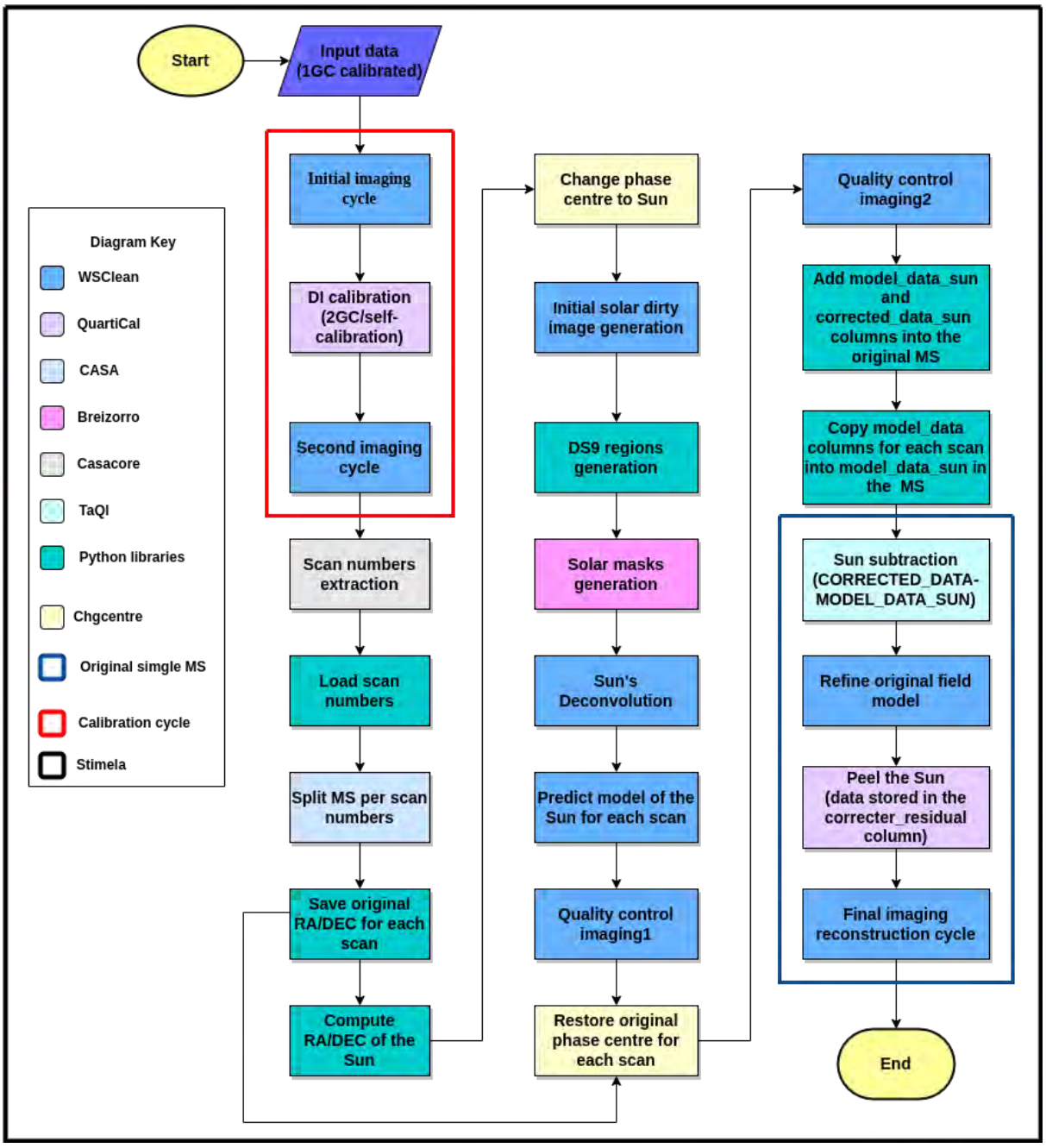


FIGURE 3.1: Proposed fully-automated pipeline workflow for solar interference mitigation in MeerKAT observations. The pipeline consists of 6 main steps: calibration, solar tracking, Sun deconvolution and model prediction, data subtraction+peeling and final image reconstruction.

3.3 SolarKAT Pipeline Workflow

SolarKAT follows a series of steps from data ingestion until the final product (output). This subsection presents the pipeline workflow (see the flowchart in Figure 3.1), outlining the step-by-step process. The SolarKAT pipeline tracks and images the Sun to mitigate solar interference.

A final product of it will be the daytime observations of MeerKAT free from radio interference from the Sun. The SolarKAT workflow is composed of the following steps:

1. **Input data:** The pipeline ingests MeerKAT data in a MS format. This must be at least partially a daytime observation that has been 1GC calibrated and is ready for self-calibration (2GC).

2. **Initial imaging cycle:** After the pipeline ingests the 1GC data and before self-calibration, the next step in the pipeline is imaging. This step generates a radio image of the observed field in the sky. This first imaging step is used to evaluate the quality of the image. These images serve as diagnostic tools to assess the data quality and identify potential contamination from the Sun. This diagnostic is conducted by visually inspecting the output images obtained after the self-calibration process, identifying any potential artefacts that may indicate the presence of solar interference affecting the data.

3. **Direction independent calibration (selfcal):** The 1GC (explained in Section 2.5 in Chapter 2) calibrated data are processed through self-calibration, which is done using QuartiCal software and imaged using WSClean (Offringa et al., 2014b) to generate initial images. In this step, we perform self-calibration to refine the calibration parameters. This step can be done in more than one round if needed.

   After we do our best possible calibration (despite solar interference), which is crucial for the following steps (especially solar imaging), we move to the next step, split the MS.

4. **Second imaging cycle:** After self-calibration, another imaging cycle is performed. After this step, another round of self-calibration can be done if needed (this is manually analysed from the images). After all the calibration and imaging processes, we proceeded to the next step, extracting scan numbers from the MS.

5. **Scan numbers extraction:** Extract the scan numbers in the MS. These are unique numbers used in the following step to split the MS. The scan numbers are stored in a text file.

6. **Load scan numbers:** After extracting the scan numbers and storing them in a text file, we load the text file and use the content to split the MS.

7. **Split the MS per scan numbers:** In this step, the pipeline uses the extracted scan numbers to split the MS into individual scans. It first loads the scan numbers and then

uses them to split the data. Splitting the MS is a crucial step in SolarKAT for several reasons: i) the Sun is in apparent movement relative to sidereal sources due to the Earth's rotation; ii) the variability of the Sun: the Sun's activity varies during the day, splitting the data allows identification of specific times of more or less solar interference. Splitting the data allows us to analyse each scan individually, giving insights into the most prominent time, duration, angle of observation and phenomena causing interference.

8. **Save original RA/DEC for each scan:** In this step, we save the original phase centre coordinates (RA and DEC) into a text file for each scan. These coordinates are further used to restore the original phase centre in the scans (see step 17).

9. **Compute RA/DEC of the Sun:** In this step, the pipeline computes the Right Ascension (RA) and Declination (DEC) of the Sun for future use to change the phase centre in each scan to the Sun position. The RA and DEC of the Sun are determined using the standard ephemeris data and the skycoords from the Python astropy library. The output coordinates of this step are stored in text files. The Sun's RA and DEC are time-dependent and computed per scan by default.

10. **Change the phase centre to the Sun:** Once we obtain the coordinates of the Sun from the previous step, the rephasing step is performed to update the phase centre of the MeerKAT observation to the Sun's position. Each scan separately uses the obtained solar coordinates from the previous step. This step was performed using chgcentre, which is a tool used to change the phase centre of the MS, recalculating the uvw-values (using the information of the antenna locations, phase centre, and time) and phase rotate the visibilities (Offringa et al., 2014b).

11. **Generate dirty images of the Sun:** To ensure that the phase centre of each scan was successfully updated and changed to the Sun, we proceed to the imaging step of each scan. In this step, the rephased data is used to generate an initial image of the Sun, which is then used as a reference for subsequent steps, such as mask creation, which is used in the deconvolution process. These Sun images are also an input for the next step (creating region files) together with the Sun's coordinates. This and all other imaging steps in the pipeline are carried out using WSClean.

12. **DS9 region generation:** In this step, SolarKAT uses the Sun coordinates to generate DS9 region files centred on the solar disk. These regions are then used to generate FITS masks on the Sun for deconvolution. Regions files isolate the region of interest for the

cleaning (deconvolution) process. To create region files, the pipeline uses input images of the Sun from the previous step.

13. **Solar masks generation:** Once the region files have been created, SolarKAT uses Breizorro to create masks of the Sun per each scan using the region files as inputs. The output of this step is a list of FITS files that contain binary values (zeros and ones), where the zeros represent the excluded regions and the ones represent the included regions (represents the region of the data to be cleaned).

14. **Sun's Deconvolution:** In this step, the pipeline uses the mask created in the previous step to deconvolve the Sun using WSClean. WSClean provides an option to use a FITS mask during deconvolution to control the image's reconstruction. In this case, we ensure we clean only the Sun's region for accurate solar model construction. The masks define which areas in the image should be included or excluded during the deconvolution process. This is crucial to ensure that the Sun will be subtracted entirely from the visibilities. Parameters such as the threshold, the mask, the number of iterations, and the multi-scale are critical to ensure the optimum deconvolution process. We can proceed to the prediction step if the deconvolution is good enough.

    To evaluate how good the deconvolution result is, we look at the residual image, which is the difference between the convolved model and the restored image. If the residual image contains emission from the Sun, it is necessary to continue with the deconvolution and ensure it is well done. It can be done by changing the values of the parameters mentioned before, especially the threshold. Because of the Sun's dynamic emission and variability in signal-to-noise ratio, we recommend using a manual threshold (set once) to ensure the balance between removing unwanted sources and, at the same time, preserving the features of the solar emission.

15. **Predict model of the Sun for each scan:** Once the data has been deconvolved, we apply the solar model prediction to fill the MODEL_DATA column with the prediction from the deconvolved image. The Sun model can be predicted using the parameter *-predict* from WSClean (Offringa et al., 2014a). After having predicted the model of the Sun, we can use the model to subtract the Sun from the visibilities.

16. **Quality control1:** This step checks if the model data was well populated during the prediction step.

17. **Restore original phase centre:** After predicting the Sun's model from the visibilities, the next step is to rephase the phase centre from the Sun field to the original (old) coordinates. Up to this step, we worked with the scans with the phase centre pointing to the Sun's field. To assess the effectiveness of the subtraction process, we have to rephase the visibilities to how they were before rephasing to the Sun. We use the same WSClean chgcentre command to rephase the visibilities but now use the original coordinates. This process brings back the original phase centre of the observation, enabling accurate imaging and analysis of the target field without any effects of the Sun's presence in the FoV.

18. **Quality control2:** This step is done to check the effectiveness of the phase centre restoration process.

19. **Add `MODEL_DATA_SUN` and `CORRECTED_DATA_SUN` into the original MS:** In this step of the pipeline we add the `MODEL_DATA_SUN` and `CORRECTED_DATA_SUN` into the MS to further store the `MODEL_DATA` columns from each scan into the `MODEL_DATA_SUN` column in the original MS (this is done in step 20. Additionally, the `CORRECTED_DATA_SUN` will serve as storage for the residual from the subtraction process (in step 21).

20. **Copy `MODEL_DATA from each scan into the MODEL_DATA_SUN in the MS`:** After performing all the processes on the scans, we store the data in the original MS. This is done by copying the model of the Sun and the corrected data sun from the scans into the respective scan number in the full MS. This is necessary to facilitate the peeling process and allow us to have all the visibility products in the same MS.

21. **Subtract the Sun:** The Model of the Sun predicted in step 15 is subtracted from the visibilities for each scan in the scans list. This is done by subtracting `CORRECTED_DATA−MODEL_DATA`, and the result of the subtraction is stored in the `CORRECTED_DATA_SUN` column. After subtracting the Sun, the pipeline performs deconvolution in the main field to reconstruct and refine the model based on the visibilities where the Sun was subtracted. To ensure good results, we use peeling. The peeling process was done using the QuartiCal software discussed in Section 3.1.1, and its significance to this work is presented in step 23. Initially, our project was designed to subtract the Sun as a simple subtraction operation of the Sun from the affected visibilities. However, as we progressed with the project, we explored an additional technique (suggested by Landman Bester) known as *peeling*, which proved highly beneficial for several reasons. Firstly, introducing the peeling technique in

the pipeline process significantly improved the final image quality compared to the simple subtracted image. This approach contributes efficiently to the overall data quality improvement.

Furthermore, including the peeling process made our pipeline more robust against different scenarios and observation conditions. It allowed us to deal more effectively with variations in the data, thereby increasing SolarKAT's adaptability. It is worth noting that while the solar subtraction process is still an integral part of the pipeline, it now acts as an initial step that prepares the data for the peeling process. This approach ensures that our initial model is refined and ready to serve as input to the QuartiCal algorithm, resulting in more accurate and reliable results.

Adopting the peeling technique strengthened SolarKAT's ability to mitigate solar interference, making it a more robust tool for radio astronomy studies. This processing step is performed after the following step (model update).

22. **Refine original field model:** After the Sun has been effectively subtracted from the corrected data, the next step involves deconvolving the `CORRECTED_DATA_SUN` column. This column contains the residual data after subtracting the Sun and is then used to produce the improved model of the main field without the presence of the Sun. This step aims to generate a refined representation of the visibilities. The result of the deconvolution process is an improved model, ideally representing celestial sources' radio emissions with greater fidelity. This improved model is critical for subsequent data analysis, image reconstruction, and scientific interpretation. The refined model is then used with the solar model for the peeling step.

23. **Peel the Sun:** Peeling off the Sun includes three main components: the improved `MODEL_DATA` column, the `MODEL_SUN_DATA` column and the original `CORRECTED_DATA` column.

    - The `MODEL_DATA` represents the improved model (the appearance of the sky without the presence of the Sun).
    - The `MODEL_SUN_DATA` represents the model of the Sun. It is derived from the accurate solar model prediction and is used as a reference to understand the expected contributions of the Sun to the observed data.
    - The `CORRECTED_DATA` is the original observed data containing astronomical sources and the Sun.

The peeling process results are stored in the `CORRECTED_RESIDUAL` column. The Sun has its directional gain term that QuartiCal is solving for. The peeling process involves subtracting the Sun and self-calibrate the main field. Next is to deconvolve the residual image to produce the final image.

24. **Image residual (final image):** This constitutes the pipeline's last step so far, where we image the results of the solar subtraction. The imaging is done with the same parameters as the initial imaging step (before the pipeline implementation in the data) for better comparison and analysis of pipeline effectiveness. The solar imaging pipeline effectiveness was evaluated by comparing the images before and after the pipeline was run.

The quality control steps in the SolarKAT pipeline are imaging steps set with the number of iterations set to 0, the same size as the image of the Sun images, with the 'no_update_model_required' enabled to preserve the visibility models. These steps are distributed as follows: i)after subtracting the Sun, ii)after rephasing the scans back to the old phase centre, and iii) just before peeling to ensure that the columns we were about to use in the peeling process are presented as expected.

3.4 Pipeline Evaluation Metrics

To evaluate the effectiveness of the SolarKAT pipeline, different metrics were employed to measure its performance and compare its results before and after its implementation. These metrics include

1. **Image quality assessment:** Visual inspection of the image quality before and after the pipeline implementation (Figures 4.1 to 5.2). Visual inspection allows an immediate evaluation of the image fidelity and reduction of the artefacts caused by the presence of the Sun in the data. This is done by comparing the outcomes from steps 4, 21 and 24.

2. **Peak flux analysis :** Another quantity we compared was the peak flux. Scatter plots from the Tool for Operations on Catalogues And Tables (TOPCAT) were generated to compare the peak flux of the observed data before and after the pipeline application. These values were obtained by generating source catalogues using PyBDSF (Mohan & Rafferty, 2015). The peak flux represents the maximum brightness of each source in the image field. This metric allows us to visualise the pipeline's impact on the flux density.

3. **Statistical values:** To assess the impact of the pipeline on the data as well as its effectivity, we compared Root Mean Square (RMS) and flux density before and after implementing the pipeline. The RMS provides insights into the noise level reduction achieved by the pipeline. The pipeline's effectiveness can be assessed by comparing the RMS values before and after Sun subtraction. A significant reduction in RMS suggests that the pipeline effectively mitigated the solar interference and improved the signal-to-noise ratio of the data.

4. **Gaussian fit on pixel flux distribution:** A Gaussian fit, also known as a normal distribution, is a statistical function that describes the pattern of a set of data points. This is another valuable metric we used to evaluate the effectiveness of the solar interference mitigation pipeline. When fitting a Gaussian distribution, there are three parameters of interest to analyse the data: amplitude, mean and standard deviation, known as σ. In this context, the amplitude is referred to as the peak of the distribution of the counts of the pixel values. The mean refers to the average (centre) pixel value, and the standard deviation or σ refers to how spread the distribution is around the mean. A small σ generally represents a narrower distribution around the mean, and a larger σ indicates a broader distribution. This metric also tells us the variability of the pixel values in the images pre and post-SolarKAT. Combining the pixel values to the Gaussian fit, we investigated how the pixel distribution changed before and after applying a solar interference mitigation pipeline by fitting Gaussian distributions to the data. This analysis aims to assess the pipeline's effectiveness in reducing or eliminating the impact of solar interference on the observed data. The Gaussian fit could provide insights into the shape and characteristics of the pixel distribution, and any changes in the fitted parameters before and after applying the pipeline may indicate the success of the interference mitigation efforts.

The Gaussian function is represented as:

$$f(x) = A \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.1}$$

where: A is the amplitude, μ is the mean and σ is the standard deviation.

Chapter 4

DATA ANALYSIS AND PIPELINE IMPLEMENTATION

Here, we describe the observations that we worked on. The pipeline was implemented on three datasets with different observed conditions. This Chapter will detail the observations and how we processed and analysed the data for each step of SolarKAT during its implementation.

4.1 Data and data analysis

The observations used in this thesis were conducted by the MeerKAT radio telescope located in the Karoo region of South Africa. MeerKAT is a new-generation radio telescope consisting of 64 dish antennas operating at a frequency range of 580 MHz to 3.5 GHz, with a bandwidth of 856 MHz.

In this research project, we present the results of three radio observations affected by the Sun using the MeerKAT telescope. Two datasets were observed in the L-band (856-1712 MHz) and one in the UHF (580-1015 MHz) band. Although the telescope was not pointed towards the Sun, its emission can still be captured due to the Sun's strong emission in the radio wavelength region, which can cause interference in radio observations even when the telescope is not pointed directly towards it. Each data presented a different level of interference. These are observations from the Large Survey Projects ThunderKAT [1] (Fender et al. (2017)), MIGHTEE [2] (Jarvis

[1]The Hunt for Dynamic and Explosive Radio Transients with MeerKAT
[2]MeerKAT International GHz Tiered Extragalactic Exploration

et al. (2017)) and LADUMA[3] (Blyth (2015)) respectively. Table **??** summarises the details of these observations, including the date, time range, observed band (L-band/UHF), integration time, number of channels, scan duration, and the number of scans for each data.

All observations were conducted at least partially during the day with the telescope tracking a single phase centre. The daytime observations are challenging because of the increased level of solar radio emission and RFI in the radio spectrum. However, daytime observations are necessary to study the Sun's dynamic behaviour and to develop mitigation strategies for solar interference. More details of the data are presented in the following subsections.

4.1.1 ThunderKAT GRS1747−213 L−band

This is a 15-minute ThunderKAT MS observed in the L-band (856-1712 MHz) on 19-Dec-2022 from 08:48 to 09:03 AM. In this MS, the Sun is about 8 degrees north of the field. This is a 1GC calibrated MS, ready to be modelled for self-calibration. This and other observations are excellent cases for this study because the Sun negatively impacted the target field. This seems mainly due to the strong sunspots; thus, a simple u, v cut probably would not fix it. Another feature of this observation is that Sagittarius A (the centre of our galaxy) is detected through the sidelobes about 2.5 degrees NNW from the field centre, requiring the image to be large enough to deconvolve and model that.

4.1.2 MIGHTEE Extended Chandra Deep Field South (E−CDFS)

This is an approximately 4-hour (3.98 hours) L-band observation from the MIGHTEE survey in the E−CDFS field. This observation was taken in February 2022 from 12:03 PM to 4:02 PM, centred on 1284 MHz with 1024 channels and an integration time of 8 seconds—this observation of the J0333-2741 target containing six scans of 30 minutes each was observed with all 64 antennas. The Sun was in the far sidelobe of the telescope between 76,6 (initial position) and approximately 76,7 (final position) degrees from the target field centre. This was a 1GC calibrated data.

[3]Looking At the Distant Universe with the MeerKAT Array

4.1.3 LADUMA E−CDFS UHF−band

This is a LADUMA UHF-band (580-1015 MHz) observation of the J033230-280757 E−CDFS target field, observed by 57 of the 64 MeerKAT antennas from 2020-03-08 10:14:15.753 to 2020-03-08 20:21:55.111 UTC. It is a 10-hour observation of a mean integration time of 8.14 s covering 1024 channels at 544 MHz. In this observation, the Sun was between 80 to 82 degrees from the telescope's pointing position. Raw data had to be corrected for instrumental and gain complex solutions. This process is called first-generation calibration (1GC) or reference calibration, which is calibrated before implementation in the pipeline.

TABLE 4.1: Datasets details.

	ThunderKAT	MIGHTEE	LADUMA
Date	2022-12-19	2022-02-19	2020-03-08
Number of Scans	1	6	23
Scan Duration (min)	15	30	20
Frequency Band	L-band	L-band	UHF
Time Range (UTC)	08:48 - 09:03 AM	12:03 - 16:02 PM	10:14 - 20:21 PM
Integration Time	8 seconds	8 seconds	8.14 seconds
Number of Channels	1024	1024	1024
Angle of separation from the Sun (degrees)	8	76.6 - 76.7	80- 82
Observer	Joe Bright	Ian Heywood	Sharmila

4.2 Data Calibration

Here, we describe the data calibration process according to what we discussed in section 2.4. All the radio data was obtained when flagged, and 1GC was calibrated with OXKAT (Heywood, 2020), which ensured the data was ready for self-calibration.

4.2.1 Direction-independent self-calibration (2GC)

The first step of the process is to make the first image of the data (1GC data) and perform self-calibration on the target data to refine the calibration solutions obtained from 1GC, reducing imaging artefacts. We used the QuartiCal software to derive the self-calibration solutions for our observed data, performing two rounds of phase-only gain self-calibration for all three datasets.

In this process, we corrected the data for phase-only calibration for all three datasets. Self-calibration is a crucial step in the pipeline to refine the antenna gain and correct for instrumental and atmospheric fluctuations' effects. This process is crucial for optimising because poor calibration can lead to poor images of the Sun. The solar features, such as sunspots and active regions, may not be correctly localised or characterised, and this can affect the reliability of the solar radio interference mitigation process. Furthermore, when solar images are used as models for Sun subtraction, poor calibration can introduce inaccuracies in the model visibilities, leading to incomplete subtraction of the Sun from the observed data. A bad calibration solution will affect the quality of the solar images and, therefore, the effectiveness of the subtraction.

The ThunderKAT data was an observation that required special treatment before and after subtracting the Sun. Additionally, we considered it a particular dataset, first because of the presence of the centre of our galaxy, Sgr A, in the FoV, second because the Sun was very close to the FoV in relation to other observations, and last because it presented an extended source which made the improvement of the data after calibration challenging. The position of the Sun in relation to the main field caused the Sun to be aliased in the main field (it can be seen in Fig 4.1), which was solved by increasing the FFT padding parameter from 1.5 to 2.0 in the second iteration of deconvolution before self-calibration. In general, aliasing in astronomical data can be solved by increasing the image size to include the source in the imaging field or by applying the parameter *padding* available in WSClean. The deconvolution was performed first with an auto-threshold and an image size of 14200×14200 pixels to include Sgr A in the image.

Hence, a careful calibration process is essential to ensure the accuracy and reliability of the solar images, which will subsequently enhance the effectiveness of the Sun's subtraction process. As we advanced with the pipeline, after the calibration process, we carried out the intrinsic steps of the pipeline, starting with the splitting step.

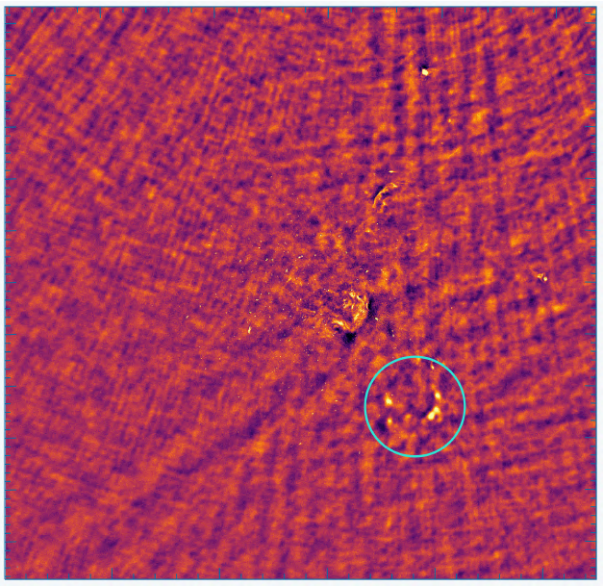


FIGURE 4.1: ThunderKAT GRS1747−213 field image with the sunspots aliased into the field. Observed on December 19, 2022, from 08:48 to 09:08 AM in L-band. The alias of the Sun is in the blue circle in the image field. The vertical banding in the image is due to PSF sidelobes from the out-of-image Sun.

4.3 Split MS

After self-calibrating the data, we split the data into small time chunks (scans). This was done for all the observations using the split task from CASA. The ThunderKAT, MIGTHEE and LADUMA observations have 1, 6 and 23 scans, respectively. Table 4.2 presents the number of scans and angle separation of the Sun for each dataset. The splitting process is done by extracting the scan numbers from the MS and storing them in a text file, which is then loaded as a list of strings and used as input to the split-ms step. Each output in the split-ms step is named based on the original MS and scan number.

TABLE 4.2: Number of scans and Eagle separation of the Sun for each observation.

Observation	**Number of Scans**	**Scan Numbers**
ThunderKAT	1	12
MIGTHEE	6	3, 5, 7, 9, 14, 16
LADUMA	23	3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49

4.4 Getting the coordinates

This step consists of two steps. Firstly, we made a backup of the original phase centre coordinates, Right Ascension (RA) and Declination (DEC) (correspondent to the main FoV), for future use. Secondly, the Sun coordinates are determined in relation to the telescope's pointing position. The primary field coordinates in our pipeline are obtained from the **PHASE_DIR** column within the MS. To present these coordinates in a format that can be used in the following steps, it passes for conversion from degrees into a sexagesimal representation, which employs hours, minutes, and seconds (HH:MM:SS) for RA and degrees, arcminutes, and arcseconds (DD:MM:SS) for Dec. This conversion allows astronomers and researchers to interpret and communicate the precise positions of celestial objects easily. After we get the RA and DEC, we use them to track the Sun, which is explained in the next Section. Both the Sun's and main field coordinates are recorded into text files.

4.5 Change phase centre to the Sun

In this step of the SolarKAT pipeline, we adjusted the phase centre of each scan to align with the position of the Sun. Initially, we used the *shift back* parameter, which is a WSClean parameter to shift the coordinate system along the tangent plane, causing the distortion observed in the right panel of Figure 4.2. This distortion exhibits an oval appearance of the Sun observed, which differs from the expected perfectly circular disk. The documentation shows this distortion is attributed to the coordinate system shift along the tangent plane. This allows for imaging in a different projection centred on the target position but with smaller w-terms [4]. Meanwhile, the chgcentre command rotates the coordinates system, adjusting the phase centre to the image centre. The difference between the function of these two parameters is shown in Figure 4.2. The chgcentre command rotates the coordinate system, whereas WSClean's shift parameter shifts the coordinate system along the tangent plane. The chgcentre page provides further information about the chgcentre [5] parameter. In other words, in this case, the most indicated is using a chgcentre. An illustration of the distortion caused by the *shift back* parameter is the right panel of Figure 4.2. Figure 4.2 compares the shifted and chgcentred images of scan 7 of the MIGHTEE data. This adjustment was crucial to isolate and subtract solar interference from the observations effectively. By aligning the observation to the Sun's position, we ensured that

[4] https://wsclean.readthedocs.io/en/latest/w_snapshot_algorithm.html
[5] https://github.com/ratt-ru/solarkat/discussions/7

the subsequent pipeline steps were performed accurately on the solar data. The phase centre adjustments to the Sun's position were carefully calculated based on the Sun's coordinates and the telescope pointing centre for each scan. The successful execution of this step is evident in Figures 4.3 and 4.4. During the initial stage of the pipeline development, we utilised the WSClean *shift back* [6] parameter to shift the phase centre to the Sun and back. However, this tool was observed to introduce distortion in the resulting images, making them appear oval-shaped instead of the expected circular shape of the solar disk. We used the **chgcentre** command to address this distortion. This tool proved effective in changing the phase centre and producing the desired results, preserving the circular shape of the Sun.

We realise that the image is distorted and does not appear as it should, as a perfectly circular disk, but with an oval shape, and that, according to the literature, is because it uses the shift command, which has a function different from the chgcentre command.

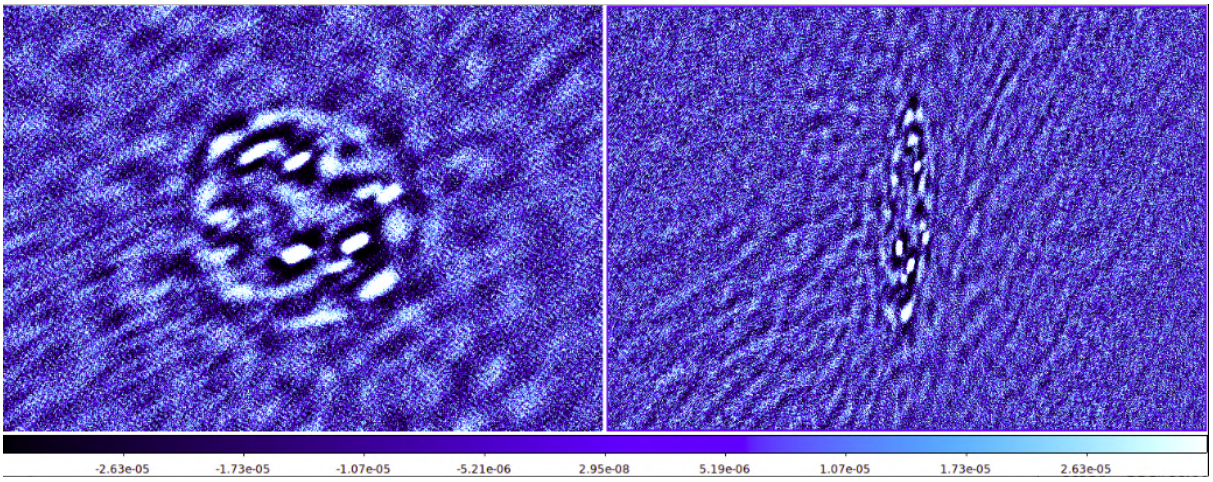


FIGURE 4.2: Comparison of images from scan 7 in the MIGHTEE project. The left panel depicts the result when updating the phase centre using the `chgcentre` command, while the right panel shows the image with the phase centre updated using the `shift` command.

4.6 Dirty image of the Sun

In this step, we generated dirty images of the Sun (see Figure 4.3 and 4.4)using the adjusted phase centre obtained from the previous shift to the Sun step. There is an image per scan for observations that have multiple scans. This step served two main objectives in the SolarKAT pipeline. The first one serves as a validation or control step to ensure the success of the phase centre adjustment in the data, and the second uses the images to guide the creation of masks of

[6] https://wsclean.readthedocs.io/en/latest/chgcentre.html

the Sun for the deconvolution process. The successful generation of the dirty image confirmed that the observation's main phase centre was accurately shifted to the Sun's position. This validation was crucial, as any misalignment could have led to inaccurate results in subsequent pipeline steps. Furthermore, the dirty image was a valuable input for generating FITS masks of the Sun. By generating regions over the Sun in the dirty image, we created FITS mask files that precisely defined the areas of interest for further analysis. These region-based masks played a significant role in subsequent deconvolution and model subtraction steps. Images obtained after the successful phase centre shift demonstrate the effectiveness of the pipeline's calculations. Images show the Sun's appearance in the dirty state for all our observations as a visual confirmation of the phase centre accuracy.

Each scan of the LADUMA observation has a duration of 20 min. In this observation, the Sun is present in the telescope FoV up to scan 33 (observed from 4:34 PM to 4:54 PM), where the Sun was located 64.86 degrees away from the telescope's pointing position. Scan 35 shows signs of the solar disk, but the emission is faint and completely disappears from the FoV. In the MIGHTEE observation of the E−CDFS field, each scan has a time resolution of 30 min, and the single scan of the ThunderKAT observation has a time resolution of 15 min. The Sun's apparent motion in the sky is of the order of an arcmin within the course of a scan; this is addressed by changing the phase centre of each scan from the main field to the FoV of the Sun. In the MIGHTEE observation, the Sun is visible in all scans; however, in scan 16, observed from 3:21 PM to 3:52 PM, the solar disk appears distorted (which could be ionospheric refraction). The Sun was 76.64 degrees from the main field centre in this scan.

Surprisingly, in the MIGHTEE observation, we can still see the Sun observation even being 76.64 degrees away, differently from the LADUMA observation, where the Sun appears for the last time in the LADUMA scan where the Sun was at around 64.85 degrees in relation to the antenna orientation, which is less than distance where the Sun is seen in the MIGHTEE data.

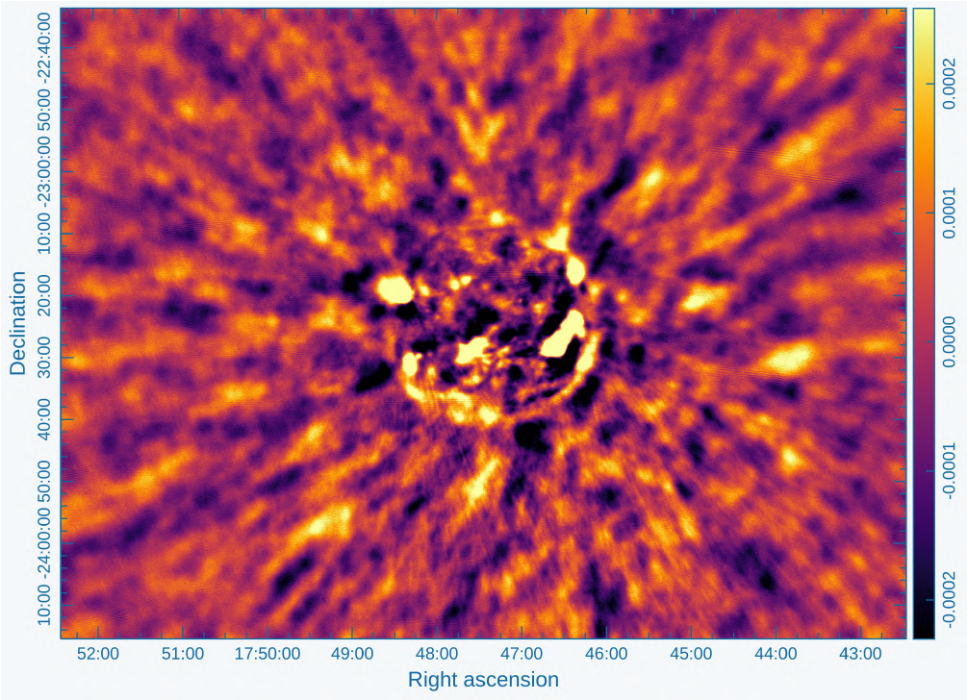


FIGURE 4.3: Dirty image of the 15-minute single scan from the ThunderKAT observation, observed on December 19, 2022, from 08:48 AM in L - band.

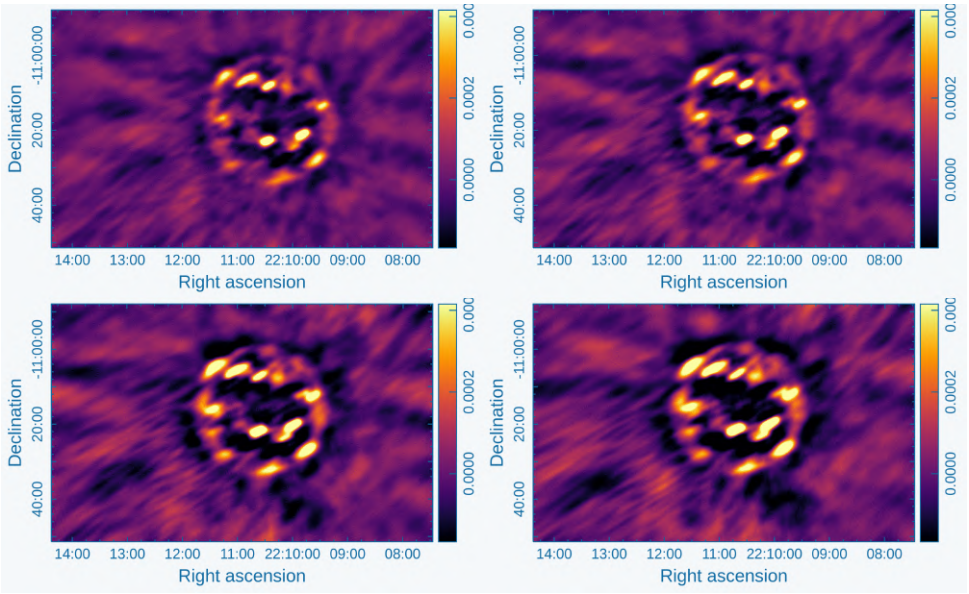


FIGURE 4.4: Dirty images of scans 3 (top left), 5 (top right), 7 (bottom left), and 9 (bottom right) in the MIGTHEE observation, observed on February 19, 2022, from 12:03 PM to 4:02 PM. Scan 3 started at 12:14 PM, scan five at 12:48 PM, scan seven at 1:21 PM, and scan nine at 1:54 PM.

4.7 Region creation

We generated region files of the Sun for each scan using the Sun coordinates (RA and DEC) determined in Section 4.4. The generated regions for each observation successfully delineated the areas of the image corresponding to the solar emission. The regions were created based on a Python function that takes the file containing the coordinates of the Sun, converts these coordinates from HMS and DMS to degrees and generates a circular region with a radius of 1188.0000 arcseconds (approximately equal to 0.33 degrees). This value was manually determined by defining a region over the Sun in CARTA and inspecting the values corresponding to the region, as the Sun's diameter default value (0.5 degrees) was higher than the actual value in the image. Figure 4.5 shows the generated regions on the solar images for the MIGHTEE dataset as a representative example.

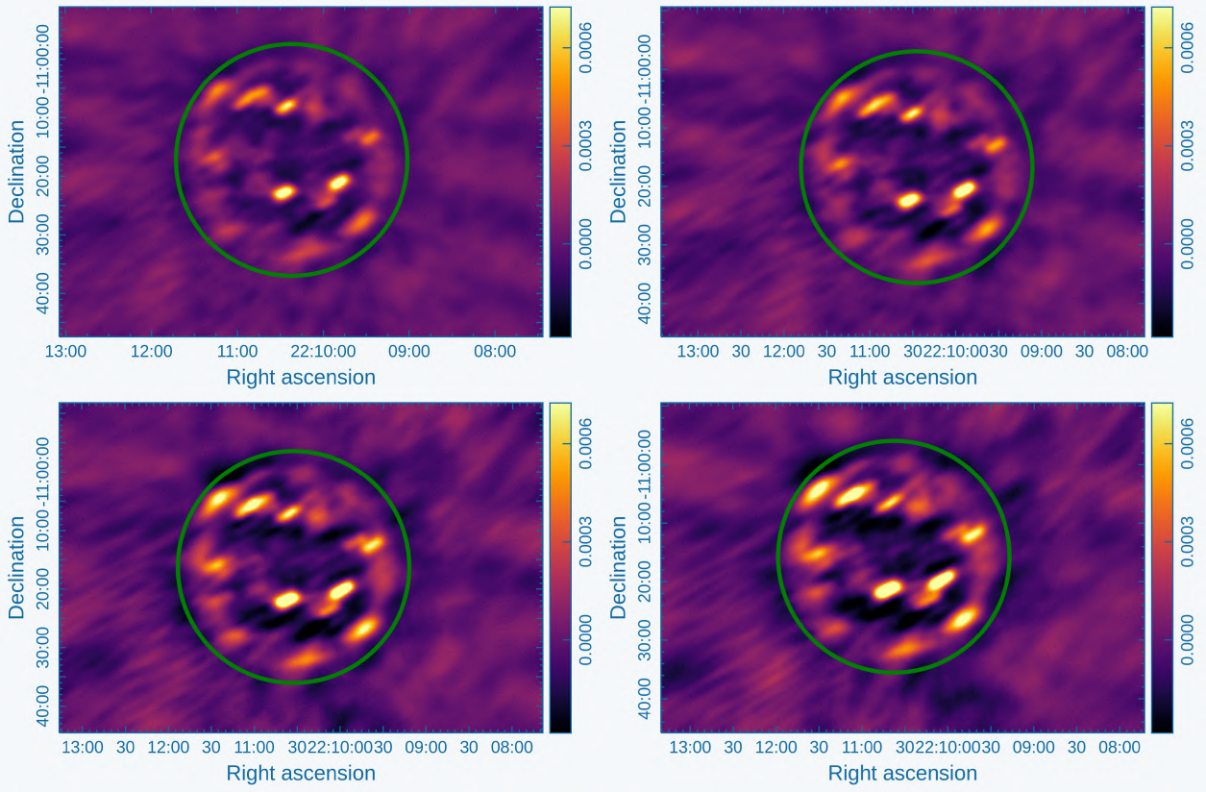


FIGURE 4.5: Regions generated for the MIGHTEE observation, observed on February 19, 2022, in L-band. Scans 3 (upper left, 12:14 PM), 5 (upper right, 12:48 PM), 7 (bottom left, 1:21 PM) and 9 (bottom right, 1:54 PM).

These generated regions were utilised to create FITS masks, crucial for accurate deconvolution and solar subtraction. This process occurs similarly for all datasets.

4.8 FITS Masks generation

As outlined in the workflow description, SolarKAT employed a mask creation process to generate masks centred on the Sun. The regions generated in the previous step generated region-based FITS masks for each time chunk. This is a crucial step of the SolarKAT pipeline as it helps to define regions of interest for subsequent data processing. For each of the three observations, the pipeline generated masks by utilising the position of the Sun in relation to the telescope-pointing centre, isolating the Sun from the rest of the field. Figures 4.6 and 4.7 are presented on the left and right, respectively. The left panel illustrates the dirty images and their respective DS9 regions, while the right panel shows the regions and their generated masks for the ThunderKAT and MIGHTEE observations. These masks were generated with a diameter of 39.6 arcmin for the ThunderKAT (L-band), and MIGHTEE was observed in L-band and 35.4 arcmin's for the LADUMA in the UHF-band. The L-band diameter of the Sun is larger because we had to ensure including the emission (in the left upper region of the Sun in Figure 4.3) from the active region in the DS9 region and mask. This demonstrates that the Sun's angular size can vary in different observation conditions, such as the observed frequency. At lower frequencies, we see thermal radio emission from the corona at higher altitudes than the solar disk.

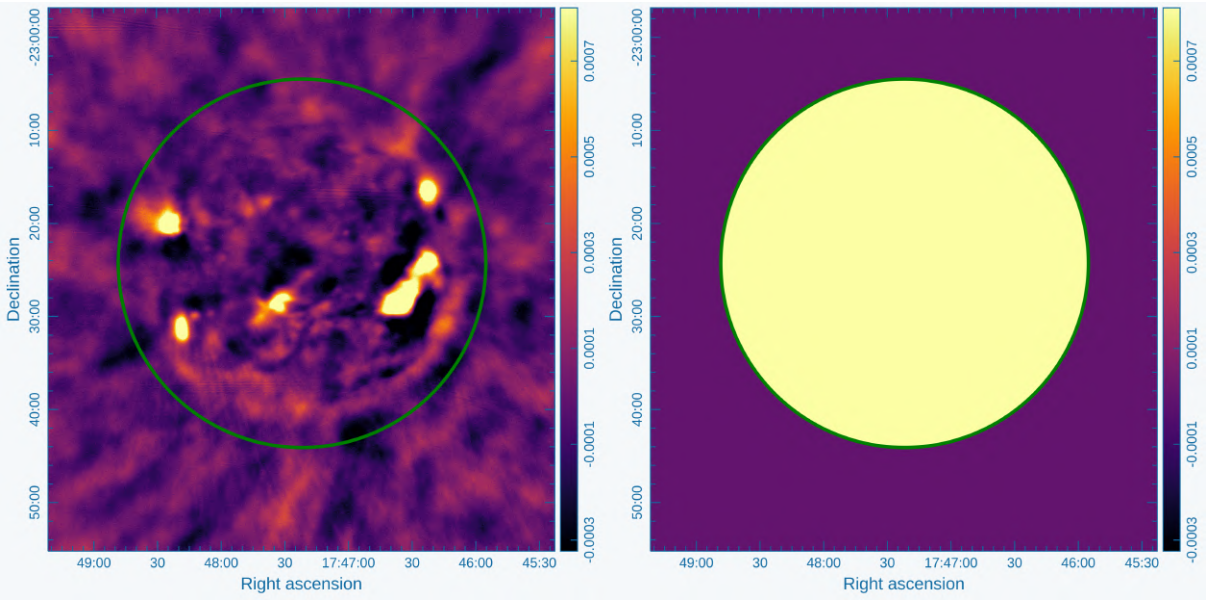


FIGURE 4.6: Mask and its corresponding region over the Sun dirty image in the ThunderKAT observation, observed on December 19, 2022, in L - band. A single scan was observed from 08:48 AM to 09:03 AM.

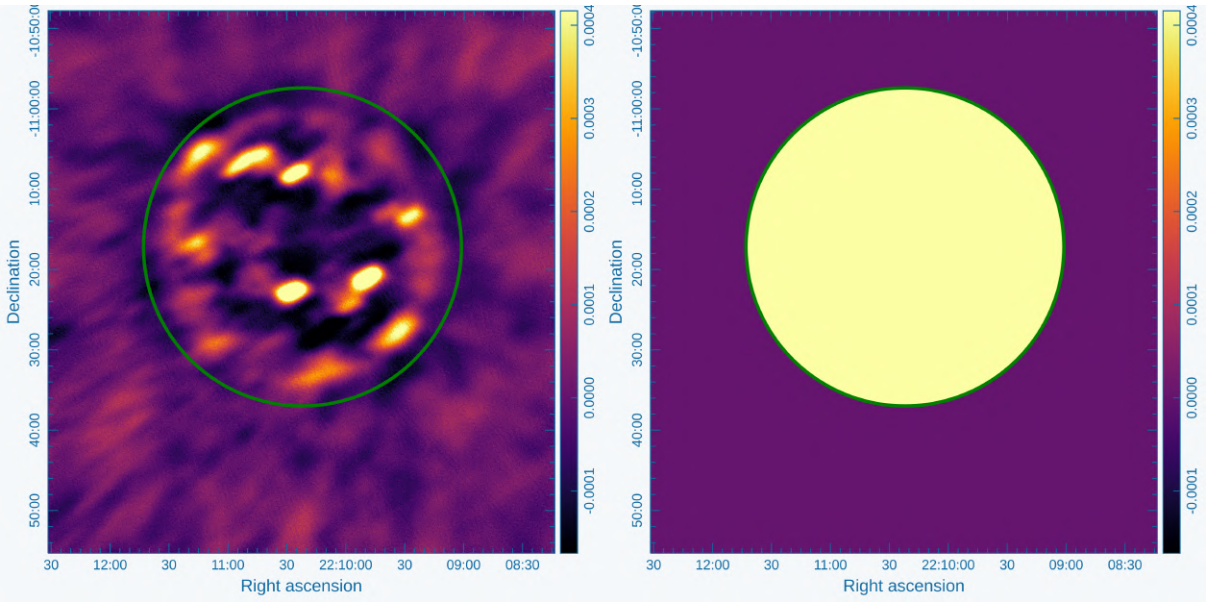


FIGURE 4.7: Mask and its corresponding region over the Sun dirty image in the MIGHTEE observation, observed on February 19, 202,2 in L - band. Scan 3, observed at 12:14 PM.

4.9 Deconvolution

After mask generation, the pipeline uses the WSClean (mentioned in Section 3.1.2) software to perform deconvolution (discussed in Section 14) centred on the Sun. We performed deconvolution using a manual threshold of 65.6 µJy, 43.8 µJy, and 92.6 µJy for the ThundeKAT, MIGTHEE and LADUMA observation. Because the Sun is an extended source with diffuse emission and present characteristics in various scales, we used WSClean multiscale deconvolution for all the scans. We used the following parameters settings: Image *size* 6000, 100000 iterations, *column* `CORRECTED_DATA` and *join channels* set to *true* for all the scans. Table 4.3 summarises the parameters used for the Sun deconvolution process. The deconvolution process occurs in several sub-bands to capture spectral variations for the model. An example of how deconvolution improved the image of the Sun is shown in the right panel of Figure 4.8.

4.10 Model prediction

After the deconvolution of the solar emission from the observed data, the next step involved the theoretical prediction of the model of the Sun using multiscale, discussed in Section 2.6.2. This process was done using WSClean software discussed in Section 3.1.2. This model (visibility

model), denominated `MODEL_DATA_SUN` in the peeling process, represents what the Sun's emission is expected to be based on its spectral and morphological characteristics and position in relation to the telescope's pointing centre.

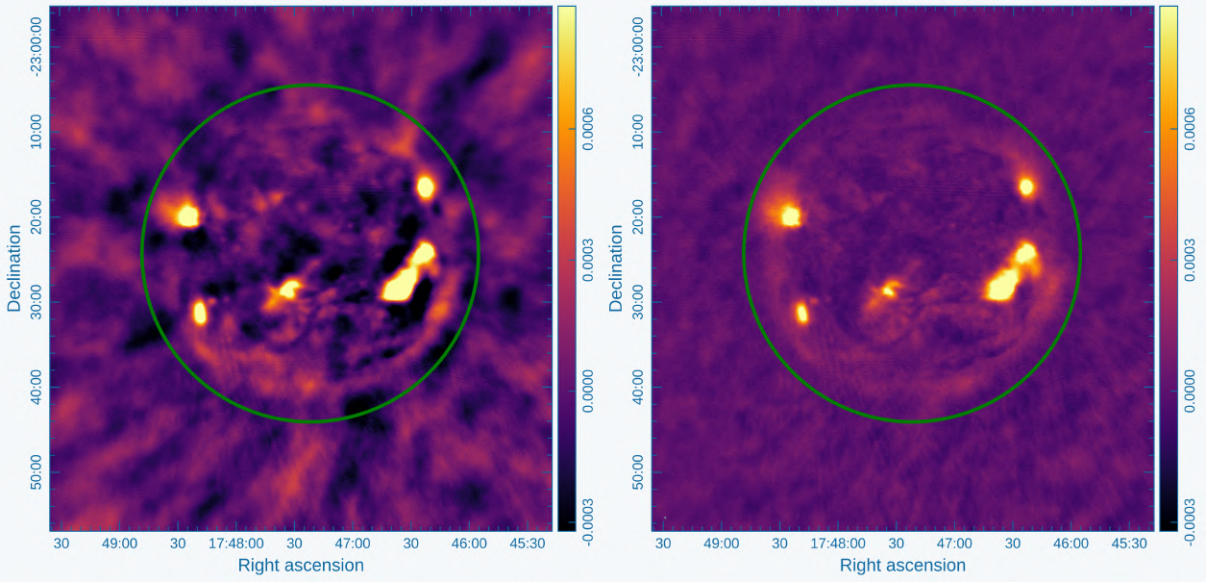


FIGURE 4.8: ThunderKAT, 15 min single scan image of the Sun observed in L-band on December 19, 2022. Left panel: Dirty image of the solar disk. Right panel: Deconvolved (restored) image of the solar disk. The solid line on the solar disk represents the DS9 region generated on the Sun. The size of the restoring beam for this image is 7.25 arcseconds for the Beam major axis (BMAJ) and 6.52 arcseconds for the Beam Minor Axis (BMIN), with the Beam Position Angle 157.87 degrees.

TABLE 4.3: Parameters used for the Sun's deconvolution step.

Dataset	**Threshold (μJy)**
ThunderKAT	65.6
MIGHTEE	43.8
LADUMA	92.6
Common Parameters	
Join Channels	true
multiscale	true
Image Size	6000
Number of Iterations	100000
Column	CORRECTED_DATA

The accuracy of the predicted model was analysed by comparing the predicted emission with the observed emission and the spatial distribution of the Sun. The model's accuracy was also analysed from the residual image in the deconvolution process, where a residual image without emission represents a good deconvolution process and good model generation. The prediction step is essential to ensure the accuracy of the subtraction and peeling process in the observed

data. The results of the model prediction step indicated a strong agreement between the predicted model and the observed emission. While the deconvolution process updates the model based on the observed data, the prediction step is essential to ensure the completeness of the model. It also confirms that the deconvolution process did not introduce unexpected artefacts or distortions in the data. The spatial distribution also aligned with the observed data, demonstrating that the predicted model successfully captured the spatial aspect of the Sun's emission distribution in the observed data. The accuracy of the model prediction enhances the reliability of the subsequent subtraction and peeling process.

4.11 Phase centre restoration

Before subtracting the Sun, we store the `MODEL_DATA` from each scan in the `MODEL_DATA_SUN` column, which is carefully added in the original MS. After all the previously performed steps in the pipeline per scan, the pipeline returned the scans to the main phase centre as it will allow us to access the changes on the main field observation after the Sun subtraction process.

4.12 Sun subtraction

After predicting the model of the Sun, the pipeline proceeded to subtract the Sun model from the corrected data. This process was done for each split scan. The 'simple' subtraction is a foundational step in refining the sky model, essential for the subsequent peeling step. Without this initial subtraction, the peeling step would not be as effective, compromising the quality of the final model. Figure 4.9 shows the deconvolved image of the Sun on the left, and the right panel shows the Sun's field after solar subtraction. The bright features in the image are regions of the corona or sunspots where the magnetic field is strong. These regions are called active regions, and they are associated with solar flares and other solar activity. Figure 4.10 shows the field of the Sun for the MIGHTEE observation after subtraction.

The diagonal stripes in the images are likely to be associated with interference caused by another source, which is not the Sun but is probably a satellite. This process underscores the importance of the "rough" subtraction phase in enhancing the accuracy of the sky model, ensuring a higher quality model for further analysis.

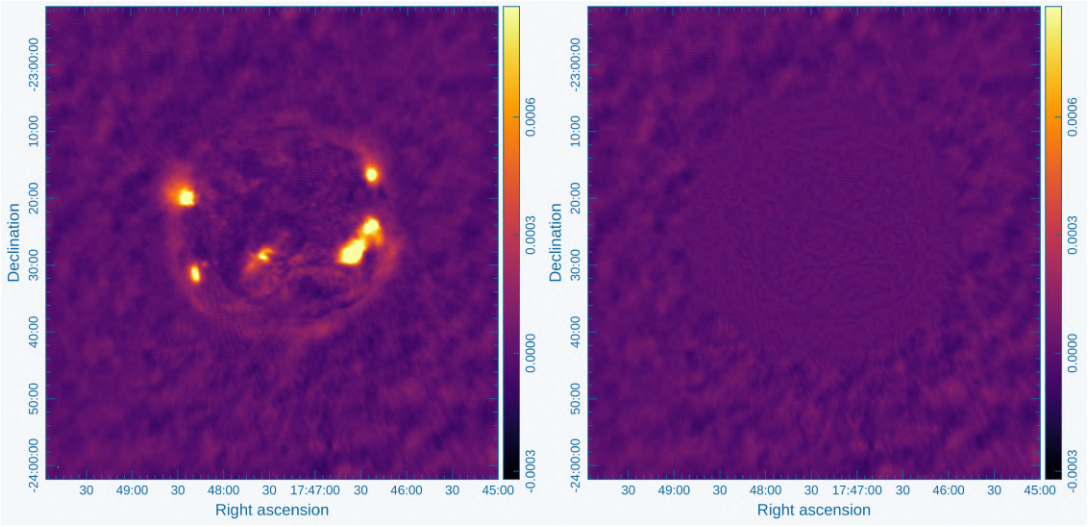


FIGURE 4.9: Image of the Sun field from the single scan ThundeKAT observation before and after subtraction in L-band. December 19, 2022, 08:48 AM.

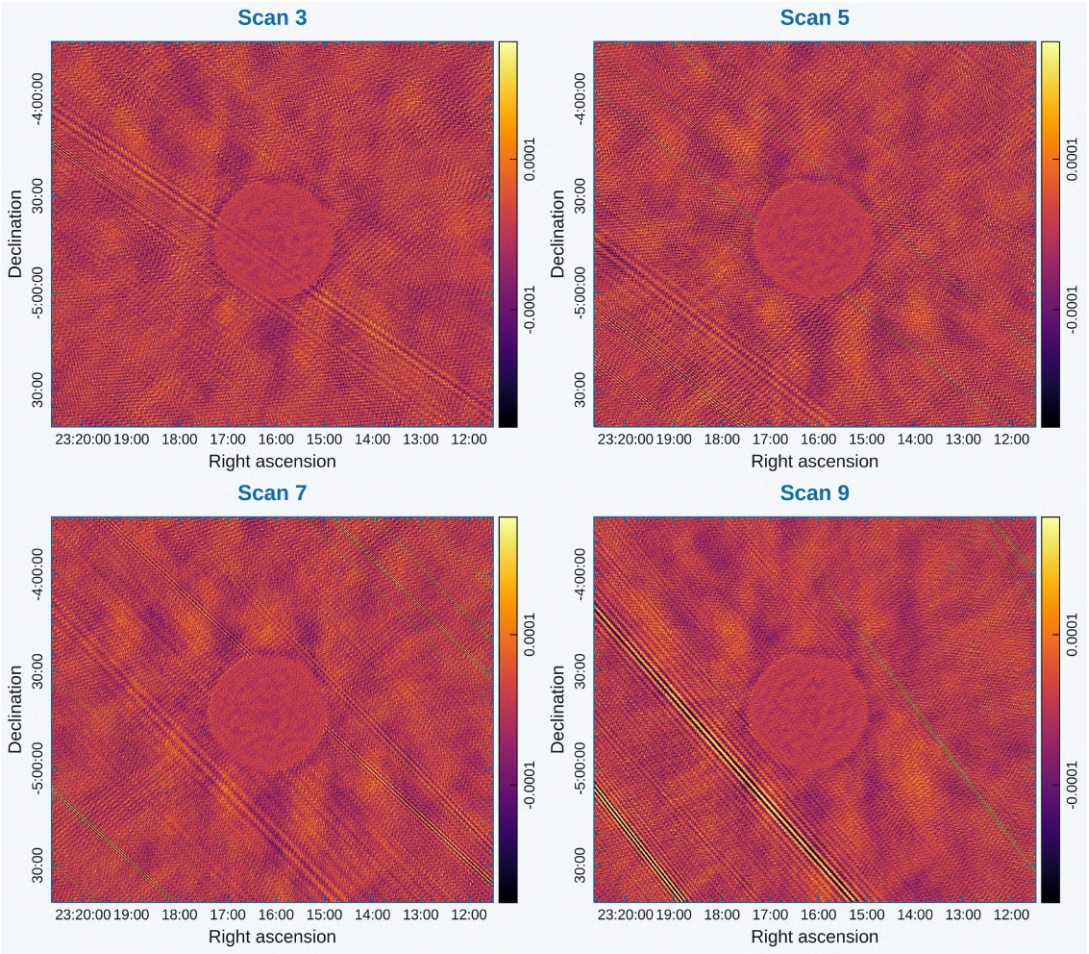


FIGURE 4.10: Snapshot images of the Sun's field after subtraction in the MIGHTEE observation in L-band. Scans 3,5,7 and 9 were observed from 12:14 PM to 2:24 PM, respectively.

4.13 Image reconstruction after subtraction

We performed deconvolution in the `CORRECTED_DATA_SUN` column (where the difference between the `CORRECTED_DATA` and `MODEL_DATA` was stored. This resulted in an improved restored image and enhanced model of the visibilities.

After subtracting the Sun's contribution, the deconvolution process focused only on the signals originating from the sources within the main field. This approach slightly improved image quality; this can be observed in the right panel of Figures 5.1, 5.4 and 5.7.

4.14 Peeling off the Sun with QuartiCal

After the process culminating in the Sun model generation and the improved model visibilities (the visibilities with the Sun subtracted), we tested the peeling process to evaluate whether peeling the Sun would give us the same or better results than subtracting it from visibilities. The peeling process was done using QuartiCal, discussed in Section 3.1.1. This process included, as input models, the predicted model (`MODEL_DATA_SUN`) of the Sun and the `MODEL_DATA`, which is the observed visibilities without the contribution of the Sun. The product of the peeling process was stored in a column named `CORRECTED_RESIDUAL`. Other parameters included the phase delay solver.

4.15 Image reconstruction after peeling

The residual data results from the peeling process, which was demonstrated to enhance the final image more than a simple subtraction operation done in Section 4.12. Figures 5.2, 5.5, and 4.6 compare the images generated by simple subtraction and peeling for the different tested observations.

Chapter 5

Solar Interference mitigation in MeerKAT

This chapter presents the results and discussion of the pipeline implementation. This chapter aims to evaluate the effectiveness of the SolarKAT pipeline in removing solar interference and improving the image quality of the observed data.

The SolarKAT pipeline was applied to three different datasets from MeerKAT observations, each presenting a unique observational scenario. As referred to in Section 4.1, in the first observation of the ThunderKAT data, the source of solar interference was located 8 degrees away from the pointing centre. The second observation (MIGHTEE) of the E − CDFS field, the Sun, was approximately 76 degrees from the telescope pointing centre. In the third observation (LADUMA dataset), the Sun was located approximately 80 degrees away from the pointing centre.

The different scenarios of these observations allowed us to assess the effectiveness of the SolarKAT pipeline in different conditions. The following sections will focus on evaluating the effectiveness of the solar interference mitigation pipeline in each dataset. This will include comparing visual images for image fidelity, peak brightness and statistical values, including RMS and total flux density. To assess these evaluation metrics, we used PYBDSF [1] for source finding and CARTA [2] (for visualisation).

[1] https://pybdsf.readthedocs.io/en/latest/ug_basics.html
[2] https://carta.readthedocs.io/en/latest/#

5.1 ThunderKAT analysis

Analysis of pipeline implementation on the ThunderKAT observation.

5.1.1 Visual inspection/Image quality improvement

Our analysis begins by comparing the images before any interference mitigation with those after interference mitigation (simple sun subtraction and peeling). Figure 5.1 shows the images of the GRS1747−312 field before and after subtracting the Sun. We compared the results of a simple subtraction operation with those from the peeling process implemented in the pipeline. This is illustrated in Figure 5.2. The initial image (left panel of Figure 5.1) shows the main field with the Sun affecting the image quality.

The vertical stripes in the left panel of Figure 5.1 are PSF sidelobes associated with the out-of-field solar radio emission. The simple subtracted image (right panel of Figure 5.1) represents the result after the initial subtraction process. Here, the Sun's contribution is reduced, but some artefacts remain in the field. Figure 5.2 compares the simple subtraction process with the image after peeling. This image shows the result after the complete application of the pipeline, where the Sun's presence and its artefacts are no longer visible. The residual sidelobes in this field (right panel in Figure 5.2) are due to incomplete deconvolution of the target field caused by the extended source's diffuse emission in the centre of the field and are not associated with the Sun.

The images revealed significant improvements in the FoV. It is notable from the right panel image illustrated in Figure 5.2. As we can see, the stripes introduced by the Sun presence were significantly reduced after applying the pipeline. The incomplete deconvolution of the extended source in the centre of the field causes the remaining structures in the final image of the GRS1747−312 field.

5.1.2 Peak Flux analysis for the GRS1747-312 field

To assess the impact of the SolarKAT pipeline on the flux measurements, we analysed the peak brightness values for the sources in the main FoV before and after pipeline implementation. To conduct these analyses, we first generated catalogues of sources in the FoV for each image before and after implementing the pipeline. These catalogues were generated using PyBDSF.

We then cross-matched the catalogues using a tolerance of one arcsec. Using the cross-matched catalogues, we plotted the peak flux for each case: i) peak in the initial image versus the peak after simple subtraction, ii) peak after simple subtraction versus after peeling and iii) initial versus after peeling.

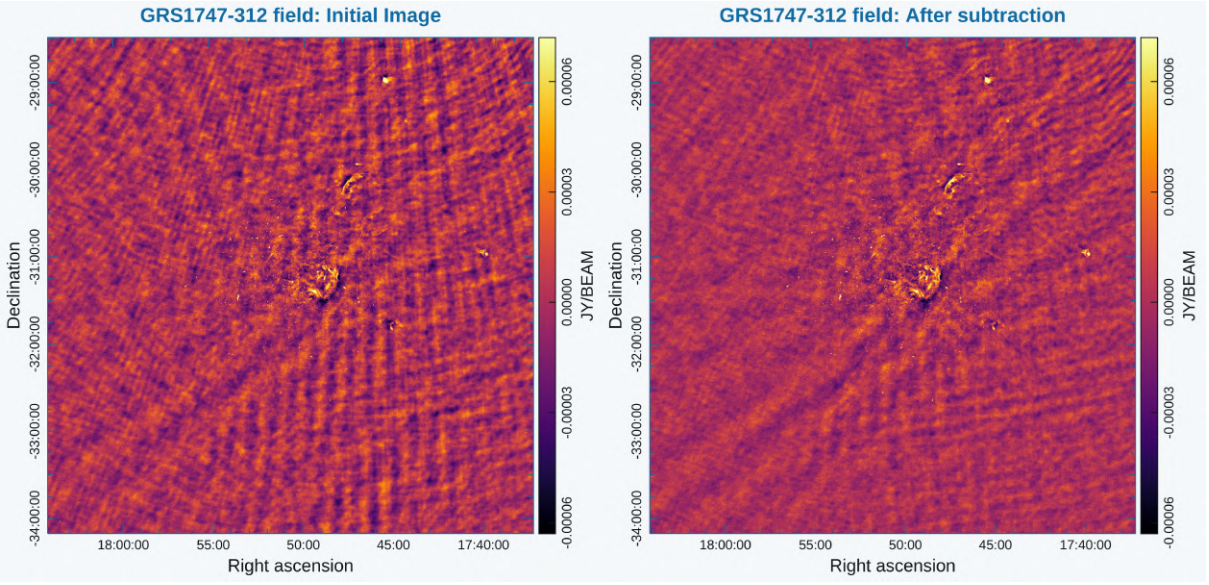


FIGURE 5.1: ThunderKAT main field before and after solar subtraction. The left panel shows the main field image before solar interference mitigation and after self-calibration. The right panel shows the same region after solar subtraction.

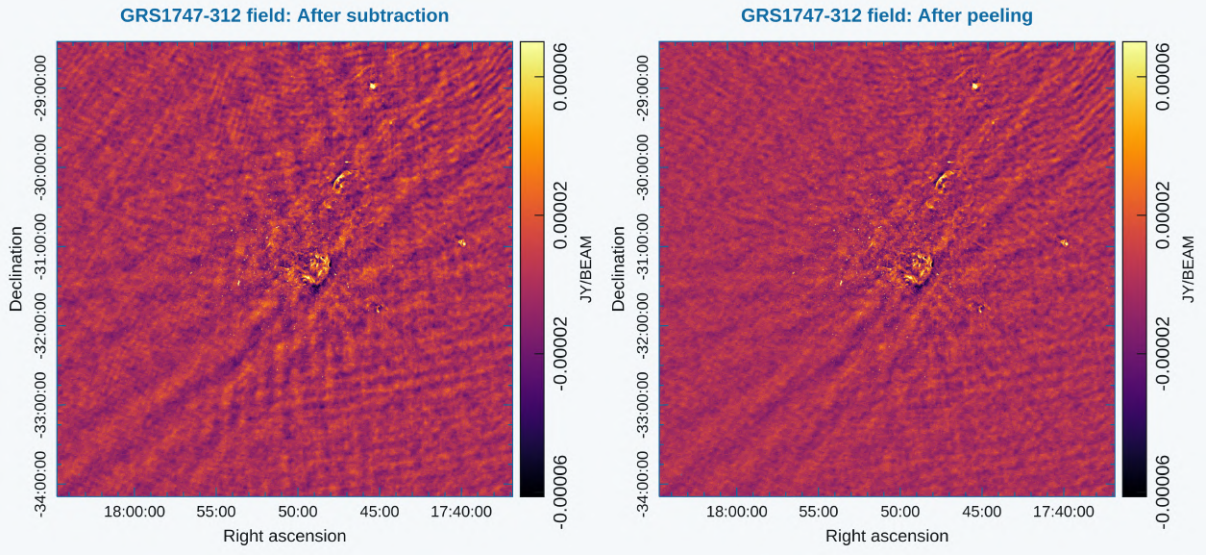
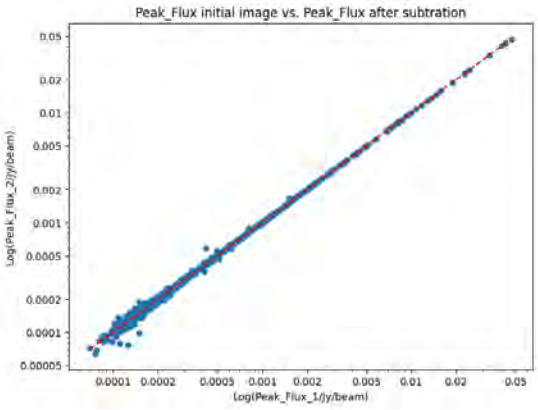


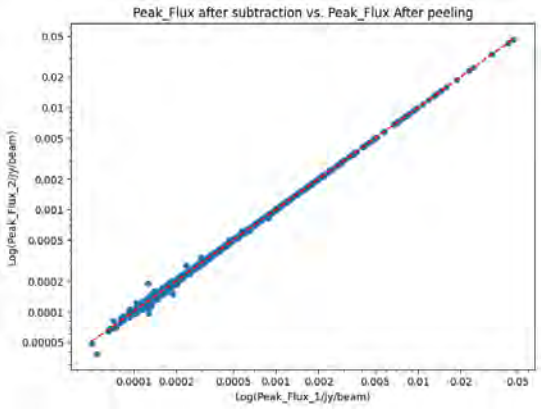
FIGURE 5.2: ThunderKAT main field after simple subtraction and after peeling. Left: After subtraction. Right: After peeling.

The plots in Figure 5.3 show the behaviour of the peak flux values initially, after subtraction and after peeling. The peak flux tells us the maximum intensity of radio waves collected from

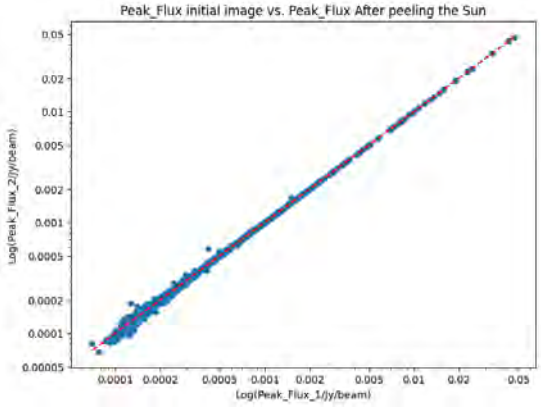
a celestial source, serving as a critical parameter to assess the impact of the SolarKAT pipeline on the data. This value is recorded in Jansky per beam (Jy/beam) units, where 1 Jy is 10^{-26} Watts per square meter per Hertz. A "beam" refers to the synthesised beam pattern of a radio telescope. The beam pattern is a model of how the antennas combine their signals to form a single, coherent image of a source.



(a) Peak flux of the initial image vs. after subtraction.



(b) Peak flux after subtraction vs after peeling.



(c) Initial peak vs after peeling.

FIGURE 5.3: Peak flux plots with 1:1 line for reference. the 1:1 diagonal line represents the perfect agreement between the peak flux values before and after the pipeline.

The plots show the relation between the peak flux values before (x-axis) and after (y-axis), subtracting the Sun and after peeling. The 1:1 line shows the agreement between the two values. Points around the line generally indicate a high degree of agreement and suggest that the pipeline preserves the overall distribution of the peak flux values. Scattered points around the line might indicate new faint sources previously hidden by the sidelobes from the Sun's radio emission (more details on the newly detected sources are discussed in Section 5.5.1). It is possible to observe a peak flux consistency in all the cases before and after applying SolarKAT.

The red dashed line is a one-on-one line representing the relationship between the plotted flux values before and after applying the pipeline. The scatter points in the lower left of the graph are sources that turned detectable after removing the contribution from the data.

After applying the pipeline, the peak flux was consistent, suggesting that the pipeline is preserving the flux measurements in the data, which aligns with our primary goal. In Figure 5.3 *a*, *b* and *c*, we observe 0.05 Jy/beam peak flux before mitigation, which remains the same after subtraction and peeling.

5.2 MIGHTEE analysis

This section presents the results and analysis of the pipeline implementation in the MIGHTEE observation.

5.2.1 Visual inspection

Figure 5.4 compares the initial and subtracted Sun images of the MIGHTEE field. The initial image (in the left panel) presents accentuated stripes as fringes in the field caused by the out-of-field solar radio emission. In this observation, the relative distance of the Sun is about 64 degrees from the telescope pointing centre. Part of the fringes remain in the image field before simple subtraction of the Sun, visible in the right panel of the image. However, the peeling process presents better results in removing the Sun and all its contributions to the data, as shown in the right panel of Figure 5.5.

5.2.2 Peak flux analysis

Similarly to the ThunderKAT observation, we compared the peak fluxes in the initial image, the subtracted image, and the peeled image to analyse their respective impacts of the pipeline on the data for the MIGHTEE field. Here, we also observe some scattered points that suggest the appearance or detection of new objects in the FoV. All the plots show a consistent peak flux of 0.07 Jy/beam after applying the pipeline.

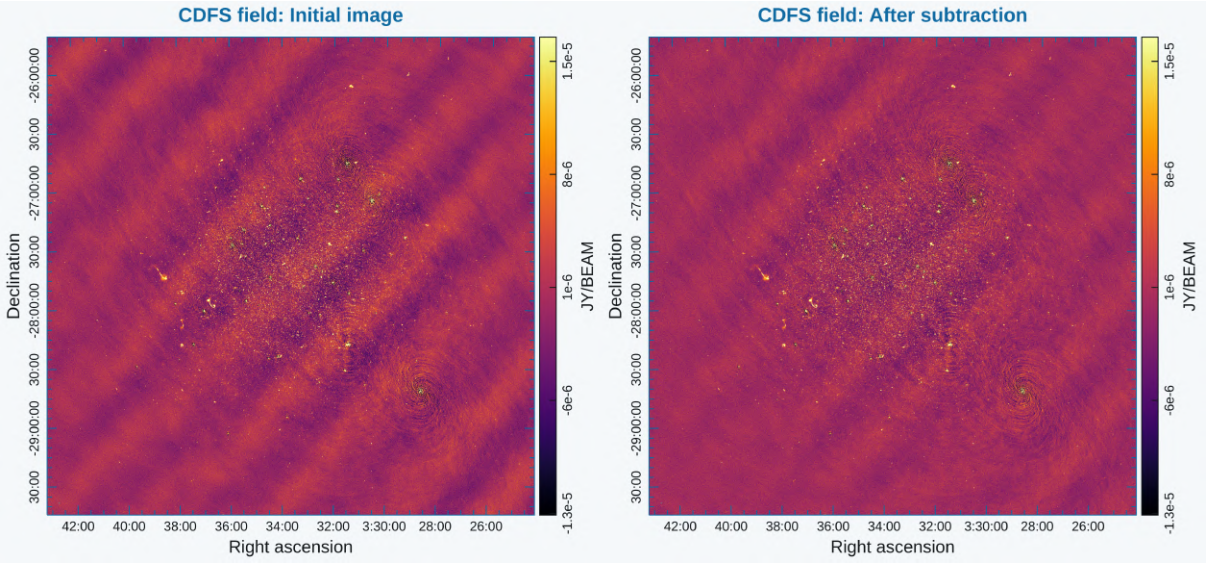


FIGURE 5.4: MIGHTEE images of the main field. Left: Initial image. Right: After simple subtraction with taql update.

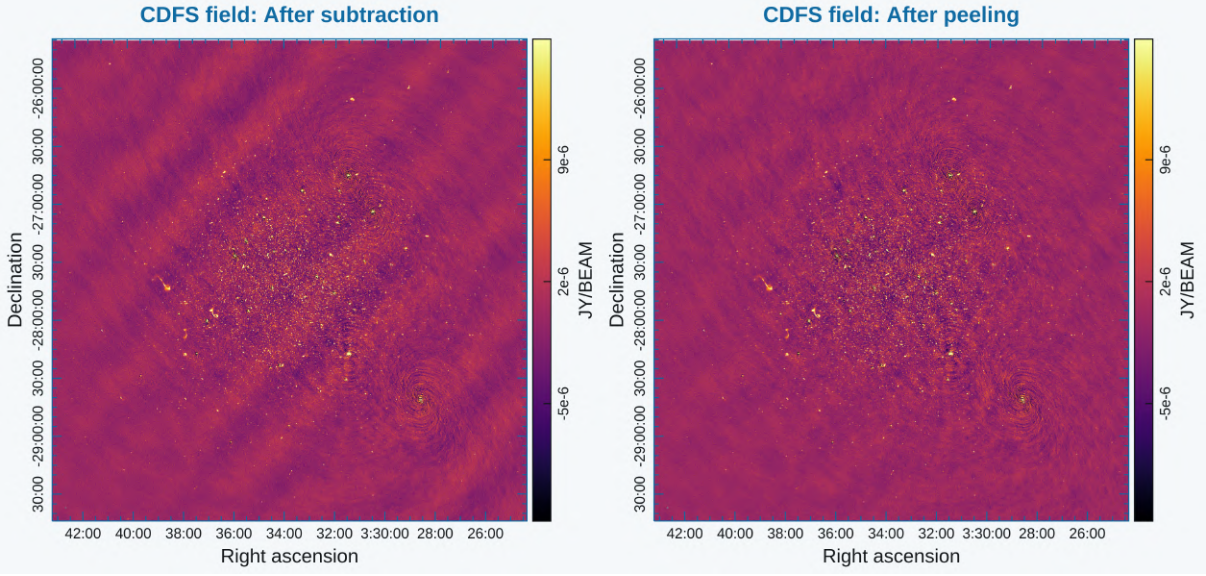
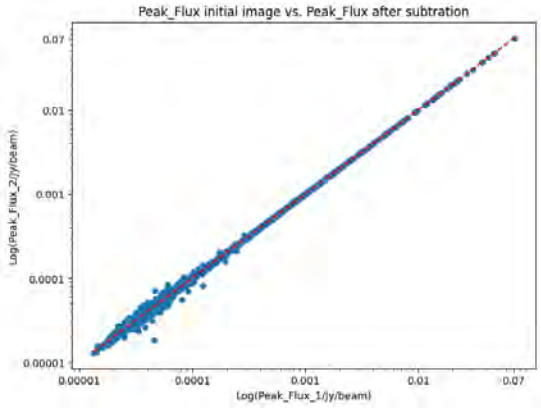
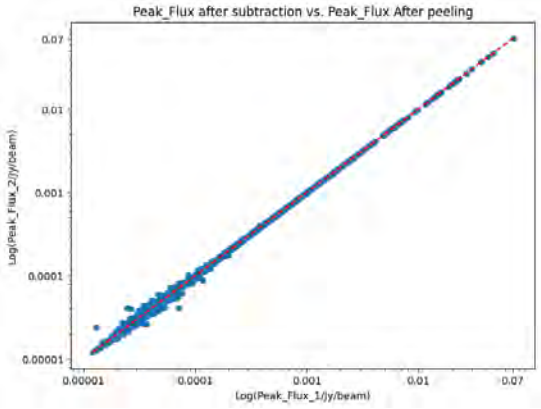


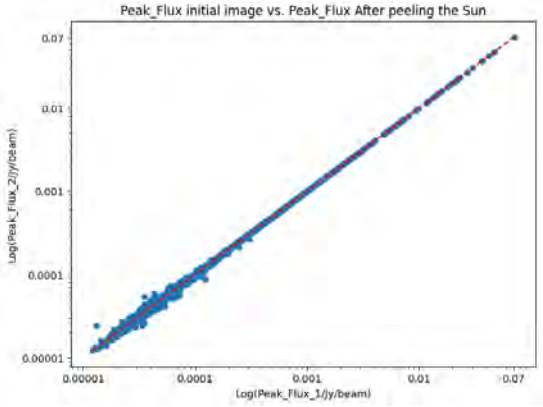
FIGURE 5.5: MIGHTEE images of the main field.Left: after subtracting the Sun. Right: after peeling the Sun with QuartiCal.



(a) Peak flux of the initial image vs. after subtraction.



(b) Peak flux after subtraction vs. after peeling.



(c) Initial peak flux vs. after peeling.

FIGURE 5.6: Peak flux plots. Peak flux plots with 1:1 line for reference.

5.3 LADUMA analysis

In this Section, we repeat the analysis conducted in Sections 5.1 and 5.2 for the LADUMA J033230-280757 target in the field. We analyse the effectiveness and impact of the pipeline in the data.

5.3.1 Visual inspection

For this MS, we also compare the visual images of the field before and after applying the solar interference mitigation pipeline. We used the pipeline with the same parameters for the deconvolution process before and after applying the pipeline. Figure 5.7 shows the initial image of the field (with the Sun) in the left panel and the image of the field after simply subtracting the Sun from the data. Figure 5.8 shows the simple subtracted and the image after peeling.

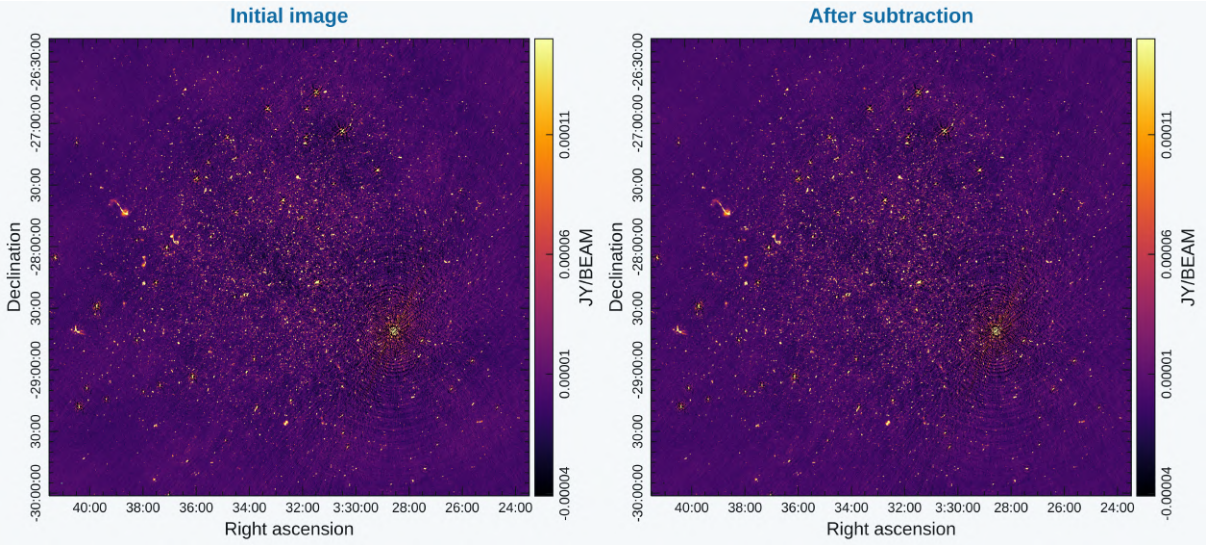


FIGURE 5.7: LADUMA field images of the main field. Left: Initial image. Right: After simple subtraction with taql update.

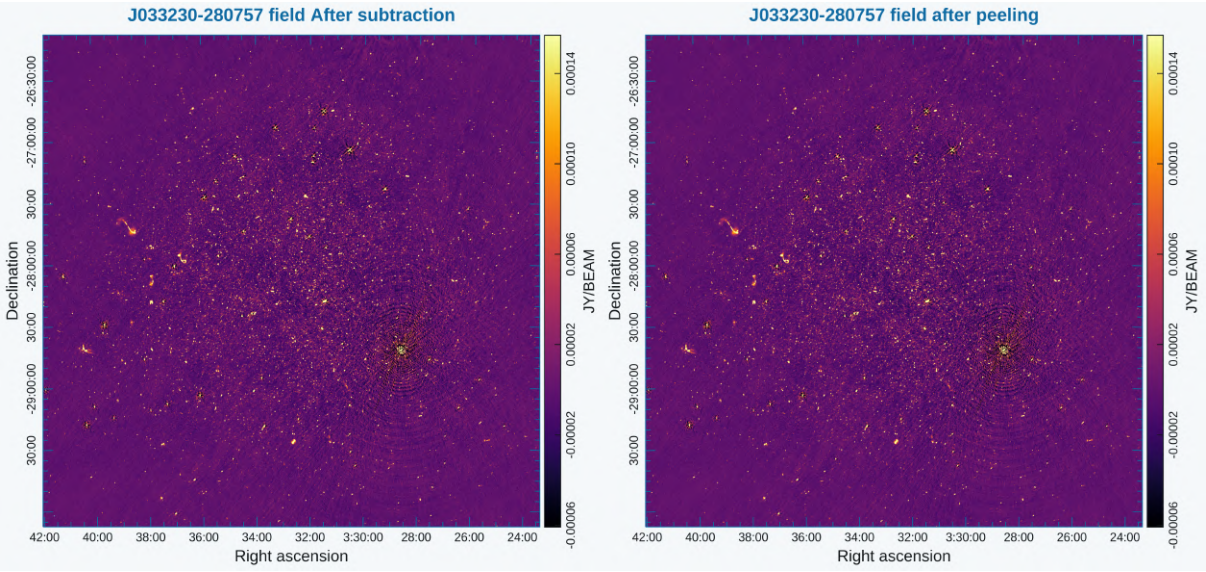


FIGURE 5.8: LADUMA field images of the main field. Left: After subtracting the Sun. Right: After peeling the Sun with QuartiCal.

Observation 3 showed less solar interference than our other observations conducted in the L-band, where the Sun was 8 degrees and 64 degrees from the main field. This is because the solar radio emission did not significantly affect this dataset. Several factors can justify this. First, the angular separation between the Sun and the telescope's pointing position significantly influences the level of solar interference in the field. When the Sun is closer to the telescope's pointing position, it is more likely to introduce noise in the FoV, causing interference. Another point could be the observed frequency band. In the case of the LADUMA dataset, the combination

of a higher angular separation (82 degrees) of the Sun from the FoV and the observed frequency band (UHF) likely resulted in very insignificant solar interference. The conditions that lead to the worst solar contamination on the data remain unclear. However, in this observation, a combination of angular separation, baseline projection, frequency and the state of the Sun, particularly in terms of its activity (in this case, minimal activity), could be what contributed to the low level of interference. While further investigation is required, these initial findings suggest that our dataset benefits from a unique combination of conditions that naturally leads to less contamination in the data.

SolarKAT incorporates two complementary techniques: simple subtraction and peeling. The peeling process was the icing on top of the cake for the SolarKAT pipeline. This is validated by the significant improvement that peeling contributed to in the images, as evident in the right panel of images 5.2, 5.5 and 5.8

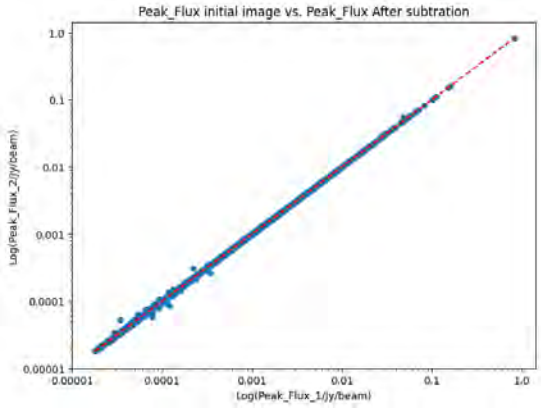
5.3.2 Peak flux analysis

After visually inspecting the visibilities, we performed peak fluxes analysis in each image as in Section 5.1.2 and 5.2.2. Similarly to the previous observations, the peak flux plots from the cross-matched catalogues for the J033230−280757 field are shown in Figure 5.9, *a*,*b* and *c*. Panel *a* shows the plot of the initial peak flux over the peak after simple subtraction over the entire FoV. Panel *b* shows a plot of the peak flux after subtraction and peeling, and panel *c* shows a plot of the initial peak flux over the peak flux after peeling. The max peak flux remains the same after applying the pipeline.

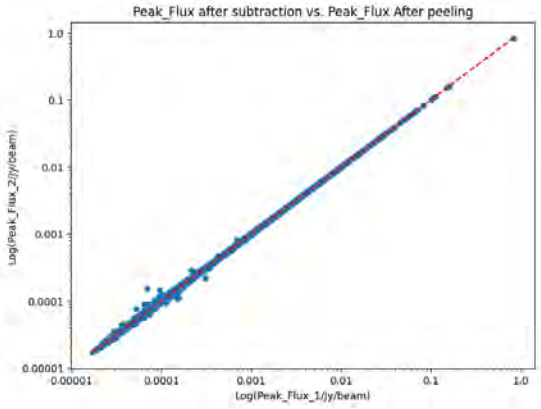
Like the previous two datasets, the LADUMA field peak flux values maintained consistency after applying SolarKAT with a max value of 0.84 Jy/beam.

The max peak flux remains the same after applying the pipeline.

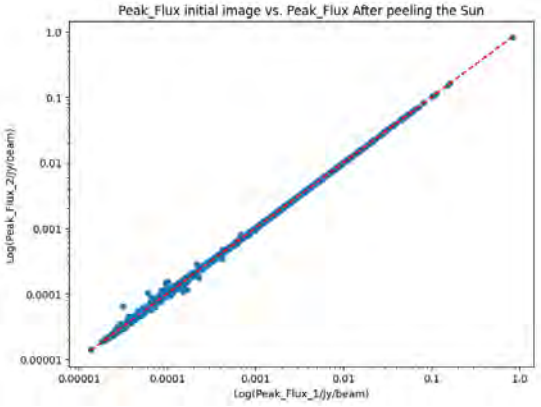
The consistency of the peak flux values is significant to ensure that the pipeline does not change the characteristics of the observed sources in the main field. This consistency shows that our pipeline effectively preserves the observed signal while removing unwanted solar interference in the data. The consistency of the flux measurements shows the pipeline's ability to preserve the data integrity, ensuring that the sources of interest remain unaffected by the mitigation process. However, this observation has noticeably low artefacts caused by the Sun. There is very similar scatter, which needs investigation to be understood.



(a) Peak flux of the initial image vs. after subtraction.



(b) Peak flux after subtraction vs. after peeling.



(c) Initial peak flux vs. after peeling.

FIGURE 5.9: Peak flux plots of the LADUMA data. Peak flux plots with 1:1 line for reference.

This section highlights the consistent nature of peak flux measurements after applying the mitigation pipeline and underscores its effectiveness in preserving the observed astronomical signal.

5.4 Statistical analysis

This Section presents the results of our solar interference mitigation process before subtraction, after subtraction and after peeling the Sun. The process was evaluated on three observations: ThunderKAT, MIGHTEE and LADUMA. Table 5.1 summarises each observation's total flux density and root mean square (RMS) values before and after subtraction and peeling the Sun. The pipeline implements subtraction and peeling to mitigate solar interference.

For the ThunderKAT data, the flux density decreased from 1.947 Jy before subtraction to 1.610 Jy after peeling. After applying the pipeline, the RMS slightly decreased from 4.210 $\times 10^{-5}$ Jy/beam to 4. 108 $\times 10^{-5}$ Jy/beam.

For the MIGHTEE data, the total flux density slightly decreased from 2.199 Jy before subtraction to 2.165 Jy after peeling. The RMS were relatively stable, slightly decreasing from 6.423 x 10^{-5} Jy/beam to 6.417 x 10^{-5} Jy/beam.

For the LADUMA data, the flux showed stability after subtraction from 8.735 Jy before subtraction to 8.736 Jy and decreased to 8.609 Jy after peeling. The RMS value in this observation remained at 7.119 Jy/beam after the mitigation process.

The statistical values reveal a consistent trend of interference reduction and improved data quality, with peeling demonstrating a more effective mitigation effect than a simple subtraction. These results show the SolarKAT pipeline's efficacy in enhancing the image's quality.

TABLE 5.1: Statistics values of the flux density and RMS of the full field for the ThunderKAT, MIGHTEE and LADUMA fields.

Project name	**Parameters**	**Before Subtraction**	**After Subtraction**	**After Peeling**
ThunderKAT	Flux Density (Jy)	1.947	1.632	1.610
	RMS ($10^{-5}\times$ Jy/beam)	4.210	4.135	4.108
MIGHTEE	Flux Density (Jy)	2.199	2.171	2.165
	RMS ($10^{-5}\times$ Jy/beam)	6.423	6.422	6.417
LADUMA	Flux Density (Jy)	8.735	8.736	8.609
	RMS ($10^{-4}\times$ Jy/beam)	7.119	7.119	7.119

The results show an overall decrease in flux density and RMS values for all observations. This suggests that SolarKAT effectively mitigated solar interference, achieving the pipeline goal.

For the ThunderKAT and MIGHTEE data, both flux and RMS values showed a decreasing trend, suggesting that the pipeline effectively removed the solar interference in the data. Additionally, the slight decrease in RMS for the MIGTHEE data suggests that the observation had a lower level of solar interference. Even though the image shows accentuated fringes in the data, the flux and RMS values show that this observation has a significantly lower level of interference than the ThunderKAT observation, which is quite curious.

In contrast, in the LADUMA field, the flux increased after subtraction, suggesting an ineffective subtraction process, but after the peeling process, a decrease in flux was observed. The RMS

value for this observation remained constant, indicating that the level of interference in this observation was negligible. This can be justified by the angular distance of the Sun from the telescope's pointing position.

The analysis of results suggests that both the subtraction and the peeling process significantly contribute to reducing solar interference in the data. However, peeling has shown to be more efficient. This can be because peeling is a complex process that includes subtraction and a calibration process that encounters the DDEs in the data. At the same time, a simple subtraction is likely to have residual noise in the data (as what happened for the LADUMA observation). This is observed in the statistical values in Table 5.1, where RMS and flux density values decrease. In conclusion, the peeling process is valuable to the effectiveness and robustness of the pipeline.

The SolarKAT pipeline has been demonstrated to be effective. For all three datasets, the results demonstrate a successful removal of solar interference using the SolarKAT pipeline. This was proved by the different evaluation metrics used to assess its effectiveness, including comparison in peak flux, RMS and flux density.

5.5 Comparative Study of Pixel Distributions in Solar Observations: Pre and Post SolarKAT Analysis

In addition to the image quality assessment, peak flux and statistical comparison of the flux density and RMS, we also fit a Gaussian model to the pixel brightness distribution of the full field as one of the evaluation metrics. This is a robust indicator of the pipeline's effectiveness in reducing artefacts/noise in the data.

The Gaussian model involves three crucial parameters. However, in our analysis, our primary focus was Gaussian standard deviation (σ) value. The sigma value tells us the variability or spread of the pixel values in our data. An increase in sigma implies a higher spread of pixel values and an increase in noise, while a decrease indicates effective noise reduction by the pipeline.

To achieve this goal, we extracted pixel values from the entire image of the main field for each dataset, discarding positive values and mirroring the opposing half. This approach is grounded in the concept that noise and artefacts, originating from far sidelobes, exhibit symmetry about zero. By discarding the positive half of the histogram, we avoid bias from actual sky emissions (solar

or otherwise), making the sigma value before and after SolarKAT implementation a reliable measure of improvement from the pipeline. The resulting σ should indicate the improved noise and artefact behaviour in a single value, making it a more robust indicator than just considering, for example, the minimum value of a map.

Figure 5.10, 5.11 and 5.12 show the pixel value distributions in solar images before (left) and after (right) SolarKAT implementation for our three different datasets. The blue lines represent the pixel plot distribution, and the red lines represent the fitted Gaussian models.

The overall results show the effectiveness of SolarKAT in reducing the negative pixel values across all observations. However, the reduction is relatively small for the MIGHTEE and LADUMA fields, which experienced lower interference than the GRS1747−312 field.

A consistent decrease in the amplitude, mean, and σ values was observed in all the cases. For example, in Figure 5.10, the standard deviation decreased from 21.4 µJy/beam to 19.4 µJy/beam after SolarKAT, indicating a reduction in the spread of the pixel values. Similar trends were observed in Figures 5.11 and 5.12, with values decreasing from 3.91 µJy/beam to 3.78 µJy/beam and from 8.00 µJy/beam to 7.88 µJy/beam respectively, emphasising the ability of SolarKAT to reduce noise in the data effectively.

In summary, these results align with the discussion on RMS in Section 5.4 and support the visual inspection of the data. The comprehensive evaluation of the results consistently demonstrates the SolarKAT pipeline's capability of mitigating solar radio interference and improving data quality while preserving the intrinsic characteristics of the sources of interest.

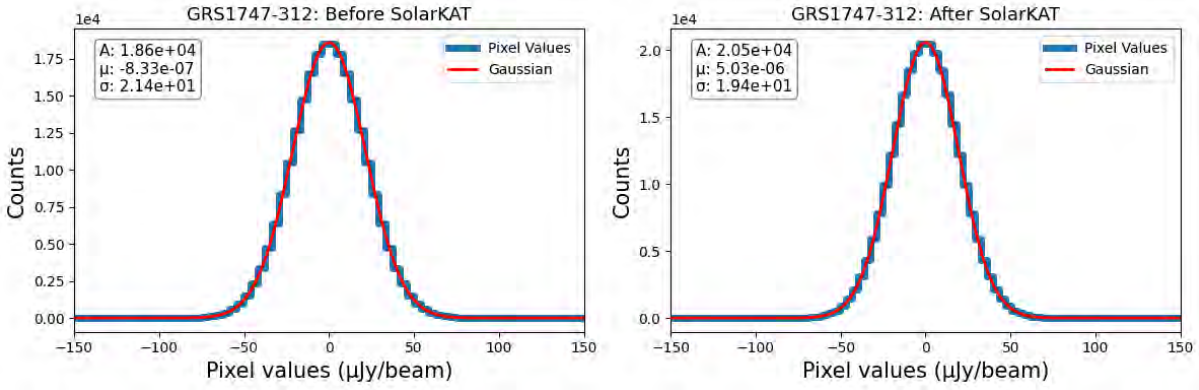


FIGURE 5.10: Pixel values distribution before and after SolarKAT in the ThunderKAT observation.

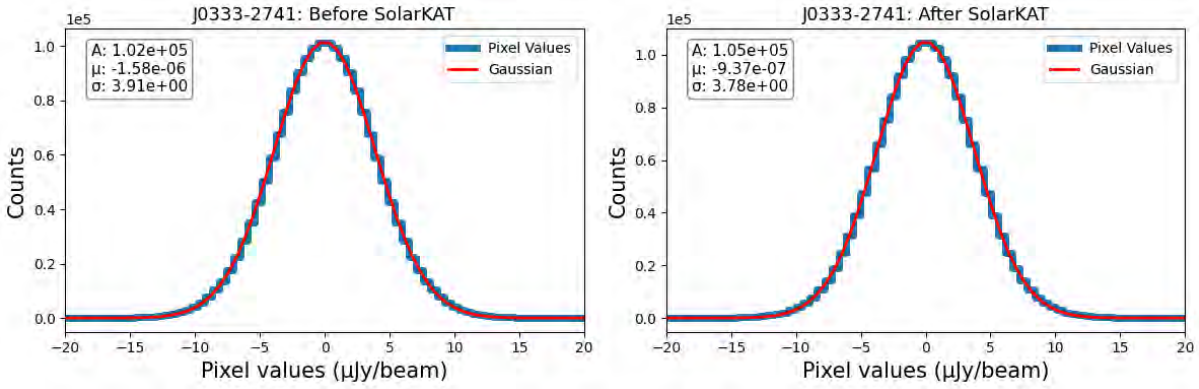


FIGURE 5.11: Pixel values distribution before and after SolarKAT in the MIGTHEE observation.

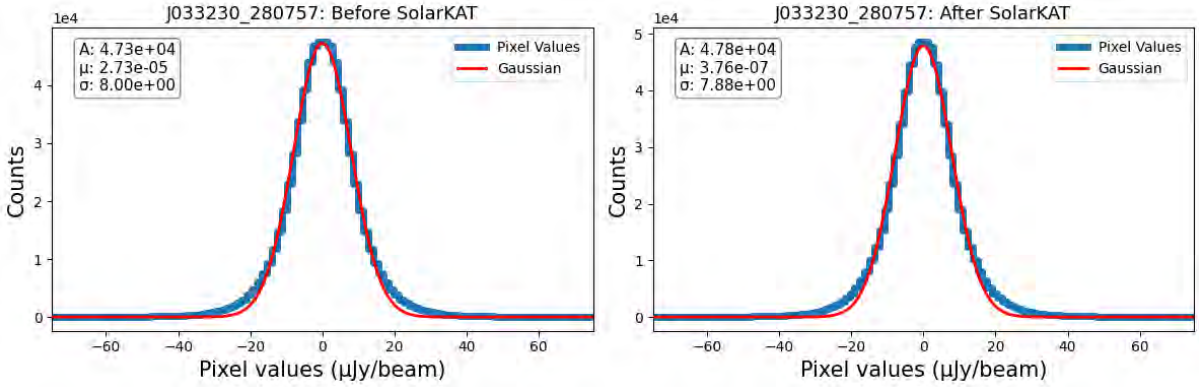


FIGURE 5.12: Pixel values distribution before and after SolarKAT in the LADUMA observation.

5.5.1 SolarKAT source revelling

The SolarKAT pipeline demonstrated the ability to reveal faint sources that were previously hidden by the presence of the Sun (see Sections 5.1.2, 5.2.2, 5.3.2). When creating the source catalogues (explained in point 2), the source count after applying the pipeline increased, which was observed as scattered points in the peak flux plots. Table 5.2 provides a comprehensive summary of the number of sources detected at various stages of the data processing pipeline and the total number of detected sources after the complete mitigation process for each dataset. This table compares the initial image (after selfcal), the image after subtraction, and the image after the peeling process, highlighting the pipeline's effectiveness in mitigating the Sun's interference.

Notably, an increase after subtraction, followed by a decrease after peeling, is observed, especially in the ThunderKAT dataset. This rise can be attributed to residual signals and noise

TABLE 5.2: Number of sources in the initial, after subtraction and after peeling images for the different datasets. The total of newly detected sources represents the total of new sources detected after the full SolarKAT mitigation process.

	ThunderKAT	**MIGHTEE**	**LADUMA**
Sources in the initial image	1285	8680	12380
Sources after subtraction	1379	9002	12501
Sources after peeling	1431	8628	12499
Total of new detected sources	146	52	119

amplification during subtraction, while the subsequent decrease results from improved modelling and reduced noise provided by the robust QuartiCal peeling process.

Further examination suggests that the scattered points likely represent sources previously hidden by the Sun. This is supported by analysing the improved images and statistical values, particularly the decreased RMS values after the pipeline application. These previously unrevealed sources are now revealed, showcasing the pipeline's positive impact on data quality and source travelling.

In practical terms, these scattered data points do not compromise the pipeline's primary objective: ensuring the integrity of astronomical observations, particularly the preservation of maximum peak values.

Our analysis indicates that the scattered data points in the lower part of the plots are a consequence of previously masked astronomical sources now unveiled by our mitigation pipeline. This discovery demonstrates the pipeline's effectiveness in enhancing data quality and expanding our knowledge of the observed field. This result can contribute to discovery in the field. An investigation can be conducted to assess whether these sources are new or known.

The SolarKAT pipeline's effectiveness in revealing previously hidden sources and enhancing data quality is a testament to its design and implementation. While the initial increase in sources due to the subtraction process's limitations is a factor to consider, the subsequent peeling process plays a crucial role in refining the model and reducing noise. This process leads to a more accurate representation of celestial sources in the data, ultimately contributing to the integrity of astronomical observations and expanding our knowledge of the Universe.

5.6 Solar images: A natural by-product of the SolarKAT Pipeline

The Sun's radio emission provides numerous unique diagnostics to estimate several physical parameters of the solar corona, which are inaccessible in other ways (Kansabanik et al., 2023) Imaging the Sun is a difficult task due to its highly dynamic solar emissions covering a wide range of wavelengths; for reasons very well explained in the paper from Kansabanik et al. (2023), MeerKAT is at the moment, possibly the best-suited telescope that can provide high-quality snapshot solar images. While Kansabanik et al. (2023) presents the first image of the Sun with MeerKAT data, SolarKAT implements an automated approach, ingesting any MeerKAT data, tracking and imaging the Sun independent of its position from the antenna pointing position. The approach we present to image the Sun differ from the approach used in the paper for the following reasons: i) the data used by them were observed to image the Sun, and the observation was made ensuring that the Sun is always 2.5 degrees away from the telescope position (means that they manually track the Sun by changing the telescope position every 15 minutes); meanwhile, SolarKAT can track the Sun by accessing data observed while the telescope where observing the same position over time with the Sun moving across the Sky.

In this Section, the solar images generated by the SolarKAT pipeline are used to ensure accuracy and consistency with reference solar images from the Solar Dynamics Observatory (SDO; Lemen et al. (2012)). We compared one scan from each observation; the other will be found in the appendices. Figures 5.13 (panel right), 5.14, and 5.15 show solar images as a natural by-product of SolarKAT compared to the images from the SDO. Figure 5.13 shows the image of the Sun in the single scan (scan 12) from the ThunderKAT L-band data, and the left panel is the SDO image. Figure 5.14 shows snapshot images (scans 3 to 16) of the Sun over time in the MIGTHEE L-band observation. The restoring beam sizes for all the solar images presented in this thesis are summarised in Appendix A.

Figure 5.15 shows the correspondent SDO images (on the left) for scans 3 and 9 of the solar images generated from the LADUMA dataset. As mentioned before, the diagonal lines and stripes in the image might be caused by another source of interference, such as a satellite. The solar disk is about 39.6 arcmin in the L-band and 35.4 in the UHF band.

We can observe a high level of agreement regarding solar features regarding sunspots and active regions, confirming the accuracy and reliability of the SolarKAT pipeline solar images. The

images show that the entire solar disk, including details of active regions, is visible. The active regions present in the SDO images are also detected in the MeerKAT images.

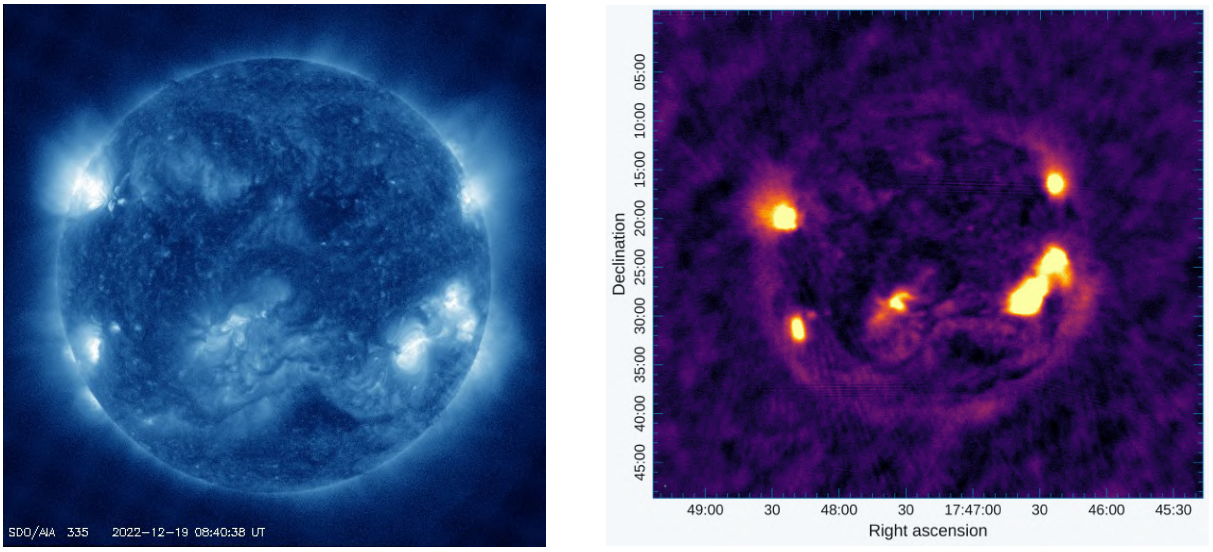


FIGURE 5.13: Comparison between the SDO image and the MeerKAT data of the ThunderKAT data. Left panel: SDO image. Right panel: Observed ThunderKAT solar radio image. In both images, we can see the correspondence in the present active regions.

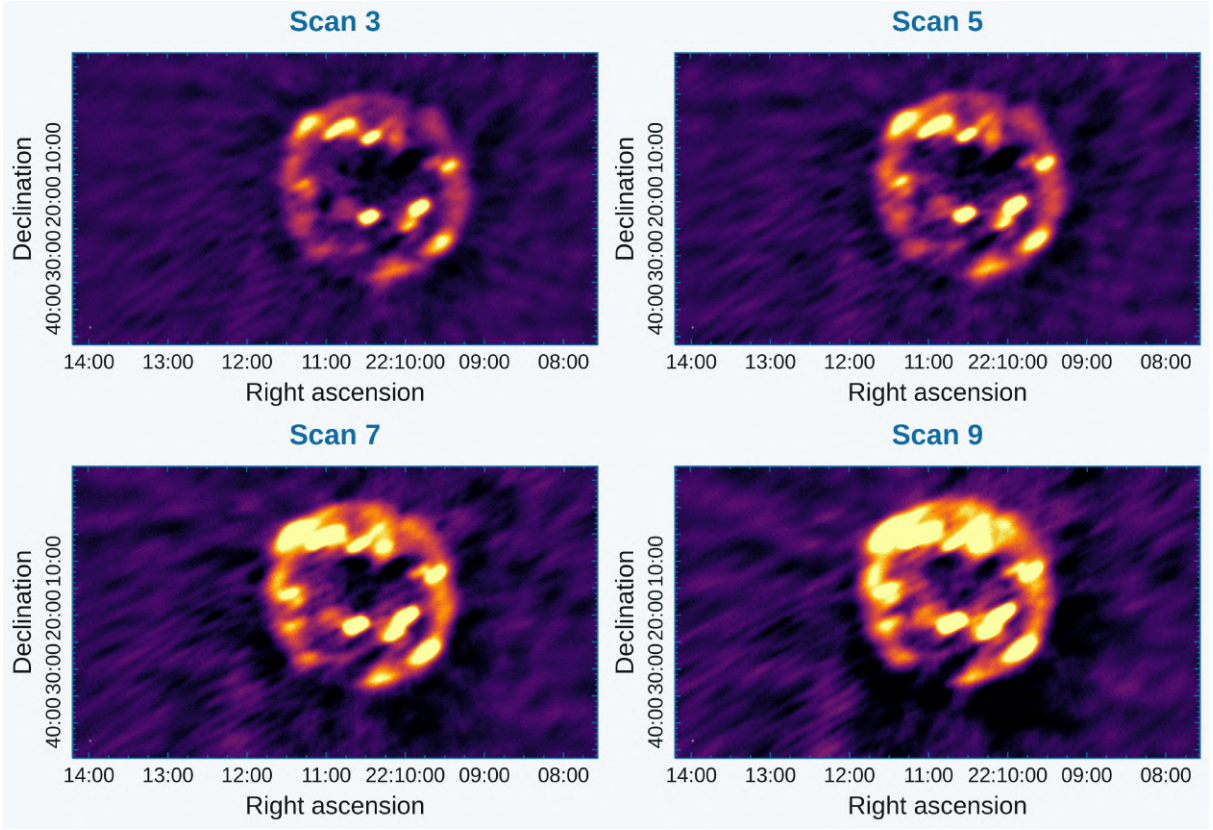


FIGURE 5.14: L-band images of different scans across the MIGHTEE observation. The Sun was located 76 degrees from the main field. Scans were recorded from 12:14 PM to 2:24 PM.

It is worth noting that during this research, a significant milestone was achieved by successfully imaging the Sun applying SolarKAT to the MeerKAT data. The data was observed in different conditions, including the angular distance of the Sun from the telescope pointing position and frequency band. The recent publication by Kansabanik et al. (2023) highlights the groundbreaking work of another research group that has independently achieved the same milestone of imaging the Sun. This publication demonstrates the collective efforts and advancements in solar interferometry using MeerKAT.

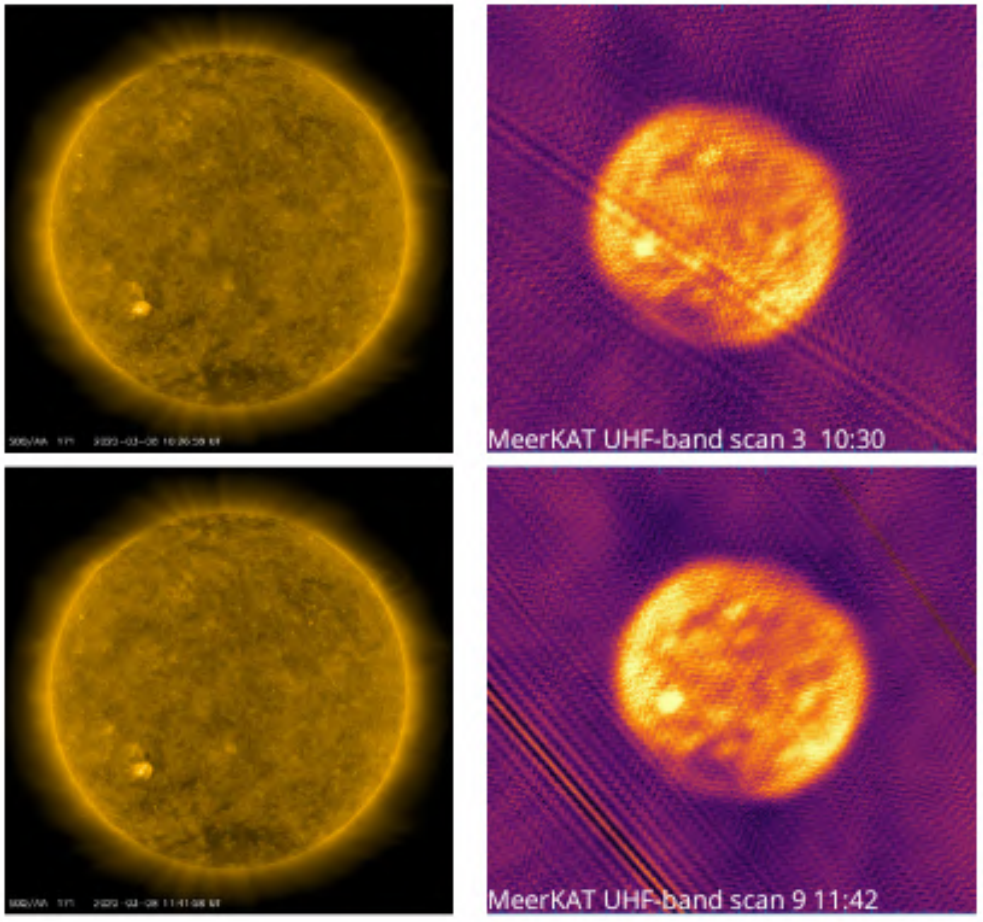


FIGURE 5.15: Comparison between the SDO image and the MeerKAT data of the LADUMA field. Left panels: SDO image. Right panel: Observed LADUMA solar radio image. In both images, we can see the correspondence in the present active regions.

However, what sets this project apart is the development of a dedicated pipeline specifically tailored for MeerKAT observations. The SolarKAT pipeline incorporates advanced techniques such as self-calibration, sun tracking, imaging, region-based analysis, and deconvolution to isolate and remove solar interference signals from observations.

5.7 SolarKAT processing time

We evaluate the processing time by measuring how long the pipeline processes a specific dataset. It is important to note that the processing time will depend on several factors, including size, duration of the observation, calibration issues and number of scans in the MS. These numbers can increase or decrease depending on the imaging and calibration cycles, varying from dataset to dataset. The time taken in the imaging cycle can also vary depending on the image size, the number of iterations, and the use of multiscale parameters in the Sun's deconvolution process. In general, the tasks that demand time are imaging, calibration, model prediction, deconvolution of the Sun, and peeling. As a reference, we used the 15-minute observation of the ThunderKAT survey to know how long it is needed to perform the mitigation process. The SolarKAT pipeline takes approximately 20 hours (0 hours corresponding to the calibration cycle) to process a 15-minute (single scan) observation with 86 GB in size. Table 5.3 provides information about the project name, dataset size, number of scans, and processing time for each project.

TABLE 5.3: Dataset information and processing time for each observation. The table shows the SolarKAT processing time plus the time taken only for the calibration cycle in the pipeline (in parenthesis).

	ThunderKAT	**MIGHTEE**	**LADUMA**
Number of scans	1	6	23
Size of Dataset	86 GB	820 GB	1.5 TB
Scans duration	15 min	30 min	20 min
Processing time (hours)	10 (+ 10 h calibration)	26 (+ 8 h calibration)	16 (+ 12 h calibration)

For the ThunderKAT dataset, the image size in WSClean also influenced the as we used an image size equal to 14100 to include SgrA in the image. The image size used for MIGHTEE and LADUMA is 10000×10000. Bigger image sizes make the processing time longer than smaller image sizes. Another aspect that can influence processing time is using software such as CubiCal or QuartiCal. Using QuartiCal for calibration and *peeling* significantly reduces the pipeline processing time. Depending on the parameters specifications (image size, number

of iterations, number of calibration and imaging cycles), the data size and specifications, the processing time can be lower. In summary, now that we have demonstrated that the workflow is viable, there's a lot of room to speed things up through optimisation in terms of QuartiCal settings and parallelisation of the process for observations that feature multiple scans.

In addition to what we mentioned previously, machine specifications can significantly affect processing time. For example, if a computer has a lot of memory (RAM) or multiple (CPUs), it can read more data and make processes such as calibration (2GC) or peeling (3GC) faster. It can handle more data at once, speeding up the process.

Chapter 6

Conclusion and Future work

In this study, our primary objective was to address the challenge of solar RFI in radio interferometric observations conducted by the MeerKAT telescope. The goal was to remove the Sun from the data, enhance the data quality, and ensure the data's reliability and integrity by developing and applying the SolarKAT pipeline. More importantly, we salvage rather than discard the affected data.

SolarKAT is shown to be capable of revealing faint sources hidden by the Sun.

The SolarKAT pipeline has proven to be an effective and robust solution for solar interference mitigation in MeerKAT observations. Through detailed statistical analysis, we observed a significant improvement in the image quality of the data, showing precise and more reliable data after the pipeline application.

Our analysis of peak fluxes revealed consistent values after applying the SolarKAT pipeline across multiple datasets. Plots illustrating the peak flux before and after mitigation consistently showed stability, indicating the pipeline's effectiveness in maintaining the integrity of peak flux measurements and demonstrating that the process does not bias the measurements.

Statistical analyses of flux density and RMS values further supported the success of the SolarKAT pipeline. The data consistently demonstrated not only the accuracy of the pipeline in mitigating interference but also its efficiency in maintaining the essential statistical properties of the observations.

The decrease in sigma value, observed in the fitted Gaussian model to the pixel values, highlights SolarKAT's effectiveness in reducing the effective image noise in a MeerKAT observation.

SolarKAT is also a valuable tool for producing detailed images of the Sun, which can help answer several solar physics questions and potentially be used for space weather forecasts if implemented as a standard component of an automated data processing pipeline. The automation of the pipeline allows it to handle large amounts of data and reduces the processing time that a manual approach can demand.

The peeling process has shown to be a powerful technique even in cases where the source of interest is out of the image's FoV, ensuring the effectiveness of the pipeline. Integrating the peeling technique into the SolarKAT pipeline further enhances its robustness. This addition automates interference mitigation for large datasets and increases data processing and analysis efficiency.

SolarKAT is built on new radio astronomy software technologies (QuartiCal, Stimela, Breizorro) and is a powerful demonstration of their capabilities. The SolarKAT pipeline is a testament to the advancements in data processing techniques, ensuring the quality and accuracy of observations from radio telescopes. In conclusion, our work demonstrates the significance of robust solar interference mitigation techniques in radio astronomy. By overcoming the challenges posed by solar interference, the SolarKAT pipeline improves the reliability and quality of radio interferometric imaging observations, opening new possibilities for advancements in our understanding of the observations.

6.1 Future Work and Collaborative Opportunities

As we conclude our study, there are several avenues for future work and opportunities for collaboration. One significant aspect is the packaging of the developed pipeline for public release.

We intend to make the tools and methodologies accessible to the broader scientific community. This involves creating user-friendly documentation (ongoing work), organising the codebase, and ensuring that researchers in the field can easily install and utilise the pipeline.

Moreover, our pipeline has the potential for broader applications beyond the specific radio telescope used in this study. In particular, we plan to explore the adaptability of the pipeline to other radio telescopes, such as ASKAP. This adaptability enhances our work's versatility and promotes collaboration with researchers using different instruments.

An essential consideration for the future development of the SolarKAT pipeline revolves around the speed-up of its processing time. This is particularly vital to enable the generation of daily solar maps in the L and UHF bands (and perhaps in S-band in future). To meet this demand, strategic enhancements and considerations will be implemented to ensure the pipeline's efficiency aligns with the evolving requirements of routine solar observations. This can be met by implementing optimization strategies such as parallel processing or algorithm enhancements (which can be easily implemented).

We aim to contribute to the collective progress in radio astronomy data analysis by sharing our pipeline and facilitating its integration with other telescopes. Collaborative efforts in this sense promise to advance our understanding of the Universe and address new scientific challenges in radio astronomy.

Appendix A

Solar Images over time

This section presents a visual representation of the solar disk over time for the MIGHTEE and LADUMA observation. We do not include the ThunderKAT image of the Sun because it only has one scan in the observed Measurement Set. These images offer a unique perspective of the changes and patterns of solar activity over time.

The observations were conducted using the MeerKAT telescope, known for its large baseline and sensitivity. This sensitivity, combined with the telescope's movement relative to the Earth's rotation, can lead to the smearing of the Sun's image in the radio images. The smearing effect is particularly noticeable in observations made during or close to sunrise or Sunset (as we can see in the right panel of Figure A.2 and left panel of Figure A.6) scan 15 when the Sun's movement across the sky is more pronounced.

The Earth's rotation changes the Sun's position; this can lead to the spreading of the Sun's image over the telescope FoV. The movement of the telescope relative to the Earth's rotation can also cause the smearing effect. Additionally, the large MeerKAT baseline makes it more sensitive to the Sun's position in the sky, which can also lead to more pronounced smearing. Other factors, such as atmospheric scattering and observation timing, can lead to smearing. Atmospheric scathing can distort the Sun's image, making it appear smeared or distorted in radio observations. Observations conducted during sunrise or Sunset, when the Sun is on the horizon, can also lead to smearing. The smearing of the Sun in radio observations, mainly when observed with large antenna arrays such as MeerKAT, is not an actual emission from the Sun. Instead, it results from how the telescope receives and processes the radio waves. Therefore, SolarKAT can effectively deal with such cases. By generating regions followed by masks in the

region of the true emission of the Sun (where the solar dish is centred), then subtracting it from the visibilities, we subtract the Sun. However, because of the effects such as smearing, after subtraction, some residual emissions remain in the data, which can be addressed by applying the peeling process. The subtraction process is essential to improve the visibilities model before peeling. Otherwise, the model would be incomplete.

Tables A.1 and A.2 present the values of the restoring beam sizes (Beam Major Axis (BMAJ), Beam Minor axis (BMIN), and Beam Position Angle (BPA) of the beam for the solar images produced from MIGHTEE and LADUMA datasets.

A.1 The MIGTHEE solar images per scan

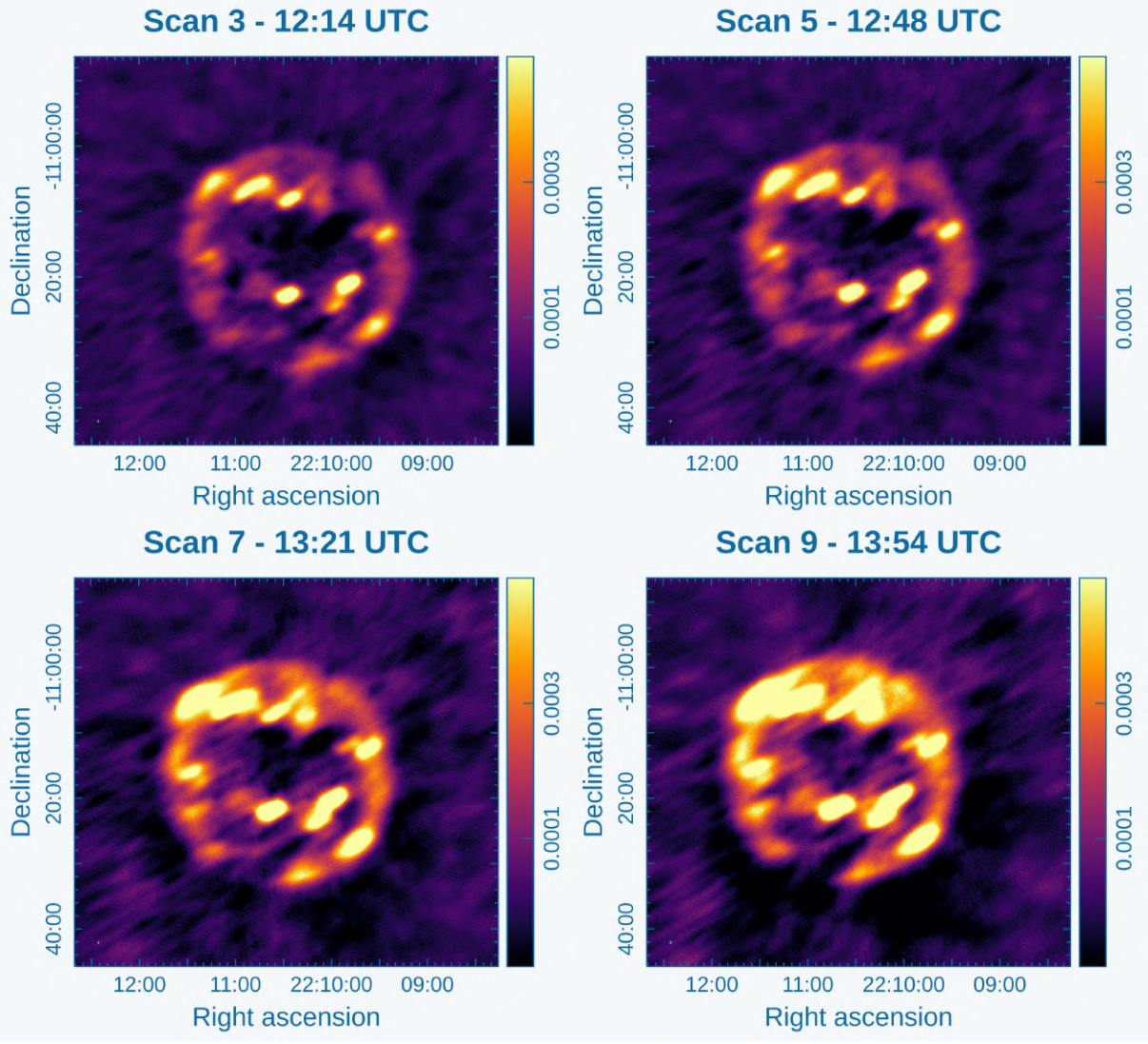


FIGURE A.1: Sun images of scans 3, 5, 7 and 9 in the L-band.

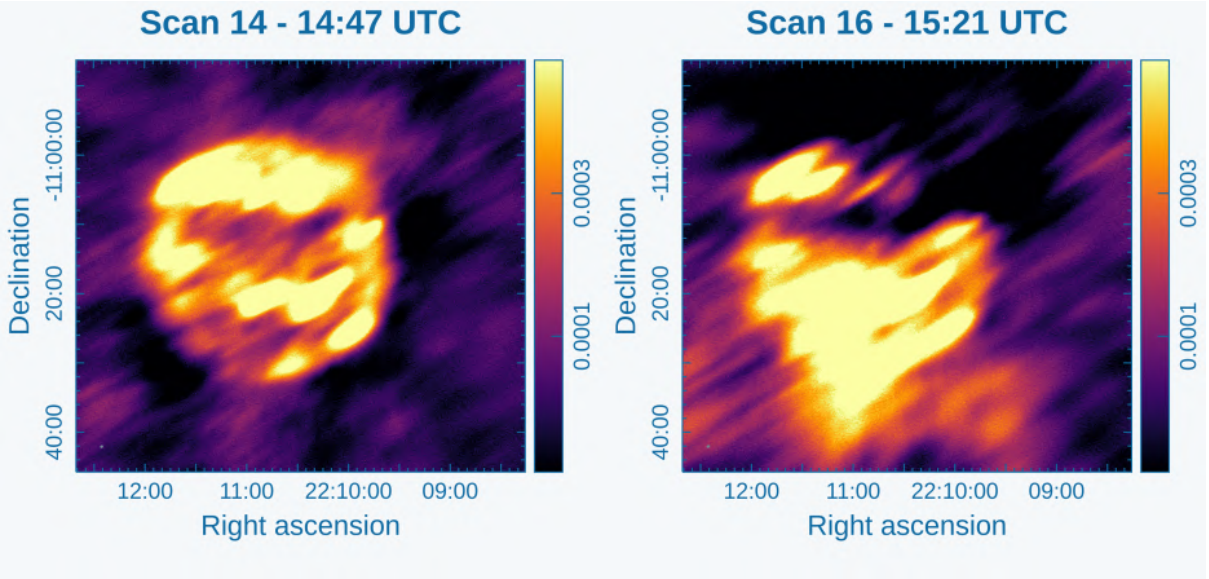


FIGURE A.2: Sun images of scans 14 and 16 in the L-band.

A.2 The LADUMA solar images per scan

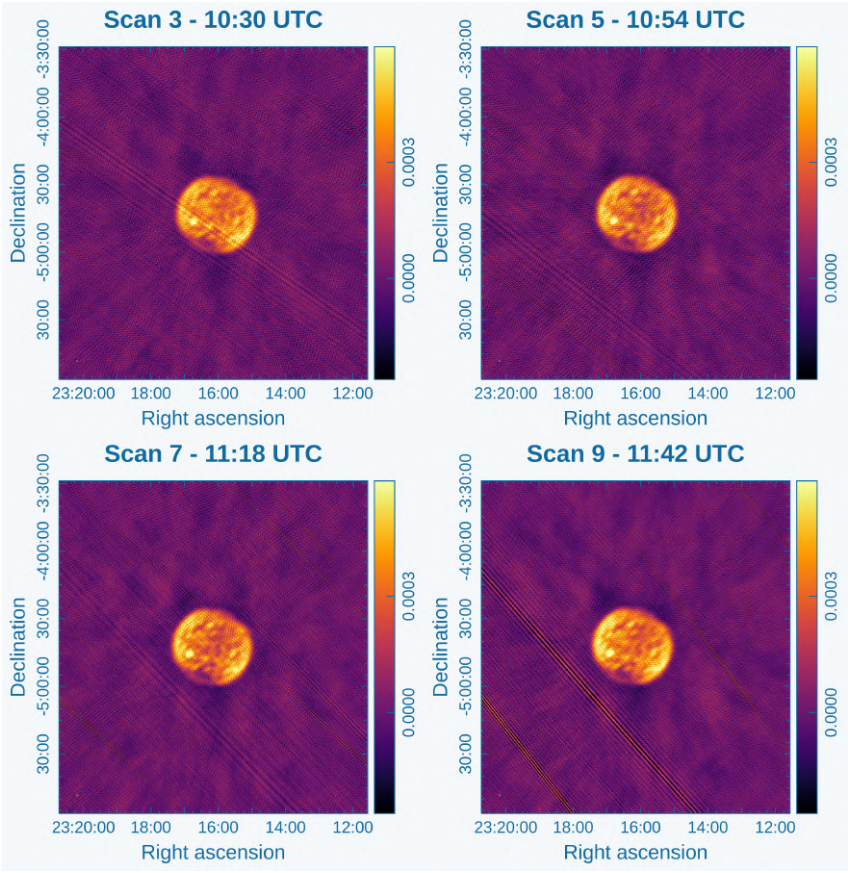


FIGURE A.3: Sun images of scans 5, 13, 15 and 17 in the UHF-band. The central frequency in this observation is 815 MHz.

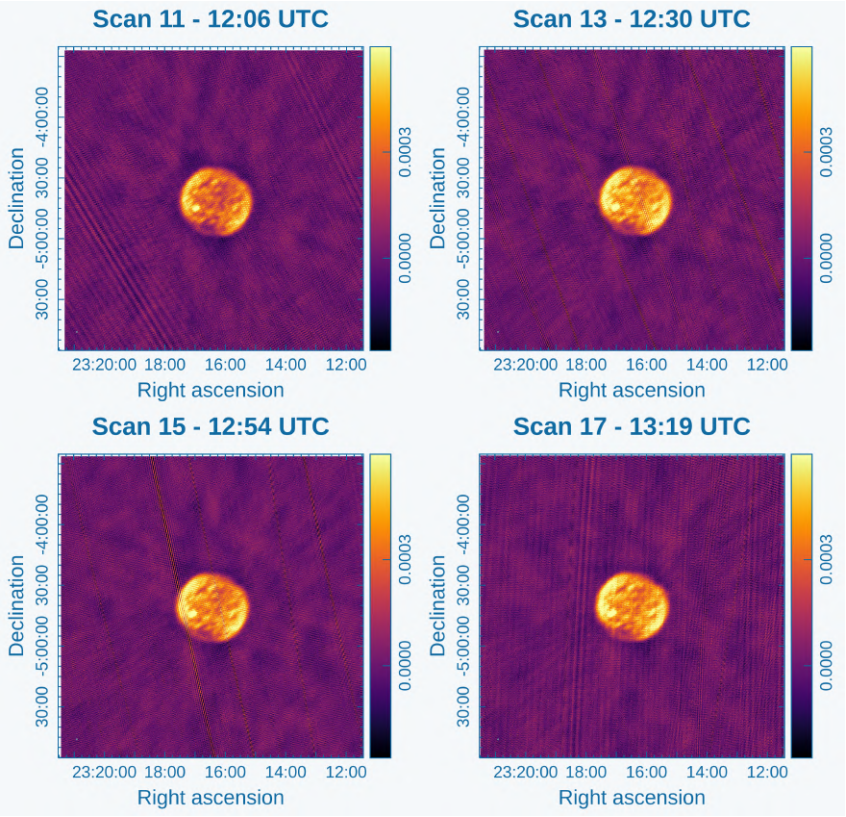


FIGURE A.4: Sun images of scans 19, 21, 25 and 27 of the UHF-band.

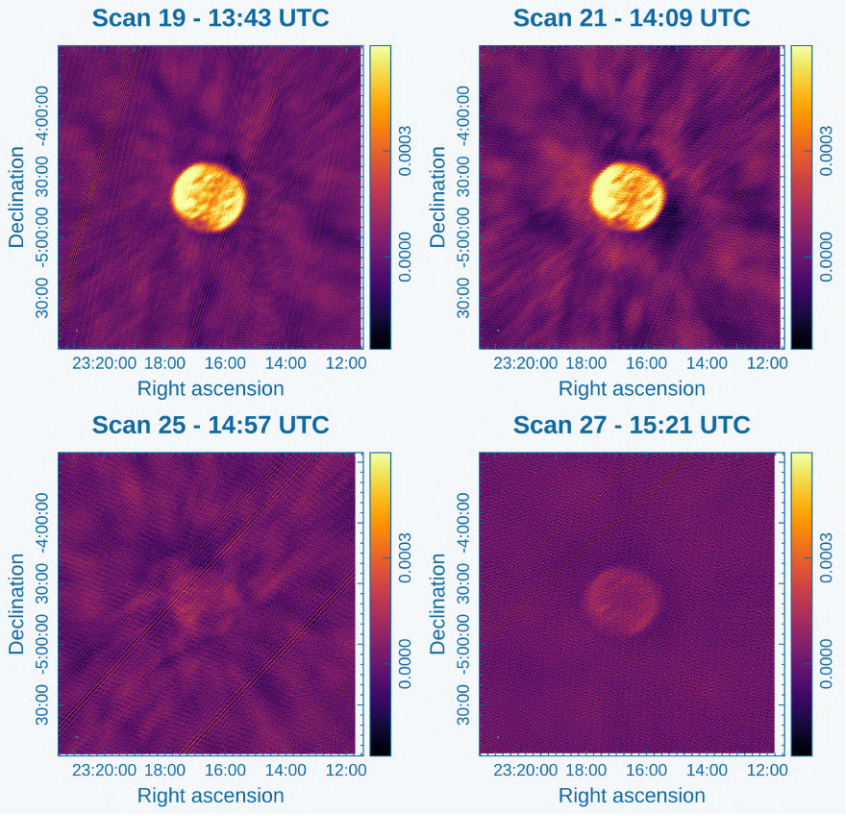


FIGURE A.5: Scans 29, 31, 33 and 35 in the UHF-band.

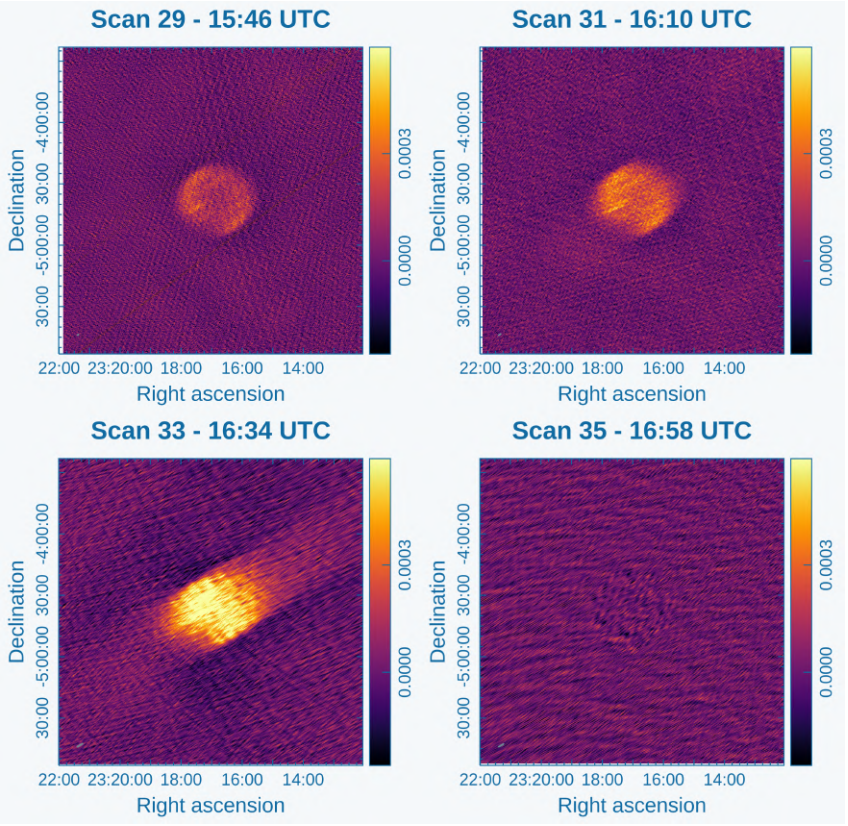


FIGURE A.6: Scans 29, 31, 33 and 35 in the UHF-band. The smearing of the Sun in scan 33 is due to Sunset.

TABLE A.1: Values of BMAJ, BMIN, and BPA for each scan of the MIGHTEE observation.

Scans	BMAJ (arcseconds)	BMIN (arcseconds)	BPA (degrees)
scan 3	8.14	6.25	162.2
scan 5	8.19	6.77	157.89
scan 7	8.27	7.83	128.2
scan 9	9.42	7.66	124.42
scan 14	12.36	7.66	119.36
scan 16	16.04	7.69	118.11

TABLE A.2: Values of BMAJ, BMIN, and BPA for each scan of the LADUMA observation.

Scans	BMAJ (arcseconds)	BMIN (arcseconds)	BPA (degrees)
scan 3	15.96	10.16	160.03
scan 5	16.03	10.22	160.47
scan 7	16.07	10.42	160.32
scan 9	16.0	10.85	159.39
scan 11	15.71	11.67	156.38
scan 13	15.82	12.41	153.11
scan 15	16.71	12.56	154.03
scan 17	17.81	12.38	149.78
scan 19	19.19	12.56	144.65
scan 21	21.21	12.81	139.25
scan 25	28.17	13.68	129.61
scan 27	34.63	13.68	125.99
scan 29	47.12	13.51	123.59
scan 31	77.02	13.66	121.82
scan 33	185.28	21.78	120.78
scan 35	173.12	20.67	119.77

Appendix B

Pipeline Code and Documentation

The SolarKAT pipeline code and documentation can be found at the following links:

- Git repository: https://github.com/ratt-ru/solarkat
- SolarKAT documentation: https://solarkat-docs.readthedocs.io/en/latest/index.html
- ASCL Code Record: https://ascl.net/2401.013

Bibliography

Almeida L. B., 1997, Product and convolution theorems for the fractional fourier transform, IEEE Signal Processing Letters, 4, 15

Andati L. A., 2020, PhD thesis, Rhodes University

Bean B., et al., 2022, Casa, common astronomy software applications for radio astronomy, Publications of the Astronomical Society of the Pacific, 134, 114501

Blyth S.-L., 2015, in SALT Science Conference 2015 (SSC2015). p. 36

Briggs D. S., Schwab F. R., Sramek R. A., 1999, in Synthesis Imaging in Radio Astronomy II. p. 127

Burke B. F., Graham-Smith F., Wilkinson P. N., 2019, An introduction to radio astronomy. Cambridge University Press

Clark B., 1980, An efficient implementation of the algorithm'clean', Astronomy and Astrophysics, 89, 377

Condon J. J., Ransom S. M., 2016, Essential radio astronomy. Vol. 2, Princeton University Press

Cornwell T. J., 2008, Multiscale clean deconvolution of radio synthesis images, IEEE Journal of selected topics in signal processing, 2, 793

Cornwell T., Wilkinson P., 1981, A new method for making maps with unstable radio interferometers, Monthly Notices of the Royal Astronomical Society, 196, 1067

Fender R., et al., 2017, Thunderkat: the meerkat large survey project for image-plane radio transients, arXiv preprint arXiv:1711.04132

Fomalont E. B., Perley R. A., 1999, in Synthesis imaging in radio astronomy II. p. 79

Ford J., Buch K., 2014. pp 231–234, doi:10.1109/IGARSS.2014.6946399

Hamaker J., Bregman J., Sault R., 1996, Understanding radio polarimetry. i. mathematical foundations, Astronomy and Astrophysics Supplement Series, 117, 137

Hey J. S., Hey J., 1973, The evolution of radio astronomy. Vol. 1, Science History Publications New York

Heywood I., 2020, Oxkat: Semi-automated imaging of meerkat observations, Astrophysics Source Code Library, pp ascl–2009

Högbom J., 1974, Aperture synthesis with a non-regular distribution of interferometer baselines, Astronomy and Astrophysics Supplement, Vol. 15, p. 417, 15, 417

Ishak B., 2019, An introduction to radio astronomy, 4th edition, Contemporary Physics, 60, 335

Jarvis M. J., et al., 2017, The meerkat international ghz tiered extragalactic exploration (mightee) survey, arXiv preprint arXiv:1709.01901

Jonas J., MeerKAT Team 2016, in MeerKAT Science: On the Pathway to the SKA. p. 1

Kansabanik D., Mondal S., Oberoi D., Biswas A., Bhunia S., 2022, Robust absolute solar flux density calibration for the murchison widefield array, The Astrophysical Journal, 927, 17

Kansabanik D., et al., 2023, First image of the sun with meerkat solar observations: Opening a new frontier in solar physics, arXiv preprint arXiv:2307.01895

Kenyon J., Smirnov O., Grobler T., Perkins S., 2018, Cubical–fast radio interferometric calibration suite exploiting complex optimisation, Monthly Notices of the Royal Astronomical Society, 478, 2399

Kenyon J. S., Perkins S., Smirnov O., 2023, Quartical: Fast radio interferometric calibration, Astrophysics Source Code Library, pp ascl–2305

Lemen J. R., et al., 2012, The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo), Solar Physics, 275, 17

Lonsdale C. J., et al., 2009, The murchison widefield array: Design overview, Proceedings of the IEEE, 97, 1497

Makhathini S., 2017, Advanced radio interferometric simulation and data reduction techniques,] 10.3847/1538-4357/ac4bba, 927, 119

Makhathini S., 2018, Advanced radio interferometric simulation and data reduction techniques, Rhodes University

Mesarcik M., Boonstra A.-J., Ranguelova E., van Nieuwpoort R. V., 2022, Learning to detect radio frequency interference in radio astronomy without seeing it, Monthly Notices of the Royal Astronomical Society, 516, 5367

Mohan N., Rafferty D., 2015, Pybdsf: Python blob detection and source finder, Astrophysics Source Code Library, pp ascl–1502

Noordam J., 2004, Lofar calibration challenges, Proc SPIE

Noordam J. E., Smirnov O. M., 2010, The meqtrees software system and its use for third-generation calibration of radio interferometers, Astronomy & Astrophysics, 524, A61

Offringa A. R., Smirnov O., 2017, An optimised algorithm for multiscale wideband deconvolution of radio astronomical images, MNRAS, 471, 301

Offringa A., et al., 2014a, Wsclean: an implementation of a fast, generic wide-field imager for radio astronomy, Monthly Notices of the Royal Astronomical Society, 444, 606

Offringa A. R., McKinley B., Hurley-Walker et al., 2014b, WSClean: an implementation of a fast, generic wide-field imager for radio astronomy, MNRAS, 444, 606

Pawsey J., 1946, Observation of million-degree thermal radiation from the sun at a wavelength of 1.5 metres, Nature, 158, 633

Ramaila A. J., Smirnov O., Heywood I., 2023, Breizorro: Image masking tool, Astrophysics Source Code Library, pp ascl–2305

Samboco V. D. G. G., 2023, in Proceedings of the 29th Young Scientists' Conference on Astronomy and Space Physics. YSC29, https://indico.cern.ch/event/1279127/contributions/5373911/

Sault R., Noordam J., 1995, Eliminating solar interference in radio interferometric spectrometry., Astronomy and Astrophysics Suppl., Vol. 109, p. 593-595 (1995), 109, 593

Sihlangu I., 2019, Master's thesis, Faculty of Science

Sirothia S., 2017, PhD thesis, RHODES UNIVERSITY

Smirnov O. M., 2011, Revisiting the radio interferometer measurement equation-i. a full-sky jones formalism, Astronomy & Astrophysics, 527, A106

Smirnov and Makhathini, inprep

Starck J.-L., Pantin E., Murtagh F., 2002, Deconvolution in astronomy: A review, Publications of the Astronomical Society of the Pacific, 114, 1051

Tasse C., 2023, killms: Direction-dependent radio interferometric calibration package, Astrophysics Source Code Library, pp ascl–2305

Tasse C., et al., 2023, Ddfacet: Facet-based radio imaging package, Astrophysics Source Code Library, pp ascl–2305

Thompson A. R., 1999, in Synthesis Imaging in Radio Astronomy II. p. 11

Thompson A. R., 2017, Interferometry and Synthesis in Radio Astronomy by A. Richard Thompson, James M. Moran, George W. Swenson Jr., 3rd ed. 2017. edn. Astronomy and Astrophysics Library, Springer Nature, Cham

Thompson A. R., Moran J. M., Swenson Jr G. W., Thompson A. R., Moran J. M., Swenson G. W., 2017, Van cittert–zernike theorem, spatial coherence, and scattering, Interferometry and synthesis in radio astronomy, pp 767–786

Van Haarlem M. P., et al., 2013, Lofar: The low-frequency array, Astronomy & astrophysics, 556, A2