

RHODES UNIVERSITY

Link between Ghost Artefacts, Source
Suppression and Incomplete Calibration
Sky Models

*Author: Nunhokee
Chuneeta Devi*

*Supervisors: Dr. T.L. Grobler &
Prof. O.M. Smirnov*

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

in the

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

March 2015

Abstract

Calibration is a fundamental step towards producing radio interferometric images. However, naive calibration produces calibration artefacts, in the guise of spurious emission, buried in the thermal noise. This work investigates these calibration artefacts, henceforth referred to as “ghosts”. A 21 cm observation with the Westerbork Synthesis Radio Telescope yielded similar ghost sources, and it was anticipated that they were due to calibrating with incomplete sky models. An analytical ghost distribution of a two-source scenario is derived to substantiate this theory and to seek answers to the related bewildering features (regular ghost pattern, points spread function-like sidelobes, independent of model flux). The theoretically predicted ghost distribution qualitatively matches with the observational ones and shows high dependence on the array geometry. The theory draws the conclusion that both the ghost phenomenon and suppression of the unmodelled flux have the same root cause. In addition, the suppression of the unmodelled flux is studied as functions of unmodelled flux, differential gain solution interval and the number of sources subjected to direction-dependent gains. These studies summarise that the suppression rate is constant irrespective of the degree of incompleteness of the calibration sky model. In the presence of a direction-dependent effect, the suppression drastically increases; however, this increase can be compensated for by using longer solution intervals.

Acknowledgements

“we don't stand alone in this world...”

To all those who played their roles well...

Almighty God for marking this “epoch” (... the astronomer in me is waking up ...) in my life. His grace and love held things together and helped me assemble my Master's degree jigsaw puzzle in a beautiful manner.

Honour thy mother and father

Mom, Dad; your little girl has finished working on a thesis using the same English alphabet you taught her (well, along with some Mathematics as well). Your scoldings and long boring speeches (did not have any other choice other than shut my mouth and pretend to be a good listener ...) have indeed helped me sail through my difficulties.

Growing up with siblings is a nice thing; however, they don't help you write your thesis. But their support strengthens you to cross the finish line in all splendour. To my brother Jhavikesh, we are a team!!! And I love you!

Talking about Rhodents (from RATT) and other mammals (every human being is a mammal), I express my heartfelt gratitude to Marcellin, Sphesihle, Liju (Lee), Theophilus, Elizabeth and Mark for their trust and encouragement. Marcel and his rules ..., I hope one day I will be able to abide by your so called “RATT rules”. Sphe - the cool dude who drags us to the most renowned RAT & PARROT; the day I decide to have a drink, I will surely join you, but I am afraid that day might never come! Charming and forever happy Lee, you made all the unpleasant moments of my life disappear in the blink of an eye. Our mathematician, Theo, you inspired me to spend sleepless nights, long office hours solving complex equations. Guys, your kid has finally made it!. Thank you for this pleasant journey.

People who got me to South Africa - Prof Oleg Smirnov and Dr Trienko Grobler. Thank you Oleg for taking me to the depths of astronomical research - reminds me of the movie ‘Inception’. Trienko - thank you for reading my thesis more than I did; that was indeed PATIENCE! Not to forget, Dr Griffin Foster for his useful (not always though) comments and suggestions. When it comes to efficient, timely office work - Mrs Ronel Groenewald is the person you want in your team! You made it look so easy.

People back in my homeland, Daryl, Shandiv, Selven, Veer and Yash, I wonder do they even realise South Africa and Mauritius are separated by an ocean? They made the distance shrink and tied our hearts with a beautiful bond of friendship.

Gratitude to Dr Noor Sookia and Dr Dinesh Somanah for their motivational guidance toward pursuing further studies.

I thank the National Research Foundation for funding my Masters research.

Grahamstown ...is awesome!

To my parents. Thank you Mom and Dad for believing in me . . .

Contents

Abstract	i
Acknowledgements	ii
Contents	v
Physical Constants	vii
1 General Introduction	1
1.1 Radio Astronomy - Unveiling the Mystery of the Unseen Universe	1
1.2 Why Radio Astronomy?	2
1.2.1 Radio interferometer	2
1.2.2 Calibration and imaging	4
1.3 Statement of Problem	4
1.4 Research Objectives	6
1.5 Delimitation	7
1.6 Significance of study	7
1.7 Thesis Outline	8
1.8 Summary	8
2 Calibration Techniques	9
2.1 Measurement Equation	9
2.2 Alternating Least Squares	12
2.2.1 Eigenvalue decomposition	13
2.2.2 Parameter estimation	14
2.3 Self-calibration (Selfcal)	15
2.3.1 Levenberg-Marquadt	16
2.3.2 Statistically efficient and fast calibration	18
2.4 Differential Gains	18
2.5 The Different Calibration Regimes	20
2.6 Conclusion	20
2.7 Summary	21
3 Ghost Phenomena in WRST Interferometric data	22
3.1 Experimental Setup	22

3.2	Configuration of WSRT	24
3.3	Theoretical Analysis	26
3.3.1	The extrapolated visibility matrix	26
3.3.2	ALS calibration	27
3.3.2.1	Essential propositions	28
3.3.2.2	ALS calibration matrix	28
3.4	Ghost Formation	31
3.4.1	Distillation	31
3.4.2	Imaging	32
3.4.3	Corrected visibilities	33
3.5	Results and Discussion	35
3.5.1	Ghost patterns	36
3.5.2	Dependence on flux ratio	39
3.5.3	Dependence on the source position and antennas	42
3.5.4	The role of array redundancy	43
3.6	Conclusion	44
3.7	Summary	45
4	Source Suppression	46
4.1	Spectral Flux Density	46
4.2	Implementation of calibration pipeline	48
4.2.1	Experimental setup	50
4.3	Source Suppression Estimates	50
4.3.1	Unweighted average	51
4.3.2	Weighted Average	51
4.3.3	Ordinary least square (OLS) estimation	52
4.3.4	Optimal estimator	52
4.4	Incomplete Calibration Sky Models and Source Suppression	57
4.4.1	Suppression with unmodelled flux	59
4.4.2	Analysis of differential gains, ΔE	61
4.5	Observations with 3C 147	62
4.5.1	Analysis of source suppression	63
4.6	Conclusion	66
4.7	Summary	66
5	General Conclusion	67
A	APPENDIX A: Lemmas and Propositions	70
	Bibliography	78

Physical Constants

Speed of light	c	$=$	$2.99792458 \times 10^8 \text{ ms}^{-1}$
Boltzmann constant	k	$=$	$1.3806488 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$
Planck's constant	h	$=$	$6.62606957 \times 10^{-34} \text{ m}^2 \text{ kg s}^{-1}$

Chapter 1

General Introduction

“Astronomy –past in the present and present in the future.”

–Anonymous

1.1 Radio Astronomy - Unveiling the Mystery of the Unseen Universe

Astronomy is one of the oldest sciences and was practised by the Greeks, Mayans, Chinese, Indians and Babylonians. Ancient astronomy revolved around the geocentric model, which described earth at the orbital centre of all celestial bodies. This model was abolished by Nicolaus Copernicus’s heliocentric theory, which paved the way for modern astronomy. With the advent of telescopes, which were used to explore the cosmos, astronomy became a diverse field in itself. Powerful telescopes backed up with modern technology aided understanding about the mystical universe we live in. Modern astronomical observations are made across the electromagnetic spectrum, which ranges from high-frequency gamma radiation to low frequency radio waves. Radio astronomy in particular operates in the low frequency region of the electromagnetic spectrum.

A frequently asked question is, “Do radio astronomers see radio sources or do they listen to them?” The terms “see” and “listen” are just figures of speech but literally speaking, the signals emitted by these radio sources belong to a part of the electromagnetic spectrum that is not directly visible to the eyes. These signals are therefore captured using radio telescopes and eventually processed into radiographs via electronics. These radiographs unveil the deepest mysteries in the cosmos. The more distant an object is, the longer the light it emits will take to reach the earth, that is, the instantaneous snapshot

that is being observed today is actually a past event in time. Astronomical distances are therefore interpreted in terms of vast spans of time and the most commonly used unit for them is light-years.

1.2 Why Radio Astronomy?

Stars and galaxies are in general distant celestial bodies that appear like point sources in the sky. These point sources can mainly be attributed to galactic emission in the radio spectrum. The radio spectrum is dominated by non-thermal emissions, unlike the visible spectrum, which is dominated by thermal emissions, thus the two complement each other. The optical spectrum measures the brightness of the stars while long radio wavelengths can probe the cosmic dust and gases hidden in the interstellar medium.

Radio telescopes can either be used individually, referred to as single-dish telescopes, or be connected together to form a radio interferometer. The antennas may be close to each other or several kilometres apart.

1.2.1 Radio interferometer

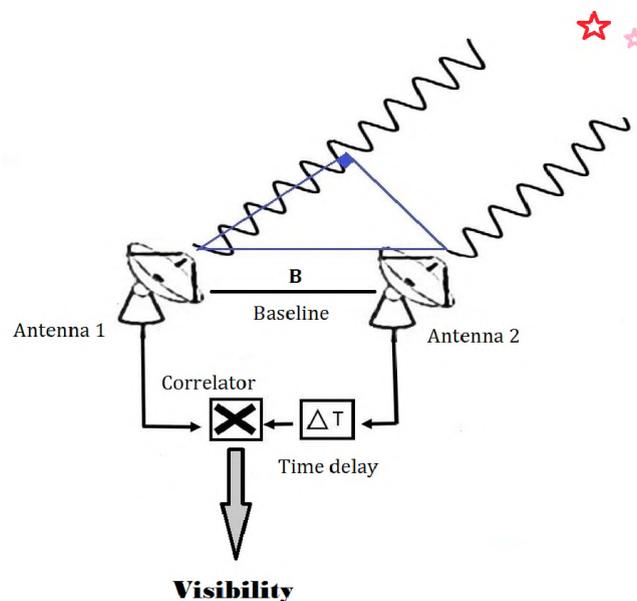


FIGURE 1.1: Two-element radio interferometer

Radio interferometers combine the radio waves emitted by an astronomical source to a wave diagnostic to the actual observation. Radio interferometry is preferred in radio astronomy as it yields high-resolution images. The angular resolution of an interferometer

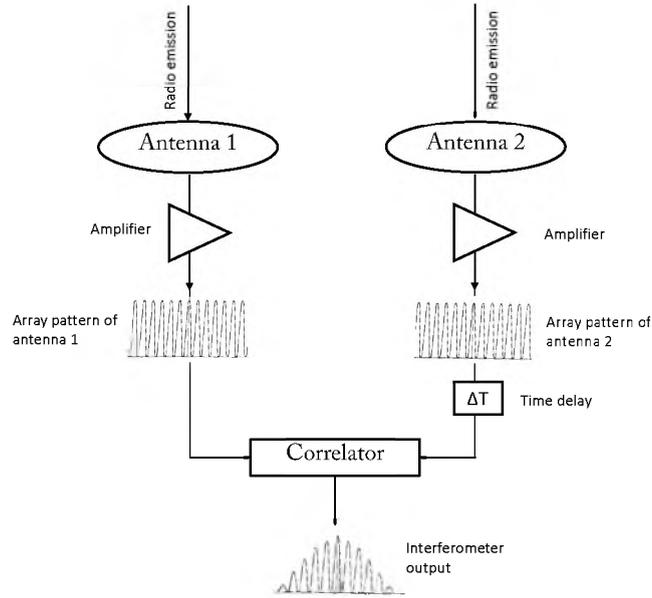


FIGURE 1.2: Block-diagram of a two-element radio interferometer

response is given by

$$\theta = \frac{\lambda}{\mathbf{B}}, \quad (1.1)$$

where λ denotes the wavelength of the signal being collected and \mathbf{B} is the baseline length measured in metres.

An illustration of a simple two-element interferometer pointing at a particular region in the sky can be seen in Fig. 1.1. By the time the signals reach the antennas, they behave as plane transverse waves. The signals are received at two unique points in time, thus introducing a delay factor. This delay is compensated for by a time delay ΔT to the signal, as shown in Fig. 1.2. From Fig. 1.2, we can see that the two waves are combined at the correlator. The correlator output is called the visibility function or simply visibilities.

The Van Cittert-Zernike theorem (Thompson et al. (2001)) states that the visibility function is the Fourier transform of the sky brightness given by

$$V(u, v, w = 0) = \int_l \int_m \frac{A_n(l, m) I(l, m)}{\sqrt{1 - l^2 - m^2}} e^{-i2\pi(ul+vm)} dl dm, \quad (1.2)$$

where $I(l, m)$ is the sky intensity and A_n is the sensitivity of the optics and antenna for a given (l, m) position. The baseline coordinate system is denoted by (u, v, w) , where u and v are the angular scales in the East-West and North-South direction and w is directed towards the phase centre. In Eq. 1.2, (l, m, n) are mapped to an (u, v, w) coordinate system where n is pointing towards the field centre and it approximates to

one. The two-dimensional Cittert-Zernike theorem can be represented as

$$V(u, v) = \int_l \int_m I(l, m) e^{-i2\pi(ul+vm)} dl dm, \quad (1.3)$$

where A_n is considered to be unity for the sake of simplicity. The source geometry is stored in the exponential term. Eq. 1.3 holds only if the signal is uncorrupted. However, this is not the case in reality, as the radiation is influenced by instrumental gains, antenna feed polarisation leakage and atmospheric conditions along the signal path. The above stated corruptions must therefore be mitigated. The process of minimising these errors is known as calibration.

1.2.2 Calibration and imaging

The corruptions associated with a radio signal can be classified as direction-independent effects (DIEs) and direction-dependent effects (DDEs). DIEs are constant irrespective of the source positions and these include the direction-independent electronic gains, while DDEs are associated with the source geometry. The most popular calibration technique used to approach the real sky distribution is self-calibration, abbreviated as selfcal (Pearson and Readhead (1984); Boonstra and van der Veen (2003); Wijnholds and Van der Veen (2009); van der Tol et al. (2007); Intema et al. (2009)).

The acquired visibility data $V(u, v)$ are continuous on a set of (u, v) coordinates, whereas the interferometer measures the intensity for a subset of the (u, v) set. The visibility function is therefore convolved by a sampling function $S(u, v)$. The sampled visibilities are now Fourier transformed back into a sky distribution estimate

$$I(l, m) = \int_u \int_v S(u, v) V(u, v) e^{i2\pi(ul+vm)} dudv.$$

The sky estimate $I(l, m)$ is termed the “dirty image”, since it convolves with the point spread function (PSF) of the instrument, which produces sidelobes. The sidelobes are due to the gaps between the measured (u, v) points. The dirty images are deconvolved to suppress these sidelobes via deconvolution algorithms; this decomposes the sky distribution into point-source components and convolves them with a clean beam (Cornwell et al. (1999)).

1.3 Statement of Problem

Naive calibration leads to distortions in the radio maps, deformation of extended sources, loss of real emission or generation of spurious sources (Linfield (1986); Cornwell and

Wilkinson (1981); Bridle and Schwab (1999); Martí-Vidal and Marcaide (2008); Martí-Vidal et al. (2010); Kazemi and Yatawatta (2013); Grobler et al. (2014)). Martí-Vidal and Marcaide (2008) found that spurious sources can be generated when selfcal is applied to white noise. A novel approach is proposed by Kazemi and Yatawatta (2013) whereby the recovered amount of flux that is retrieved is maximised by using a t -distribution.

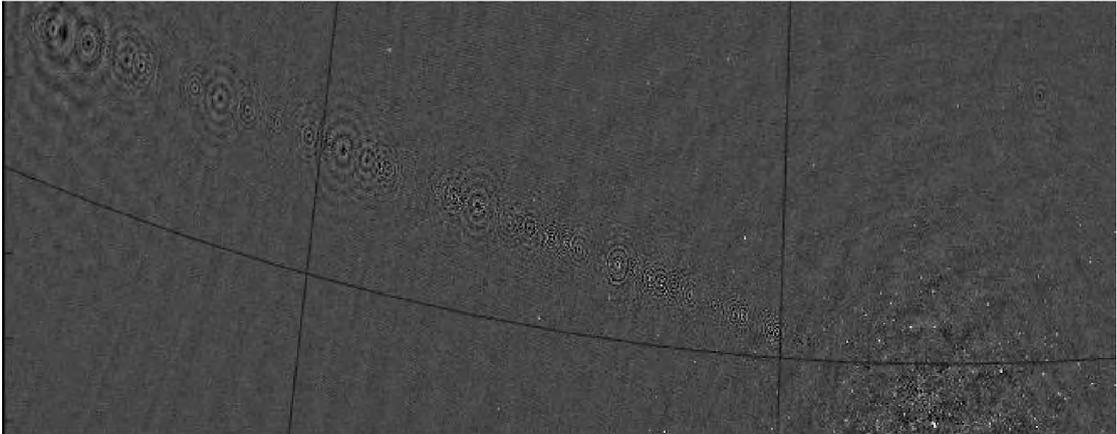


FIGURE 1.3: Ghost sources in a 92 cm WSRT observation of $J1819+3845$. The target field is in the lower-right corner of the image and Cyg A is to the upper left, just outside the image (Credit de Bruyn).

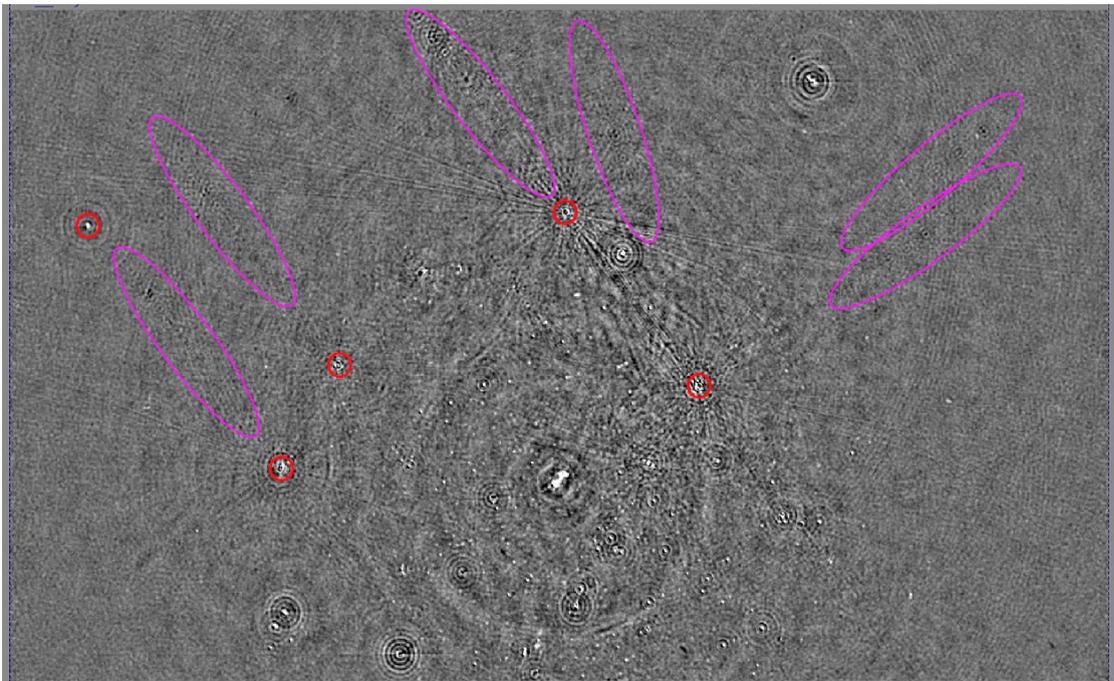


FIGURE 1.4: Residual map with ghost sources in a 21 cm WSRT observation of the QMC2 field. The sky sources have been subtracted, therefore the bright sources in the images are in fact relatively faint. The positions of the subtracted sources are indicated by the red circles and the string of ghosts are highlighted by the ellipses. Note how the strings are firmly associated with the brightest sources. (Credit O. Smirnov)

During a 92 cm Westerbork Synthesis Radio Telescope (WSRT) observation of $J1819 + 3845$, carried out by de Bruyn, a strange phenomenon was observed, illustrated in Fig. 1.3. The maps obtained after applying selfcal exhibited a string of point-source like objects referred to as “ghosts”. They were found along the line linking the brightest source with Cyg A, which was located 20° away from the brightest object in the field of view (FoV). The ghosts followed a specific pattern shown in Fig. 1.3 and their positions were independent of frequency. This puzzle was not thoroughly investigated until the “Quality Monitoring Committee (QMC)” project in 2010 (Smirnov (2010)). During the QMC project, a large pointing error was deliberately added and several strings of ghost sources appeared, as pictured in Fig 1.4. The problem was further analysed through a set of simulations by Smirnov (2011a) and the following observations were pointed out:

- The appearance of the ghost sources in the calibrated maps were due to incomplete sky models or because of DDEs. In the case of the $J1819 + 3845$ observation, the missing flux or inaccurate modelling of Cyg A was responsible for the formation of the spurious sources, while for QMC, this behaviour was caused by the large pointing errors. Surprisingly, the ghosts disappear when differential gains are applied to solve for the DDEs.
- A simple simulation of a two-source sky model of 1 Jy and 1 mJy, with the 1 Jy source acting as the dominator in the calibration model and the other source acting as the contaminator, yields a similar ghost pattern in the residual maps. The highest peak of the ghost pattern seems to be directly proportional to the flux of the contaminator, but independent of the model flux. This indicates that ghost sources will always appear when incomplete sky models are being used. However, they are often buried in the thermal noise because of their low intensities.
- The ghost sources are formed along the line(s) joining the brightest sources with the unmodelled sources. Ghosts with different intensities occur at specific rational fractions (i.e $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{5}$, etc).
- The ghost sources produce PSF-like sidelobes but these sidelobes are not identical to the PSF of the radio telescope.

1.4 Research Objectives

The purpose of this study is to achieve a deeper understanding of the ghost phenomenon observed in Fig. 1.3 and 1.4. This work is basically an extension of our previously published paper Grobler et al. (2014). It aims at attaining a good analytical framework

that explains the ghost sources, marking their positions and intensities. The influence of these ghost sources on the real emission is then studied. Furthermore, this work attempts to find a link between the incomplete sky models and source suppression, where source suppression refers to the amount of flux lost during calibration. Lastly, the effect of DDEs on source suppression is investigated.

1.5 Delimitation

This entire work is based on observations using the East-West (EW) WSRT. The WSRT is known for its regular antenna layout. The description of the ghost patterns is therefore restricted to EW regularly spaced arrays.

1.6 Significance of study

Recent advances in technology have promoted the development of interferometers with large FoVs, low instrumental errors and reasonable angular resolutions. These instruments are thus able to provide measurements of the cosmological dark ages and Epoch of Reionisation EoR (Zaroubi (2012)), though the stated applications are not feasible without efficient and precise calibration algorithms. The cosmological dark ages are measured through intensity mapping (Villaescusa-Navarro et al. (2014)) and stacking (Delhaize et al. (2013)). Intensity mapping is the technique of accumulating radio waves emitted by hydrogen gas in large volumes of space, while stacking combines weak galactic signals into a stronger one. These aforementioned techniques are prone to calibration errors. EoR marks the transition of the primordial hydrogen from the neutral to ionised state. It involves detecting EoR signals, probing large-scale structures at the end of reionisation. This calls for the use of incomplete sky models since complete removal of the foreground is tough to achieve. In addition, given that the detection of transient sources entails very fast-on-the-fly calibration, the use of incomplete sky models is inevitable. Future radio telescopes, such the Square Kilometre Array (SKA), probing faint emissions, call for a thorough study of the ghost sources caused by the use of incomplete sky models, mentioned in Section 1.3.

1.7 Thesis Outline

The remaining part of the thesis is divided into four chapters:

Chapter 2: Literature Review on Calibration Techniques: This chapter commences with the introduction of the measurement equation (ME) (Sault et al. (1996)) and then proceeds with some direction-independent antenna-based calibration techniques: alternating least squares (ALS) (Boonstra and van der Veen (2003)), self-cal (Pearson and Readhead (1984); Boonstra and van der Veen (2003); Wijnholds and Van der Veen (2009); van der Tol et al. (2007); Intema et al. (2009)) and SteFCal (Salvini and Wijnholds (2014)). It continues with solving for DDEs by applying differential gains and lastly it states the criterion by which these calibration techniques are designed.

Chapter 3: Ghost Phenomena in WRST Interferometric data: This chapter focuses on the investigation of ghost phenomenon as described in Grobler et al. (2014). It provides a theoretical framework, describing the ghost distribution. The theoretical results are compared with simulated outputs obtained using MeqTrees package (Noordam and Smirnov (2010)). The optimisation algorithms used are Alternating Least Squares (ALS) and Least Squares (LS). My contribution towards Grobler et al. (2014) revolved around performing simulations and experiments to scrutinise the discrepancies between theoretically predicted and simulated results.

Chapter 4: Source Suppression: The fourth chapter investigates the influence of incomplete sky models on the recovered flux density from simulated maps. Source suppression is then studied as function of the number of sources subjected to DDEs and the solution time interval.

Chapter 5: General Conclusion: The last chapter draws conclusions based on the results and lists suggestions for future work.

1.8 Summary

This chapter acquaints us with the world of astronomy, in particular radio astronomy. The concept of radio interferometry was also briefly discussed. The purpose of this study was then introduced, followed by the breakdown of the thesis. Chapter 2 gives an overview of the calibration algorithms that will be useful in the investigation.

Chapter 2

Calibration Techniques

“ Astronomy is the science of the
harmony of infinite expanse.”

–John Scott Russell

Electronic imperfections generally lead to polarisation leakages in radio dishes, manifesting themselves as direction-independent antenna-based gains. Antenna-based gain calibration is a key step in the creation, of radio interferometric maps. However, it is not sufficient since the emissions are also influenced by the primary beam and the atmospheric conditions, which depend on the direction of the source. Solving for the direction-dependent factors mentioned above is a major challenge for radio astronomy. This chapter delineates some of the frequently used optimisation algorithms to solve for the direction-independent and direction-dependent parameters.

2.1 Measurement Equation

This section acquaints us with the measurement equation derived in [Sault et al. \(1996\)](#). The measurement equation describes the response of an interferometric array. Considering a single point source, the electromagnetic field in the xyz plane, with z pointing in the direction of radiation is assumed to be a plane wave and is therefore denoted by

$$\mathbf{e} = \begin{pmatrix} e_x \\ e_y \end{pmatrix}.$$

The receivers then convert the electromagnetic radiation into voltages. If all the propagation effects associated with the signal are linear with respect to \mathbf{e} , then

$$\mathbf{v} = \mathbf{J}\mathbf{e}, \quad (2.1)$$

where \mathbf{J} is a 2×2 matrix, which represents all the transformations along the signal path. The matrix \mathbf{J} is known as the *Jones matrix*. It can be broken up into

$$\mathbf{J} = \mathbf{J}_1 \mathbf{J}_2 \mathbf{J}_3 \dots \mathbf{J}_m, \quad (2.2)$$

with each of its factors being a unique effect along the signal path. The factors in Eq. 2.2 are arranged according to the physical order in which the transformations occur along the path. The order of the *Jones* matrix needs to be preserved, as matrices do not generally commute with each other.

Coming back to the two-element interferometer discussed in Subsection 1.2.1, the voltages measured by any two antennas p and q are

$$\begin{aligned} \mathbf{v}_p &= \mathbf{J}_p \mathbf{e}_p \\ \mathbf{v}_q &= \mathbf{J}_q \mathbf{e}_q. \end{aligned} \quad (2.3)$$

The corresponding cross correlations are

$$\begin{aligned} V_{xx} &= \langle v_{px}, v_{qx}^* \rangle \\ V_{xy} &= \langle v_{px}, v_{qy}^* \rangle \\ V_{yx} &= \langle v_{py}, v_{qx}^* \rangle \\ V_{yy} &= \langle v_{py}, v_{qy}^* \rangle, \end{aligned}$$

where $\langle \cdot \rangle$ represents averaging over time and v^* is the complex conjugate of v . The correlated values obtained for any baseline pq can be packed into a matrix:

$$\begin{aligned} \mathbf{V}_{pq} &= 2 \left\langle \begin{pmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{pmatrix} \right\rangle \\ &= 2 \begin{pmatrix} \langle v_{px} v_{qx}^* \rangle & \langle v_{px} v_{qy}^* \rangle \\ \langle v_{py} v_{qx}^* \rangle & \langle v_{py} v_{qy}^* \rangle \end{pmatrix} \\ &= 2 \left\langle \begin{pmatrix} v_{px} \\ v_{py} \end{pmatrix} \begin{pmatrix} v_{qx}^* & v_{qy}^* \end{pmatrix} \right\rangle \\ &= 2 \langle \mathbf{v}_p \mathbf{v}_q^H \rangle. \end{aligned} \quad (2.4)$$

This matrix is known as the *visibility matrix*. The factor 2 is introduced as a convention and will be explained later in this section. Substituting Eq. 2.3 into Eq. 2.4 results in

$$\mathbf{V}_{pq} = 2\langle (\mathbf{J}_p \mathbf{e}_p)(\mathbf{J}_q \mathbf{e}_q)^H \rangle = 2\langle \mathbf{J}_p (\mathbf{e}_p \mathbf{e}_q^H) \mathbf{J}_q^H \rangle = 2\mathbf{J}_p \langle \mathbf{e}_p \mathbf{e}_q^H \rangle \mathbf{J}_q^H.$$

The Jones factors \mathbf{J}_p and \mathbf{J}_q are taken out of the inner brackets as they are assumed to be constant over the averaging interval. Sault et al. (1996) showed that the inner product $\langle \mathbf{e}_p \mathbf{e}_q^H \rangle$ can be expressed as the Stokes parameters,

$$2\langle \mathbf{e}_p \mathbf{e}_q^H \rangle = \begin{pmatrix} I + Q & U + iV \\ U - iV & I - Q \end{pmatrix} = \mathbf{C}, \quad (2.5)$$

where \mathbf{C}_{pq} denotes the coherency matrix. By definition,

$$I = \langle |e_x|^2 + |e_y|^2 \rangle \quad ; \quad Q = \langle |e_x|^2 - |e_y|^2 \rangle.$$

In the case of an unpolarised point source with unit power, that is, $I = 1$ and $Q = U = V = 0$,

$$|e_x|^2 = |e_y|^2 = \frac{1}{2}.$$

Now consider the corresponding interferometer output. The two conventions worth considering are (Smirnov (2011b)):

Convention- $\frac{1}{2}$: Each correlation corresponds to $\frac{1}{2}$.

$$\mathbf{V}_{pq} = \begin{pmatrix} \langle |e_x|^2 \rangle & 0 \\ 0 & \langle |e_y|^2 \rangle \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Convention-1: Each correlation corresponds to unity.

$$\mathbf{V}_{pq} = 2 \begin{pmatrix} \langle |e_x|^2 \rangle & 0 \\ 0 & \langle |e_y|^2 \rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Convention- $\frac{1}{2}$ is the correct one in theory, and conforms to IEEE definitions. However, and somewhat unfortunately, convention-1 is much more widely used in practice, including in the most common software packages (see discussion in Smirnov (2011b)). We therefore adopt convention-1 through the rest of this work. The per-baseline interferometer response is hence equal to

$$\mathbf{V}_{pq} = \mathbf{J}_p \mathbf{C} \mathbf{J}_q^H. \quad (2.6)$$

As per Eq. 2.2, the per-baseline visibility in Eq. 2.6 can be expanded into

$$\mathbf{V}_{pq} = \mathbf{J}_{pm}(\dots(\mathbf{J}_{p2}(\mathbf{J}_{p1}\mathbf{C}\mathbf{J}_{q1}^H)\mathbf{J}_{q2}^H)\dots)\mathbf{J}_{qm}^H. \quad (2.7)$$

The source and antenna positions introduce a phase difference between \mathbf{v}_p and \mathbf{v}_q , as explained in Subsection 1.2.1. The phase delay at the p^{th} antenna is given by $\mathbf{K}_p = e^{-2\pi i(u_p l + v_p m + w_p(n-1))}\mathbf{I}$, where \mathbf{I} is the identity matrix. In addition to the phase delay \mathbf{K}_p , the signal is subjected to direction-independent gains (electronic gains), \mathbf{G}_p and direction-dependent gains (beam gains) \mathbf{E}_p due to polarisation leakages. Consider a sky distribution of N sources, then Eq. 2.6 can be restated as

$$\begin{aligned} \mathbf{V}_{pq} &= \sum_s \mathbf{J}_{sp} \mathbf{C}_s \mathbf{J}_{sq}^H \\ &= \mathbf{G}_p \sum_s \left(\mathbf{E}_{sp} \mathbf{K}_{sp} \mathbf{C}_s \mathbf{K}_{sq}^H \mathbf{E}_{sq}^H \right) \mathbf{G}_q^H. \end{aligned} \quad (2.8)$$

Since the electronic gain \mathbf{G}_p is source-independent, it can be taken out of the summation. Note that a noise-free measurement is being considered here. In practice, the measurement V_{pq} is corrupted by additive uncorrelated Gaussian noise (Thompson et al. (2001)),

$$\mathbf{V}_{pq} = \mathbf{G}_p \sum_s \left(\mathbf{E}_{sp} \mathbf{K}_{sp} \mathbf{C}_s \mathbf{K}_{sq}^H \mathbf{E}_{sq}^H \right) \mathbf{G}_q^H + \mathbf{N}_{pq}; \quad \mathbf{N}_{pq} \in \mathbb{C}^{2 \times 2}. \quad (2.9)$$

Eq. 2.9 is known as the radio interferometry measurement equation (RIME). The process of estimating the parameters listed in Eq. 2.9 is known as calibration. Some of the calibration approaches are discussed in the coming sections. Note that from now onwards, unpolarised visibility data are assumed for simplicity, unless specified otherwise.

2.2 Alternating Least Squares

Alternating Least Squares (ALS) (Boonstra and van der Veen (2003)) is one of the pioneering calibration techniques that has been developed to solve for the unpolarised direction-independent gain complex scalar g_p . To solve for the electronic gains, a single point source with known sky brightness, situated at the phase centre, is observed. The measurement equation in Eq. 2.9 then reduces to

$$\tilde{V}_{pq} = g_p B g_q^* + n_{pq}, \quad (2.10)$$

where \tilde{V}_{pq} denotes the unpolarised visibility along baseline pq , and $B = \mathbf{K}_p \mathbf{C} \mathbf{K}_q^H$. The complex conjugate of g is indicated by g^* . Given n antennas, the visibilities \tilde{V}_{pq} can be

packed into a matrix \mathbf{R} such that

$$\mathbf{R} = \mathbf{G}\mathbf{B}\mathbf{G}^H + \mathbf{N}, \quad (2.11)$$

with $\mathbf{g}\mathbf{g}^H \odot \mathbf{B} = \mathbf{G}\mathbf{B}\mathbf{G}^H$, where $\mathbf{G} = \text{diag}(\mathbf{g}) = \text{diag}\{g_1, g_2, \dots, g_n\}^T$ and \mathbf{N} is the noise matrix. The visibility matrix \mathbf{R} can be estimated by averaging over N samples of cross-correlated interferometric measurements or voltages \mathbf{v}

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{i=1}^N \mathbf{v}_i \mathbf{v}_i^*. \quad (2.12)$$

The estimates of \mathbf{G} and \mathbf{N} are calculated using least square minimisation

$$\{\mathbf{G}, \mathbf{N}\} = \arg \min \|\hat{\mathbf{R}} - (\mathbf{G}\mathbf{B}\mathbf{G}^H + \mathbf{N})\|_F^2, \quad (2.13)$$

where $\|\cdot\|_F$ is the Frobenius norm. We use eigenvalue decomposition to solve for \mathbf{G} and \mathbf{N} .

2.2.1 Eigenvalue decomposition

Eigenvalue decomposition (also known as matrix diagonalisation) refers to the factorisation of a square matrix into three matrices. Let \mathbf{A} be an $n \times n$ square matrix. The eigenvalues and eigenvectors of \mathbf{A} are denoted by $(\lambda_1, \lambda_2, \dots, \lambda_n)$ and $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ respectively. Taking $\mathbf{\Lambda}$ as a diagonal matrix with $(\lambda_1, \lambda_2, \dots, \lambda_n)$ as its diagonal terms, and \mathbf{X} as a matrix with \mathbf{x}_j as its j^{th} column, the diagonalisation of \mathbf{A} is given by

$$\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}. \quad (2.14)$$

It should be noted that \mathbf{X} is arranged such that each eigenvector is multiplied by its appropriate eigenvalue. Since the eigenvectors of \mathbf{A} are linearly independent,

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}. \quad (2.15)$$

In other words, \mathbf{A} can be decomposed into three matrices:

$$\mathbf{A} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nn} \end{bmatrix}.$$

If a complex Hermitian matrix is used, then its eigenvalue decomposition is given by

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^H, \quad (2.16)$$

with $\mathbf{X}\mathbf{X}^H = \mathbf{X}^H\mathbf{X} = \mathbf{I}$. Though the eigenvalues are real, the corresponding eigenvectors are complex. With all the necessary equations in place, we proceed with the optimisation of the unknown parameters.

2.2.2 Parameter estimation

The parameters \mathbf{G} and \mathbf{N} are solved interchangeably. ALS solves for \mathbf{G} , keeping the diagonal matrix \mathbf{N} fixed, and then solves for \mathbf{N} , keeping \mathbf{G} constant. Let $\mathbf{N}^{(k)}$ be the estimate of \mathbf{N} at the k^{th} iteration. The gains $\hat{\mathbf{G}}$ are then estimated using

$$\hat{\mathbf{G}}^{(k)} = \arg \min_{\mathbf{G}} \|\hat{\mathbf{R}} - \mathbf{G}\mathbf{G}^H - \hat{\mathbf{N}}^{(k)}\|_F^2. \quad (2.17)$$

Eq. 2.17 is solved by using the following eigenvalue decomposition

$$\hat{\mathbf{R}} - \hat{\mathbf{N}}^{(k)} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^H,$$

where the diagonal matrix $\mathbf{\Lambda}$ has the eigenvalues of $\hat{\mathbf{R}} - \hat{\mathbf{N}}^{(k)}$, λ_i as its diagonal terms. The corresponding eigenvectors, \mathbf{x}_i , are stored in the \mathbf{X} matrix. The optimal gain estimate is therefore given by

$$\hat{\mathbf{G}}^{(k)} = \mathbf{x}_1\sqrt{\lambda_1}, \quad (2.18)$$

with λ_1 and \mathbf{x}_1 referring to the largest eigenvalue and its corresponding eigenvector respectively. The system noise matrix \mathbf{N} is minimised using

$$\hat{\mathbf{N}}^{(k+1)} = \arg \min_{\mathbf{N}} \|\hat{\mathbf{R}} - \hat{\mathbf{G}}^{(k)}\hat{\mathbf{G}}^{(k)H} - \mathbf{N}\|_F^2. \quad (2.19)$$

In Eq. 2.19, the negative entries of \mathbf{N} are set to zero so that the matrix has only non-negative diagonal terms. Subtracting $\hat{\mathbf{G}}^{(k)}\hat{\mathbf{G}}^{(k)H}$ from $\hat{\mathbf{R}}$ and discarding off-diagonal terms result in an estimate of \mathbf{N} , which is equal to

$$\hat{\mathbf{N}}^{(k+1)} = \text{diag}\|\hat{\mathbf{R}} - \hat{\mathbf{G}}^{(k)}\hat{\mathbf{G}}^{(k)H}\|_F^2. \quad (2.20)$$

Eq. 2.17 and 2.19 are iterated until a monotonic convergence is reached. The convergence rate is very slow, especially if a proper initial point is not chosen. However, ALS has not been popular because of lack of implementation.

2.3 Self-calibration (Selfcal)

Selfcal is a method introduced to solve for antenna-based errors, whereby the complex antenna gains are treated as free parameters. The procedure for selfcal is described in Fig. 2.1.

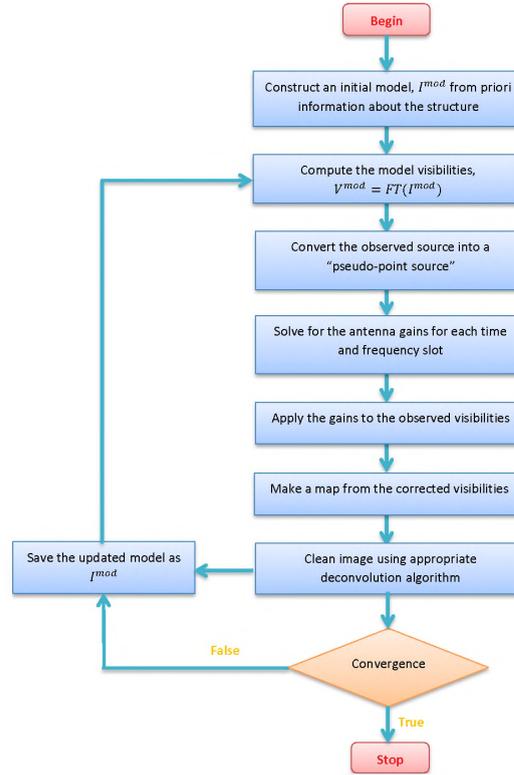


FIGURE 2.1: Self-calibration scenario

A calibration sky model, I^{mod} , is considered. This is constructed based upon prior knowledge about the source or sky distribution. The model can be initialised with a single point source. The sky distribution is Fourier transformed into model visibilities. The difference between the predicted visibilities and model visibilities can be minimised in an LS sense, as explained in Schwab (1980)

$$\sum_k \sum_{p,q} w_{pq}(t_k) |V_{pq} - g_p(t_k) V_{pq}^{mod} g_q^*(t_k)|^2, \quad (2.21)$$

where $w_{pq}(t_k)$ are weights applied to the various interferometer pairs. The weights are assigned such that they are inversely proportional to the variance of the measured visibilities. Eq. 2.21 can be rewritten as

$$\sum_k \sum_{p,q} w_{pq}(t_k) |V_{pq}^{mod}| |X_{pq} - g_p(t_k) g_q^*(t_k)|^2, \quad (2.22)$$

where the “pseudo-point source”

$$X_{pq} = \frac{V_{pq}}{V_{pq}^{mod}}. \quad (2.23)$$

Since the antenna gains are treated as free parameters, the data can be processed simultaneously rather than dealing with the amplitude and phase individually. However, the flux density of the final map has a tendency to decrease, since the total intensity of the model is less than the actual intensity. Selfcal uses an incomplete model field, therefore some of the fainter structures might not be recovered. Furthermore, a reasonable integration time interval should be used in order to avoid signal loss.

The following subsections deal with the common optimisation algorithms used to solve for the antenna-based gains.

2.3.1 Levenberg-Marquadt

Consider n stations of an array, forming $\frac{n(n-1)}{2}$ distinct interferometers, which measures data as specified in Eq. 2.9. The calibration process involves estimating J_p along each possible baseline, and for the individual sources in the sky. Because of the earth’s rotation, the propagation effects, stored as *Jones* factor J_p , are functions of time and frequency. We therefore assumed J_p to be constant within a specific time interval τ .

With K discrete sources in the sky, the measurement equation in Eq. 2.9 can be written as an unpolarised optimisation problem in vector form

$$\tilde{\mathbf{v}}_{pq} = \sum_{s=1}^K \mathbf{s}_{pqs}(\theta) + \mathbf{n}_{pq}, \quad (2.24)$$

where $\mathbf{s}_{pqs} = J_{qs}^* \otimes J_{ps} C_{pqs}$, $\tilde{\mathbf{v}}_{pq} = [v_{11}, v_{12}, v_{nn}]^T$ and $\mathbf{n}_{pq} = [n_{11}, n_{12}, \dots, n_{nn}]^T$.

The propagation effects in J_p are the set of parameters θ for the optimisation problem. Since they are a complex number, both real and imaginary parts need to be estimated. Hence the total number of parameters to solve for is $n \times K \times 2$ ($\theta \in \mathbb{C}^{2Kn}$). Eq. 2.24 can therefore be stacked up as (Yatawatta et al. (2012)):

$$\begin{aligned} \mathbf{s}_s(\theta) &= [\text{Re}(\mathbf{s}_{12s}^T), \text{Im}(\mathbf{s}_{12s}^T), \text{Re}(\mathbf{s}_{13s}^T), \text{Im}(\mathbf{s}_{13s}^T), \dots]^T; & \mathbf{s}_s(\theta) &\in \mathbb{R}^{\tau n(n-1)} \\ \mathbf{y} &= [\text{Re}(\mathbf{v}_{12}^T), \text{Im}(\mathbf{v}_{12}^T), \text{Re}(\mathbf{v}_{13}^T), \text{Im}(\mathbf{v}_{13}^T), \dots]^T; & \mathbf{y} &\in \mathbb{R}^{\tau n(n-1)}. \end{aligned}$$

The values of n , K and τ depend on the radio array, sky coverage and synthesis time. The maximum likelihood of the unknown θ is equal to

$$\hat{\theta} = \arg \min_{\theta} \left\| \mathbf{y} - \sum_{s=1}^K \mathbf{s}_s(\theta) \right\|^2. \quad (2.25)$$

The process iterates until the observed visibilities converge to the predicted ones. The cost function $f(\theta)$ associated with Eq. 2.25 is a mapping from \mathbb{R}^{2nK} to \mathbb{R}

$$f(\theta) = \sum_{s=1}^K \mathbf{s}_s(\theta).$$

If θ_0 is the initial value, then the estimate $\theta^{(k+1)}$ at the $(k+1)^{th}$ iteration

$$\theta^{(k+1)} = \theta^{(k)} - \left(\nabla_{\theta} \nabla_{\theta}^T f(\theta) + \lambda \mathbf{H} \right)^{-1} \nabla_{\theta} f(\theta), \quad (2.26)$$

as stated in [Madsen et al. \(2004\)](#). The damping factor, λ , is defined as the change at each iteration in Eq. 2.25 and $\mathbf{H} = \text{diag} \left(\left(\nabla_{\theta} \nabla_{\theta}^T f(\theta) \right) \right)$, where ∇_{θ} is the gradient with respect to θ given by

$$\nabla_{\theta} = \left[\frac{\partial f(\theta)}{\partial \theta_1}, \frac{\partial f(\theta)}{\partial \theta_2}, \dots \right]^T.$$

The main bottleneck of this algorithm is the slow convergence and high computational cost involved in the solving of non-linear equations. It is possible to break down the set of parameters, θ into individual components θ_i , where i represents a unique direction

$$\tilde{\mathbf{v}}_{pq} = \sum_{s=1}^K \mathbf{s}_{pqs}(\theta_i) + \mathbf{n}_{pq_i}. \quad (2.27)$$

Eq. 2.27 allows the use of the expectation minimisation (EM) and the space alternating generalized expectation maximization (SAGE), discussed in ([Yatawatta et al. \(2012\)](#); [Kazemi et al. \(2011, 2013\)](#)). EM and SAGE are relatively faster, computationally cheaper and more accurate than Levenberg-Marquadt (LM). However, they involve partitioning of θ according to source properties, which is a quite a challenge, especially for large sky models.

The statistically efficient and fast calibration (StEFCal) algorithm, derived by [Salvini and Wijnholds \(2014\)](#) is a linear optimisation approach to improving convergence at a relatively low computational cost.

2.3.2 Statistically efficient and fast calibration

StEFCal attempts to optimise the non-linear problem in Eq.2.21 by linearising the system. The antenna gain g_q^* associated with antenna q is first kept constant and the antenna gain g_p associated with antenna p is solved for. Similarly, keeping g_p fixed, g_q^* is estimated. The antenna gains are therefore calculated using

$$\hat{g}_p^{(i)} = \sum_k \sum_{p,q} w_{pq}(t_k) |V_{pq} - g_p^{(i-1)}(t_k) V_{pq}^{mod} g_q^*(t_k)|^2,$$

where $g_p^{(i-1)}$ is the gain estimate from the previous iteration. The initial value of $g_p^{(0)}$ is assumed to be unity.

Unlike LM, the gain solutions are loaded only at the last iteration, which reduces the data processing time. Typically, for n antennas, the model scales to a complexity of $O(n^2)$. The LM solver increases the complexity to $O(n^3)$ while StEFCal preserves the order. Storage of data in the random-access memory (RAM) is no longer an issue as there is no need for derivatives.

The calibration methods discussed in the above sections attempt to solve for the direction-independent gain solutions. Though direction-dependent gains are tedious to solve for, they cannot be ignored, especially for high dynamic range (DR) fields. The proposed methods to fight against these DDEs are discussed in the section below.

2.4 Differential Gains

All telescopes are subjected to mispointing errors caused by gravitational load, thermal expansion or wind pressure. These errors can be denoted by the direction-dependent factor \mathbf{E}_p , in the RIME in Eq. 2.9

$$\mathbf{E}_p(l, m) = \mathbf{E}_p(l + \delta l, m + \delta m).$$

The pointing offsets, δl and δm , can be solved in the antenna-based gains fashion by modifying the selfcal algorithm, which was proposed by Bhatnagar et al. (2004). The modified selfcal is known as pointing selfcal (Bhatnagar et al. (2004); Smirnov (2011c)). Pointing selfcal finds the true value of the offset, however, it corrects only for pointing errors. Another way to correct for DDEs is *peeling*, proposed by Noordam (2004). The peeling algorithm absorbs all the DDEs in $\mathbf{E}_p(l, m)$ and solves for them by subtracting the sources in order of their decreasing brightness as follows:

1. Perform traditional selfcal. The brightest source s_0 will tend to absorb the \mathbf{G}_p solutions.
2. Subtract the predicted visibilities of s_0 from the data,

$$\mathbf{D}_{pq}^{(1)} = \mathbf{D}_{pq} - \mathbf{G}_p \mathbf{C}_{s_0 pq} \mathbf{G}_q^H,$$

where $\mathbf{C}_{s_0 pq}$ represents the coherency matrix associated with s_0 .

3. Solve for the DDEs associated with s_0 .
4. Phase-shift $\mathbf{D}_{pq}^{(1)}$ visibilities to the next brightest source s_1 .
5. Average the visibilities over time and frequency to cut down contributions from other sources.
6. Repeat step 1- 5 with s_1 .

The main disadvantage of peeling is the extensive computational cost. Also, at step. 1, the gain solutions are contaminated (Smirnov (2011c)). The approach of differential gains ΔE , introduced in Smirnov (2011c) accounts for selfcal contamination by solving for the DDEs and DDEs simultaneously, on small and large time/frequency scales respectively,

$$\mathbf{V}_{pq} = \mathbf{G}_p \sum_s \left(\Delta \mathbf{E}_{sp} \mathbf{K}_{sp} \mathbf{C}_s \mathbf{K}_{sq}^H \Delta \mathbf{E}_{sq}^H \right) \mathbf{G}_q^H + \mathbf{N}_{pq}; \quad \mathbf{N}_{pq} \in \mathbb{C}^{2 \times 2}. \quad (2.28)$$

The differential gains $\Delta \mathbf{E}$ are normally solved for a subset of fainter sources, allowing for the degrees of freedom to proliferate. This issue can be partially resolved by using large solution intervals for the differential gains.

The breakthrough in calibration approaches has tremendously aided radio astronomy over the past years. Calibrating for the antenna-based gains began by using a unit Jy calibrator in the form of ALS (Boonstra and van der Veen (2003)). Then came the iterative selfcal (Pearson and Readhead (1984); Boonstra and van der Veen (2003); Wijnholds and Van der Veen (2009); van der Tol et al. (2007); Intema et al. (2009)), which uses an initial model and updates it at each iteration. The optimisations used to solve for the antenna gains are LM (Levenberg (1944); Marquardt (1963)) and StEFCal (Salvini and Wijnholds (2014)). As for the DDEs, the differential gain approach seems to dominate over pointing selfcal and peeling (Smirnov (2011c)).

In general, the calibration algorithms are designed based upon the array designs as discussed below.

2.5 The Different Calibration Regimes

This section discusses four different calibration regimes based on spatial scales. The spatial scales include the size A of the array, the ionospheric fluctuations S and the projected FoV, V . A schematic view of the four regimes (Lonsdale (2005)) is illustrated in Fig. 2.2.

Regime 1: Small FoV with short baselines

The interferometric array in Fig. 2.2a resembles a single-dish array, owing to its compact structure. Thus, it abides by the principles of a single-dish telescope. Both antennas encounter the same ionospheric fluctuations, which implies that direction-independent calibration is sufficient. This regime only holds for very small arrays like KAT-7.

Regime 2: Small FoV with long baselines

The antennas are placed far apart, as illustrated in Fig. 2.2b. Different propagation effects or ionospheric variations are therefore associated with the antennas, but the lines of sight within the same FoV are subjected to the same ionosphere.

Regime 3: Large FoV with short baselines

The line of sight faces the same propagation effects at the stations as shown in Fig 2.2c. However, the ionospheric fluctuations are not constant throughout the FoV, each source samples a different ionosphere. The non-varying instrumental error can be handled easily. The Murchison Widefield Array (MWA) obeys this regime.

Regime 4: Large FoV with long baselines

Fig. 2.2c displays the general array configuration, prone to both DIEs and DDEs. The ionosphere depends on the antenna positions and the viewing direction, which brings about severe deformations in the source structure. The Atacama Large Millimeter Array (ALMA) and Low Frequency Array (LOFAR) fall under this regime, and it will also be applied to the SKA.

2.6 Conclusion

Calibration is the correction of the errors associated with the interferometric response, also regarded as minimisation between the observed visibilities and the predicted visibilities. The calibration approach to be used depends on the array layout and its corresponding FoV. Most of the calibration processes deal with non-linear optimisation, which is the main impediment since it slows down convergence, thus leading to high

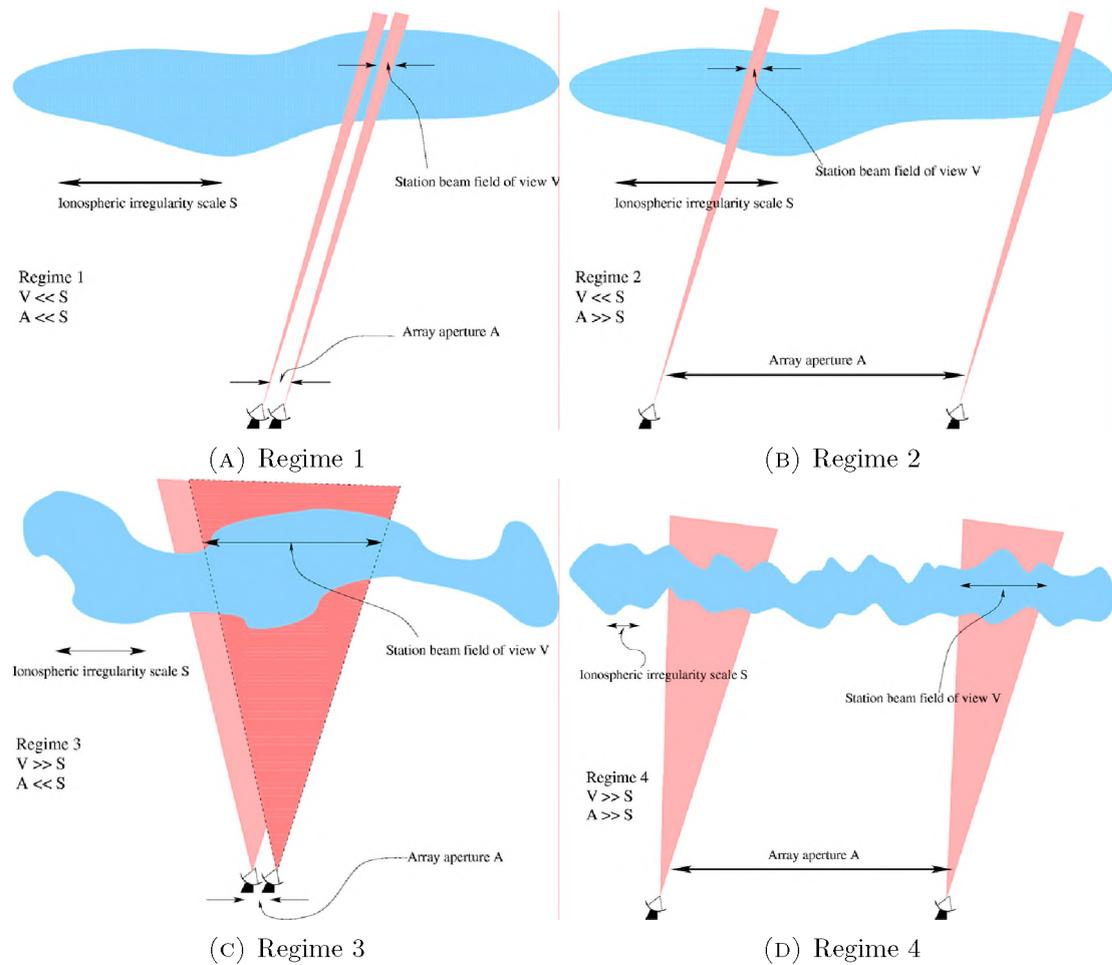


FIGURE 2.2: Array regimes with S representing the ionospheric irregularity scale, V representing the field of view and A representing the aperture size. (Credit Lonsdale)

computational cost. StEFCal is the most eminent for the time being, since it uses a linearised approach to optimising the solutions, while differential gains account for the DDEs.

2.7 Summary

This chapter gave an overview of some common calibration techniques. These calibration methods deal with the solving of the DIEs and DDEs. Lastly, it describes four calibration regimes.

Chapter 3

Ghost Phenomena in WRST Interferometric data

“Astronomy compels the soul to look upwards and leads us from this world to another”

–Plato

Calibration is one of the major steps involved in interferometric data processing. However, naive calibration leads to distortions in the resulting maps. These distortions are generally termed calibration artefacts, manifesting themselves in the form of misleading or spurious sources as well as loss of real emission. This chapter focuses on the underlying cause of the puzzling behaviour of the fainter misleading sources classically termed ghosts, which were mentioned in Section 1.3. The advancement in sensitive astronomical tools necessitates a theoretical understanding of these calibration artefacts.

3.1 Experimental Setup

From Chapter 2, we learnt that the selfcal algorithm requires a nominal (true) sky and a calibration sky model. We therefore chose a nominal sky that has two unpolarised point sources. The sources are referred to as the primary (the dominator) and secondary (the contaminator) source with fluxes A_p Jy and A_s Jy respectively. The primary source is located at the phase centre ($l = 0^\circ, m = 0^\circ$) while the secondary source is located at ($l = l_0^\circ, m = m_0^\circ$); ($l_0^\circ, m_0^\circ \neq (0^\circ, 0^\circ)$). The initial calibration model was taken to be a single point source of 1 Jy placed at the phase centre. The true sky and calibration sky model can be mathematically described as

$$I_{\mathcal{R}}(l, m) = A_p \delta(l, m) + A_s \delta(l - l_0, m - m_0)$$

$$I_{\mathcal{M}}(l, m) = A_p \delta(l, m),$$

respectively. Without any loss of generality, the skies can be normalised such that

$$I_{\tilde{\mathcal{R}}}(l, m) = A_1 \delta(l, m) + A_2 \delta(l - l_0, m - m_0)$$

$$I_{\tilde{\mathcal{M}}}(l, m) = A_1 \delta(l, m),$$

where $A_1 = 1$ Jy and $A_2 = \frac{A_s}{A_p}$ Jy. The visibilities corresponding to the skies can be obtained by a forward Fourier transform, $(l, m) \mapsto (u, v)$. The observed visibilities, $r(u, v)$, were therefore given by

$$\begin{aligned} r(u, v) &= \mathcal{F}_m \{ \mathcal{F}_l \{ A_1 \delta(l, m) + A_2 \delta(l - l_0, m - m_0) \} \} \\ &= A_1 + A_2 e^{-2\pi i (ul_0 + vm_0)} \\ &= A_1 + A_2 e^{-2\pi i \mathbf{u} \cdot \mathbf{s}_0}, \end{aligned} \quad (3.1)$$

where $\mathbf{u} = (u, v)$ and $\mathbf{s}_0 = (l_0, m_0)$. The visibilities produced by the calibration sky model are equal to

$$\begin{aligned} m(u, v) &= \mathcal{F}_m \{ \mathcal{F}_l \{ A_1 \delta(l, m) \} \} \\ &= A_1. \end{aligned} \quad (3.2)$$

The scenario considered suitable for our analysis is illustrated in Fig. 3.1, where $A_p = 1$

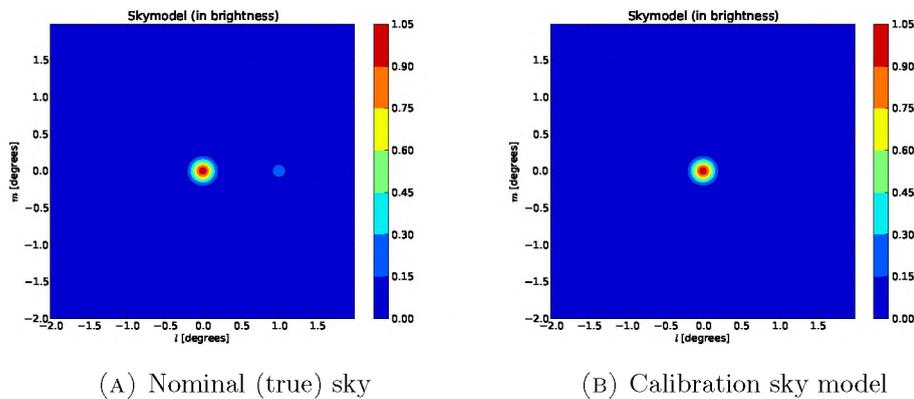


FIGURE 3.1: Two-source scenario experimental setup

Jy and $A_s = 0.2$ Jy. Normalising these fluxes give $A_1 = 1$ Jy and $A_2 = \frac{0.2}{1} = 0.2$ Jy. The secondary source was placed one degree away from the phase centre, that is, $l_0 = 1^\circ$ and $m_0 = 0^\circ$. The substitution of the flux and position values into Eq. 3.1 and Eq. 3.2

results in

$$r(u, v) = 1 + 0.2e^{-2\pi i \mathbf{b} \cdot \mathbf{s}_0} \quad (3.3)$$

$$m(u, v) = 1. \quad (3.4)$$

The scenario described above is used throughout the investigation unless stated otherwise.

3.2 Configuration of WSRT

It is essential to familiarise ourselves with the layout of WSRT. A traditional configuration (36, 108, 1332, 1404 m), illustrated in Fig. 3.2, was considered throughout this study. WSRT is a regular EW array with high redundancy. The regularity of the array allowed us to establish an analytical relationship between the antennas.

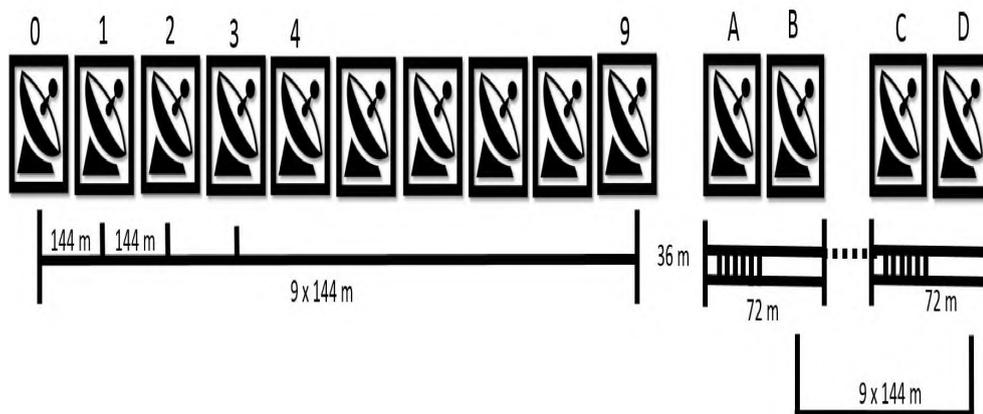


FIGURE 3.2: WSRT Configuration

Consider $\{\mathbf{u}_p\}$ to be the set of antenna positions. The set $\{\mathbf{u}_p\}$ is said to be regularly spaced if a *common quotient baseline* (CQB) \mathbf{b}_0 exists such that $\mathbf{u}_p = \phi_p \mathbf{b}_0$, where ϕ_p is a positive integer. Thus, for the WSRT configuration in Fig. 3.2 the coordinates u and v for each baseline can be represented as integer multiples of \mathbf{b}_0 . For instance, consider three WSRT antennas RT0, RT3 and RT5 and a frequency of 1460 MHz. The CQB \mathbf{b}_0 is indicated by the red dashed line in Fig. 3.3. Using this imaginary track, the uv coordinates for baselines RT3-RT5 (blue line), RT0-RT3 (green line) and RT0-RT5 (yellow line) are calculated. The CQB \mathbf{b}_0 is a function of time due to the rotation of the earth and it is chosen such that the greatest common divisor (gcd) of all the baselines is equal to one. Note that the actual visibility data are not altered by using the linear dependence substitution ϕ_p .

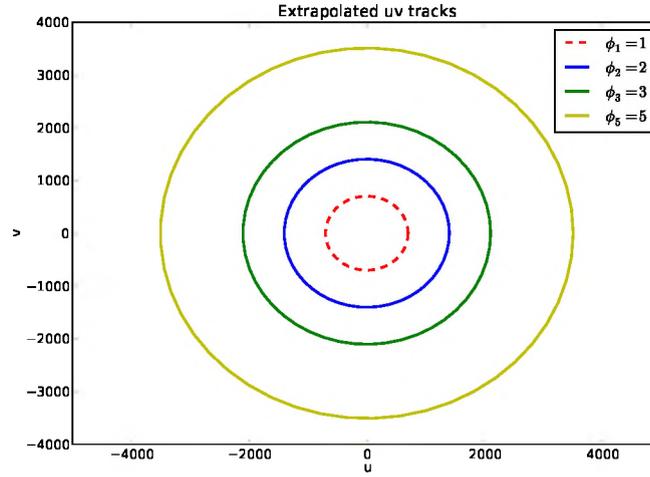


FIGURE 3.3: Extrapolated uv tracks. The imaginary track, \mathbf{b}_0 is indicated by the red dashed lines. The real tracks are represented by the lines.

The linear dependency between any two antennas p and q can be stacked into a single $n \times n$ matrix

$$\mathbf{\Phi} = \begin{pmatrix} \phi_{00} & \phi_{01} & \cdots & \phi_{0n} \\ \phi_{10} & \phi_{11} & \cdots & \phi_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n0} & \phi_{n1} & \cdots & \phi_{nn} \end{pmatrix}, \quad (3.5)$$

with entries $\phi_{pq} = \phi_q - \phi_p$ and n denoting the number of antennas. The $\mathbf{\Phi}$ for the WSRT configuration in Fig. 3.2 is equal to

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 9.25 & 9.75 & 18.25 & 18.75 \\ -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 8.25 & 8.75 & 17.25 & 17.75 \\ -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 7.25 & 7.75 & 16.25 & 16.75 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 6.25 & 6.75 & 15.25 & 15.75 \\ -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 5.25 & 5.75 & 14.25 & 14.75 \\ -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 4.25 & 4.75 & 13.25 & 13.75 \\ -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 3.25 & 3.75 & 12.25 & 12.75 \\ -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 2.25 & 2.75 & 11.25 & 11.75 \\ -8 & -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 1 & 1.25 & 1.75 & 10.25 & 10.75 \\ -9 & -8 & -7 & -6 & -5 & -4 & -3 & -2 & -1 & 0 & 0.25 & 0.75 & 9.25 & 9.75 \\ -9.25 & -8.25 & -7.25 & -6.25 & -5.25 & -4.25 & -3.25 & -2.25 & -1.25 & 0.25 & 0 & 0.5 & 9 & 9.5 \\ -9.75 & -8.75 & -7.75 & -6.75 & -5.75 & -4.75 & -3.75 & -2.75 & -1.75 & -0.75 & -0.5 & 0 & 8.5 & 9 \\ -18.25 & -17.25 & -16.25 & -15.25 & -14.25 & -13.25 & -12.25 & -11.25 & -10.25 & -9.25 & -9 & -8.5 & 0 & 0.5 \\ -18.75 & -17.75 & -16.75 & -15.75 & -14.75 & -13.75 & -12.75 & -11.75 & -10.75 & -9.75 & -9.5 & -9 & 0.5 & 0 \end{pmatrix} \times 4 \quad (3.6)$$

The scalar factor can be removed if $\mathbf{b}_0 = 36m$ instead. Here, we made use of the old configuration in order to be consistent with the configuration used for the simulations

3.3 Theoretical Analysis

This section sheds light on the calibration artefacts buried in thermal noise using a mathematical ALS framework. This framework guides us to the underlying cause of the ghost sources whilst predicting the spatial ghost distribution. In the past, inadequate attention has been paid to the ghost sources because of their low intensities. However, the advent of more sensitive radio telescopes encourages investigation of this phenomenon. The following subsections proceed mathematically through the different stages of interferometric data processing.

3.3.1 The extrapolated visibility matrix

Consider a perfect regularly spaced array with unity gains and integration intervals sufficiently small for smearing to be negligible. The per-baseline visibilities observed at each time step at coordinates $\mathbf{u}_{pq} = \phi_{pq}\mathbf{b}_0$ are given by

$$\begin{aligned} V_{pq}^{\mathcal{R}} &= r_{pq}(\mathbf{u}_{pq}) = r(\phi_{pq}\mathbf{b}_0) \\ &= A_1 + A_2 e^{-2\pi i \phi_{pq}\mathbf{b}_0 \cdot \mathbf{s}_0}, \end{aligned} \quad (3.7)$$

with $\mathbf{b}_0 \cdot \mathbf{s}_0$ denoting a dot product and $\mathbf{b}_0 = \mathbf{b}_0(t)$ denoting the CQB. The notation $V_{pq}^{\mathcal{A}} = a_{pq}(\mathbf{b})$ is used interchangeably throughout to represent the visibility data. For any baseline pq , the corresponding visibility data are given by $r(\phi_{pq}\mathbf{b}_0)$. Since $\phi_{pq}\mathbf{b}_0(t)$ follows an elliptic path, as shown in Fig. 3.3, the measurements subjected to calibration are restricted within an elliptical track. However, by introducing a free variable \mathbf{b}

$$V_{pq}^{\mathcal{R}} = r_{pq}(\mathbf{b}) = r(\phi_{pq}\mathbf{b}), \quad (3.8)$$

we can extrapolate the visibilities over the entire uv -plane. Similar to the Φ matrix, the observed visibilities are stacked into a matrix

$$\mathcal{R}(\mathbf{b}) = \begin{pmatrix} V_{00}^{\mathcal{R}} & V_{01}^{\mathcal{R}} & \cdots & V_{0D}^{\mathcal{R}} \\ V_{10}^{\mathcal{R}} & V_{11}^{\mathcal{R}} & \cdots & V_{1D}^{\mathcal{R}} \\ \vdots & \vdots & \ddots & \vdots \\ V_{D0}^{\mathcal{R}} & V_{D1}^{\mathcal{R}} & \cdots & V_{DD}^{\mathcal{R}} \end{pmatrix}.$$

Further, Eq. 3.7 can be interpreted as a coordinate scaling relationship between $r(\mathbf{u})$ and $r_{pq}(\mathbf{b})$. The map yielded by $\mathcal{R}(\mathbf{b})$ is thus a scaled version of the actual map.

3.3.2 ALS calibration

ALS is considered ideal for deriving the analytical expression, which outlines the ghost sources. It treats the antenna gains and noise components separately. Noise-free data were assumed, since noise does not influence the results (Smirnov (2010)). Hence, ALS needs to solve only for the gain components. The gain vector $\mathbf{g}(\mathbf{b})$ was calculated by minimising

$$\|\mathcal{R}(\mathbf{b}) - \mathbf{G}(\mathbf{b})\mathcal{M}\mathbf{G}(\mathbf{b})^H\|_F, \quad (3.9)$$

where $\|\cdot\|_F$ represents the Frobenius norm and $\mathbf{G}(\mathbf{b}) = \text{diag}(\mathbf{g}) = \text{diag}([g_1 \ g_2 \ \dots \ g_n]^T)$. The unpolarised extrapolated visibility matrix $\mathcal{R}(\mathbf{b})$ is of rank two, and \mathcal{M} is the visibility matrix generated from the calibration sky model with entries m_{pq} . Eq. 3.9 can be reformulated as

$$\|\mathcal{R}(\mathbf{b}) - \mathbf{g}(\mathbf{b})\mathbf{g}(\mathbf{b})^H \odot \mathcal{M}\| = \|\mathcal{R}(\mathbf{b}) - \mathcal{G}(\mathbf{b})^H \odot \mathcal{M}\|, \quad (3.10)$$

where \odot denotes element-wise matrix multiplication and $\mathcal{G}(\mathbf{b}) = \mathbf{g}(\mathbf{b})\mathbf{g}(\mathbf{b})^H$ is the product of the gain solutions with its own Hermitian transpose, that is,

$$\mathcal{G}(\mathbf{b}) = \begin{pmatrix} g_1 g_1^* & g_1 g_2^* & \cdots & g_1 g_n^* \\ g_2 g_1^* & g_2 g_2^* & \cdots & g_2 g_n^* \\ \vdots & \vdots & \vdots & \vdots \\ g_n g_1^* & g_n g_2^* & \cdots & g_n g_n^* \end{pmatrix}. \quad (3.11)$$

Without any loss of generality, \mathcal{M} was assumed to be unity, that is, all elements of \mathcal{M} is one. Eq. 3.10 can thence be restated as

$$\|\mathcal{R}(\mathbf{b}) - \mathbf{g}(\mathbf{b})\mathbf{g}(\mathbf{b})^H\| = \|\mathcal{R}(\mathbf{b}) - \mathcal{G}(\mathbf{b})\|. \quad (3.12)$$

The square matrix $\mathcal{G}(\mathbf{b})$ was decomposed into

$$\mathcal{G}(\mathbf{b}) = \mathbf{x}(\mathbf{b})\boldsymbol{\lambda}(\mathbf{b})\mathbf{x}(\mathbf{b})^H, \quad (3.13)$$

where $\lambda(\mathbf{b})$ is the largest eigenvalue of $\mathcal{R}(\mathbf{b})$ and $\mathbf{x}(\mathbf{b})$ is the associated normalised eigenvector. Define $\mathcal{G}(\mathbf{b})^\top$ as

$$\mathcal{G}(\mathbf{b})^\top = \begin{pmatrix} \frac{1}{g_1 g_1^*} & \frac{1}{g_1 g_2^*} & \cdots & \frac{1}{g_1 g_n^*} \\ \frac{1}{g_2 g_1^*} & \frac{1}{g_2 g_2^*} & \cdots & \frac{1}{g_2 g_n^*} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{g_n g_1^*} & \frac{1}{g_n g_2^*} & \cdots & \frac{1}{g_n g_n^*} \end{pmatrix}. \quad (3.14)$$

Some essential propositions are listed in the next subsection that will be used later on.

3.3.2.1 Essential propositions

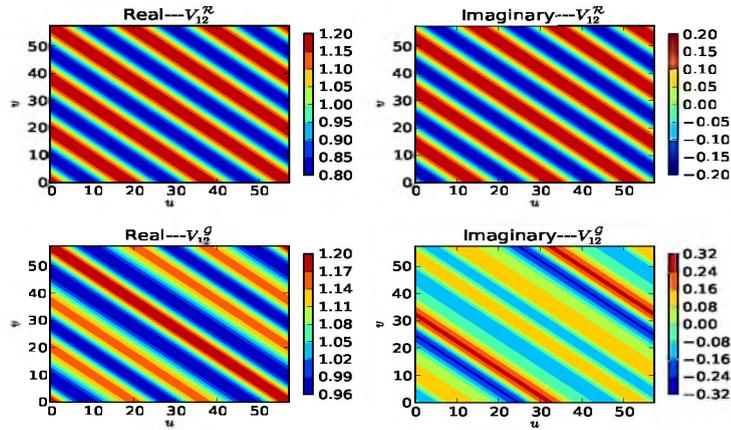
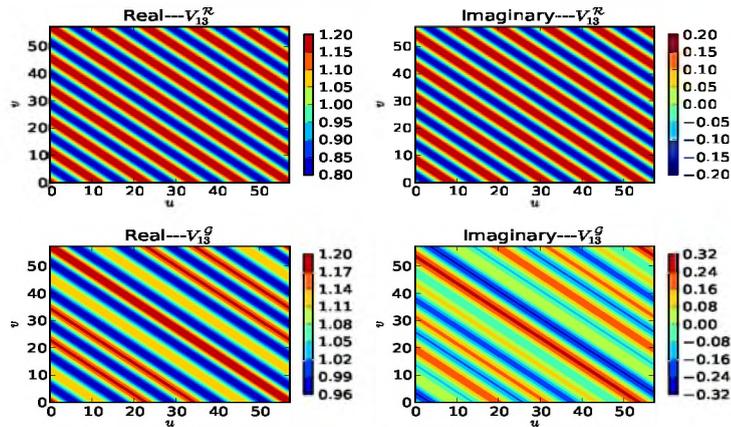
The matrices $\mathcal{R}(\mathbf{b})$, $\mathcal{G}(\mathbf{b})$, and $\mathcal{G}(\mathbf{b})^\top$ defined in Section 3.3.2 obey some properties that are significant for the derivation of ghosts. The propositions listed below are proven in Appendix A.

- (i) Proposition A.4: The rank of $\mathcal{R}(\mathbf{b})$ is two. This proposition quantifies the amount of information being lost during the computation of $\mathcal{G}(\mathbf{b})$.
- (ii) Proposition A.5: The elements of the function-valued matrix $\mathcal{G}(\mathbf{b})$ are periodic, effectively one-dimensional, differentiable, Hermitian functions.
- (iii) Proposition A.6: It follows from Proposition A.5 that the elements of $\mathcal{G}(\mathbf{b})$ can be effectively expressed as a one-dimensional Fourier series (which ultimately leads to the formation of ghosts).
- (iv) Proposition A.7: Since the elements of $\mathcal{G}(\mathbf{b})^\top$ are periodic, effectively one-dimensional, differentiable, Hermitian functions, they can be written as a one-dimensional Fourier expansion according to Proposition A.6.

3.3.2.2 ALS calibration matrix

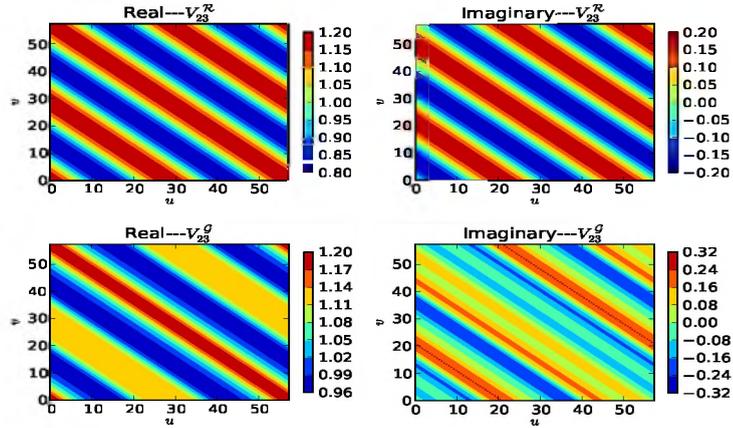
As a preliminary step towards deriving the ALS calibration matrix, three WSRT antennas RT0, RT3 and RT5, were considered. From the WSRT scaling matrix Φ in Eq. 3.6, the subset associated with the three-element interferometer is given by

$$\Phi_{035} = \begin{pmatrix} 0 & 3 & 5 \\ -3 & 0 & 2 \\ -5 & -2 & 0 \end{pmatrix}. \quad (3.15)$$

FIGURE 3.4: The functions $r_{12}(\mathbf{b})$ and $g_{12}(\mathbf{b})$.FIGURE 3.5: The functions $r_{13}(\mathbf{b})$ and $g_{13}(\mathbf{b})$.

The 0.2 Jy contaminator was placed at $l_0 = 1^\circ$ and $m_0 = 1^\circ$. The graphical representation of the resulting $\mathcal{R}(\mathbf{b})$ and $\mathcal{G}(\mathbf{b})$ matrices can be seen in Fig. 3.4, Fig. 3.5 and Fig. 3.6. The functions $r_{pq}(\mathbf{b})$ are continuous, differentiable, Hermitian and periodic in the u and v directions with periods $\frac{1}{\phi_{pq}l_0}$ and $\frac{1}{\phi_{pq}m_0}$ respectively. Moreover, they are consistent along each line, $v = -\frac{l_0}{m_0}u + c$ for any $c \in \mathbb{R}$. The $\mathcal{G}(\mathbf{b})$ matrix is also continuous, differentiable, Hermitian and periodic with a periodicity of $\frac{1}{l_0}$ and $\frac{1}{m_0}$ along u and v respectively. The secondary harmonic, determined by the linear dependence ϕ_{12}, ϕ_{15} and ϕ_{23} portrayed in Eq. 3.15, differs from baseline to baseline. This results in different per-baseline calibrated visibilities $g_{pq}(\mathbf{b})$. Since $g_{pq}(\mathbf{b})$ is a one-dimensional Hermitian function, it can be written as a Fourier series, as stated in Proposition A.6.

$$g_{pq}(\mathbf{b}) = \sum_{j=-\infty}^{\infty} c_{j,pq} e^{-2\pi i j \mathbf{b} \cdot \mathbf{s}_0}. \quad (3.16)$$

FIGURE 3.6: The functions $r_{23}(\mathbf{b})$ and $g_{23}(\mathbf{b})$.

The real Fourier coefficients, $c_{j,pq}$, are calculated using the following integral:

$$c_{j,pq} = |l_0||m_0| \int_{-\frac{1}{2|m_0|}}^{\frac{1}{2|m_0|}} \int_{-\frac{1}{2|l_0|}}^{\frac{1}{2|l_0|}} g_{pq}(u, v) e^{2\pi i j (ul_0 + vm_0)} du dv. \quad (3.17)$$

Eq. 3.17 is formally derived in Proposition A.6 found in Appendix A. Note that constraining the infinite sum in Eq. 3.16 yields better solutions. The next question that arose was what the sky distribution is that corresponds to the calibrated visibilities $g_{pq}(\mathbf{b})$. The Fourier inversion of $g_{pq}(\mathbf{b})$ gives

$$\mathcal{F}^{-1}\{g_{pq}\}(\mathbf{s}) = \sum_{j=-\infty}^{\infty} c_{j,pq}^{\mathcal{G}} \delta(\mathbf{s} - j\mathbf{s}_0), \quad (3.18)$$

which is a summation of delta functions located at integer multiples of \mathbf{s}_0 .

$$\begin{aligned} I_{pq}^{\mathcal{G}}(\mathbf{s}) &= \mathcal{F}^{-1}\left\{g_{pq}\left(\frac{\mathbf{b}}{\phi_{pq}}\right)\right\} \\ &= \sum_{j=-\infty}^{\infty} c_{j,pq}^{\mathcal{G}} \delta\left(\mathbf{s} - \frac{j\mathbf{s}_0}{\phi_{pq}}\right). \end{aligned} \quad (3.19)$$

The sky distribution $I_{pq}^{\mathcal{G}}$ physically represents the sky distribution whose Fourier transform gives the predicted visibilities along the uv track given by $\phi_{pq}\mathbf{b}_0(t)$. It should be noted how the scaling factor in Eq. 3.8 influences the resulting sky distribution in Eq. 3.19. After applying the best-fitting gains, the sky distribution produced by the predicted calibrated visibilities is consistent with a sky of delta functions spaced at intervals of $\frac{\mathbf{s}_0}{\phi_{pq}}$, with amplitudes given by $\{c_{j,pq}\}$. These delta functions are actually the fundamental constituents of the ghost sources observed in Fig.1.3. This explanation is broadened in Section 3.4.

3.4 Ghost Formation

The Fourier series described in Eq. 3.19 indicates that the corrected visibilities $\mathcal{R}^{(c)}$

$$\mathcal{R}^{(c)} = \mathcal{G}^{-1} \mathcal{R} \mathcal{G}^{-H} = \mathcal{G}^{\top} \odot \mathcal{R} \quad (3.20)$$

contain ghost sources. The gain matrix \mathcal{G} , containing the per-baseline ghost patterns, was found to be of rank one. Since \mathcal{G} is of lower rank than \mathcal{R} , some information was lost during its computation. The de-ranking of \mathcal{G} brings about drastic alterations in the Fourier characteristics of the original \mathcal{R} matrix. This signifies that the optimisation approach (LS fit) involved in the evaluation of \mathcal{G} is the main culprit in the ghost formation. The ghost sources are often buried in the noise or masked by the sidelobes of the real emission, which calls for distilled radio maps.

3.4.1 Distillation

The residual visibilities can be calculated using

$$\mathcal{R}^{\Delta} = \mathcal{R}^{(c)} - \mathcal{R}, \quad (3.21)$$

which cancels out most of the actual sources and the noise terms (if any). The images thus yield the “distilled ghost” sources. In practice, the ghost sources would not be clearly apparent in the residuals \mathcal{R}^{Δ} , since the actual sources would not cancel out, unlike our two-source simulation whereby a perfect interferometer was assumed.

Returning to our two-source scenario experiment, the extrapolated visibility matrix \mathcal{R} can be represented by

$$\mathcal{R} = A_1 \mathbf{1} + A_2 \mathbf{K},$$

where \mathbf{K} is the Fourier kernel matrix of complex phase terms corresponding to the offset of the secondary source. The residual visibilities \mathcal{R}^{Δ} can now be reformulated as

$$\begin{aligned} \mathcal{R}^{\Delta} &= A_1 \mathbf{G}^{-1} \mathbf{1} \mathbf{G}^{-1} + A_2 \mathbf{G}^{-1} \mathbf{K} \mathbf{G}^{-1} - A_1 \mathbf{1} - A_2 \mathbf{K} \\ &= A_1 (\mathcal{G}^{\top} - 1) + A_2 (\mathcal{G}^{\top} - 1) \odot \mathbf{K}. \end{aligned} \quad (3.22)$$

The residual matrix \mathcal{R}^{Δ} can be looked upon as the convolution of $\mathcal{G}^{\top} - 1$ with the true sky. The matrix \mathcal{G}^{\top} has the same ghost patterns, but different ghost amplitudes. When the flux of the primary source $A_1 = 1$ Jy is subtracted from \mathcal{G}^{\top} , the resulting visibilities $\mathcal{G}^{\top} - 1$ are known as the distilled visibilities, which produce the *distilled ghost patterns*. Eq. 3.22 reveals two ghost patterns; the first realisation corresponds to the

dominator and the second realisation corresponds to the contaminator, scaled by A_1 and A_2 respectively. The second ghost pattern is shifted to the position of the contaminator. Since $A_1 \gg A_2$, the ghost pattern associated with the primary source dominates over the other one. In the case of WSRT, the positions of the ghost sources in the two patterns coincide with each other, as will be discussed in the next subsection.

3.4.2 Imaging

Given that an interferometer samples the spatial Fourier transform of the sky intensity, the per-baseline observed visibilities can be mathematically expressed as

$$V_{pq} = S_{pq} \mathcal{F}\{I_{pq}\}, \quad (3.23)$$

where S_{pq} is the sampling function of baseline pq . Traditional interferometric imaging involves the summation of all per-baseline visibilities onto a “common” uv -plane using convolutional gridding. The visibilities are then Fourier transformed back to a sky distribution estimate, known as the “dirty image”

$$\begin{aligned} I_D &= \mathcal{F}^{-1}\left\{\sum_{pq} S_{pq} \mathcal{F}\{I_{pq}\}\right\} \\ &= \sum_{pq} P_{pq} \circ I_{pq}, \end{aligned} \quad (3.24)$$

where ‘ \circ ’ denotes convolution and $P_{pq} = \mathcal{F}^{-1}\{S_{pq}\}$ is the unnormalised PSF associated with baseline pq . In a situation where all baselines view a common sky, Eq. 3.24 becomes

$$I_D = \left(\sum_{pq} P_{pq}\right) \circ I, \quad (3.25)$$

where $P = \sum_{pq} P_{pq}$ corresponds to the PSF of the array. Letting the least common multiple of all ϕ_{pq} ’s be ϕ_0 , Eq. 3.19, which represents a string of delta functions spaced at intervals of $\frac{\phi_0}{\phi_{pq}}$, can be reformulated as

$$I_{pq}^{\mathcal{G}}(\mathbf{s}) = \sum_{k=-\infty}^{\infty} d_{k,pq}^{\mathcal{G}} \left(\mathbf{s} - \frac{k\mathbf{s}_0}{\phi_{pq}}\right), \quad (3.26)$$

where

$$d_{k,pq}^{\mathcal{G}} = \begin{cases} c_{j,pq}^{\mathcal{G}} & \text{if } \exists \text{ an } j \in \mathbb{Z} \text{ such that } k\phi_{pq} = j\phi_0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.27)$$

Substituting Eq. 3.24 in the above gives

$$\begin{aligned} I_D^{\mathcal{G}} &= \sum_{pq} P_{pq} \circ \left(\sum_{k=-\infty}^{\infty} d_{k,pq}^{\mathcal{G}} \delta_k \right) \\ &= \sum_{k=-\infty}^{\infty} \left(\sum_{pq} d_{k,pq}^{\mathcal{G}} P_{pq} \right) \circ \delta_k, \end{aligned} \quad (3.28)$$

with $\delta_k = \left(\mathbf{s} - \frac{k\mathbf{s}_0}{\phi_0} \right)$. The following deductions can be made from Eq. 3.28:

1. Each delta function spaced at regular intervals of $\frac{k\mathbf{s}_0}{\phi_0}$ is convolved with its own *ghost spread function* (GSF)

$$P_k^{\mathcal{G}} = \sum_{pq} d_{k,pq}^{\mathcal{G}} P_{pq}. \quad (3.29)$$

2. The GSF does certainly resemble the nominal PSF but they are not identical.
3. The ghost sources are formed at integer multiples of $\frac{\mathbf{s}_0}{\phi_0}$, that is the ghost positions depend on the array configuration and source positions.

These aforementioned observations provide a reasonably good explanation for the puzzling behaviour seen in Fig. 1.3.

3.4.3 Corrected visibilities

In practice, radio astronomers would image the corrected visibilities $\mathcal{R}^{(c)}$ given by Eq. 3.20. The corrected visibilities $\mathcal{R}^{(c)}$ were Fourier transformed back into an estimate of the corrected sky

$$I^{(c)} = \mathcal{F}^{-1}\{\mathcal{R}^{(c)}\}. \quad (3.30)$$

To study the effect of \mathcal{G} on $\mathcal{R}^{(c)}$, an intermediate step is required. We defined a “ \mathcal{G}^{\top} -sky” such that its Fourier transform is consistent with the visibility data given by g_{pq}^{-1} . Proposition A.7 states that $g_{pq}^{-1}(\mathbf{b})$ can be expressed as

$$g_{pq}^{-1}(\mathbf{b}) = \sum_{j=-\infty}^{\infty} c_{j,pq}^{\top} e^{2\pi i j \mathbf{b} \cdot \mathbf{s}_0}. \quad (3.31)$$

This implies that “ \mathcal{G}^{\top} ” will have a similar representation to Eq. 3.19,

$$I_{pq}^{\mathcal{G}^{\top}}(\mathbf{s}) = \sum_{k=-\infty}^{\infty} d_{k,pq}^{\top} \delta_{pq}, \quad (3.32)$$

with $d_{k,pq}^\top$ denoting the new set of Fourier coefficients. When the extrapolated visibilities are convolved with \mathcal{G}^\top , the components of the matrix $\mathcal{G}^\top \odot \mathcal{R}$ are denoted by

$$r_{pq}^{(c)} = g_{pq}^{-1} \cdot r_{pq}. \quad (3.33)$$

In accordance with Eq. 3.30, the corrected sky is attained by performing a Fourier inverse on Eq. 3.33

$$\begin{aligned} \mathcal{F}^{-1}\{r_{pq}^{(c)}\} &= \mathcal{F}^{-1}\{g_{pq}^{-1}\} \circ \mathcal{F}^{-1}\{r_{pq}\} \\ I_{pq}^{(c)} &= I_{pq}^{\mathcal{G}^\top} \circ I^{\mathcal{R}}. \end{aligned} \quad (3.34)$$

From the above equation, it is obvious that the real sky $I^{\mathcal{R}}$ is convolved with the string of delta functions specified in Eq. 3.31. These delta functions account for the puzzling behaviour observed in Fig. 1.3 and 1.4. The corresponding residual sky is equal to

$$I_{pq}^\Delta = I_{pq}^{\mathcal{G}^\top - 1} \circ I^{\mathcal{R}} = (I^{\mathcal{G}^\top} - \delta) \circ I^{\mathcal{R}}, \quad (3.35)$$

where $I_{pq}^{\mathcal{G}^\top - 1}$ denotes the per-baseline sky associated with the distilled ghost pattern $\mathcal{G}^\top - 1$. Using Eq. 3.28, the subsequent dirty image is given by

$$I_D^{\mathcal{G}^\top - 1} = \sum_{k=-\infty}^{\infty} \left(\sum_{pq} \hat{d}_{k,pq}^\top P_{pq} \right) \circ \delta_k. \quad (3.36)$$

where

$$\hat{d}_{k,pq}^\top = d_{k,pq}^\top - 1 \neq \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}. \quad (3.37)$$

In the case of our two-source sky model the corrected dirty image is a superposition of two patterns scaled by A_1 and A_2 . The ghost pattern related to A_2 shifts in position according to \mathbf{s}_0 . The ghost sources from the individual ghost pattern overlap with each other at specific positions, since they occur at discrete intervals of $\frac{s_0}{\phi_0}$. Consequently, the summation of those corresponding Fourier coefficients at those particular positions provides the amplitudes of those ghost sources. Using Eq. 3.28, the corrected sky can be described by

$$I_D^{(c)} = \sum_{k=-\infty}^{\infty} \left(\sum_{pq} A_1 d_{k,pq}^\top + A_2 d_{k-\phi_0,pq}^\top \right) P_{pq} \circ \delta_k. \quad (3.38)$$

With the aid of Eq. 3.35, we obtain the residual image by subtracting one from the coefficients $d_{k,pq}^\top$ defined in Eq. 3.37:

$$I_D^\Delta = \sum_{k=-\infty}^{\infty} \left(\sum_{pq} (A_1 \hat{d}_{k,pq}^\top + A_2 \hat{d}_{k-\phi_0,pq}^\top) P_{pq} \right) \circ \delta_k. \quad (3.39)$$

Another interesting quantity to study is the amplitude of the ghost sources as a fraction of the unmodelled flux A_2 . Consider the k -th ghost situated at $\frac{ks_0}{\phi_{pq}}$. Being convolved by the PSF, the amplitude of the ghost is given by

$$\sum_{pq} \hat{d}_{k,pq} P_{pq}(0), \quad (3.40)$$

where $P_{pq}(0)$ is the per-baseline weight and is ultimately determined via the imaging weights. For our analysis, natural weighting (Sramek and Schwab (1999)) is being considered, whereby the density weighting function for each baseline is taken to be unity. This weighting scheme is favourable for detecting faint point sources, as it optimises the root mean square error (RMSE) of the radio images. The ghost amplitudes in the distilled image are therefore given by

$$\zeta_k = \langle \hat{d}_{k,pq}^\top \rangle_{pq}, \quad (3.41)$$

where $\langle \cdot \rangle_{pq}$ denotes averaging over all baselines pq . Since the k -th ghost source in the residual image is a superposition of the two distilled ghost patterns, the residual ghost amplitudes are equal to

$$\zeta_k^\Delta = A_1 \zeta_k + A_2 \zeta_{k-\phi_0}. \quad (3.42)$$

In addition, the ratio $\frac{\zeta_k^\Delta}{A_2}$ can be identified as the amplitude of the ghost sitting on top of the unmodelled source. This quantity has been studied since the invention of traditional selfcal. The evaluated ratio gives the amount of flux that was lost during calibration. This loss in flux density is referred to as **source suppression** and is due to the same mechanism as the ghost formation. Source suppression is further investigated in Chapter 4.

3.5 Results and Discussion

A theoretical framework predicting the distribution and amplitudes of ghost sources for a two-source scenario was constructed in Section 3.4. This section predicts the ghost formation for a specific scenario and compares the results with simulations. The experimental sky settings are as follows: $l_0 = 1^\circ$, $m_0 = 0^\circ$, $A_1 = 1$ Jy and $A_2 = 0.2$

Jy unless specified otherwise. A monochromatic 12-hour synthesis at a frequency of 1.45 GHz is considered. Note that all the results in this section are generated from the traditional WSRT configuration (36, 108, 1332, 1404 m) illustrated in Fig. 3.2.

The visibility data obtained under the aforementioned assumptions are then converted into sky distributions. Three different approaches are used to predict the ghost distribution:

- The theoretical framework developed in Section 3.4.
- A customised implementation of ALS calibration explicitly explained in Section 3.3.2. The visibilities are imaged using a Fast Fourier Transform-based imager, *lwimager*, implemented in the Common Astronomy Software Application (CASA) libraries.
- The LS approach found in the MeqTrees Package (Noordam and Smirnov (2010)). The imaging tool is once again *lwimager*.

3.5.1 Ghost patterns

The theoretical distilled ghost distribution for a set of baselines (9A: 36 m, 01 and 12: 144 m, 05 and 16: 720 m, 0D: 2.7 km) are displayed in Fig. 3.7. The simulated distilled ghost distributions are obtained as follows:

- (i) ALS and LS are applied to a set of simulated visibilities using the MeqTrees Package.
- (ii) Unity (since the highest flux is 1 Jy) is subtracted from the calibrated visibilities. The distilled visibilities are then imaged for the set of baselines described above using *lwimager*.
- (iii) The fluxes at the ghost positions, illustrated in Fig. 3.10, are measured for the individual baseline. The flux measurements are displayed in Fig. 3.8 and 3.9.

Although Fig. 3.8 and 3.9 are qualitatively similar, slight variations in amplitudes can be observed, which is not at all surprising, as they are outcomes of two different optimisation approaches. One obvious difference is the absence of the ghost at 0° in Fig. 3.9. The centre ghost in Fig. 3.8 signifies that ALS suppresses the modelled source as well. The total flux in the sky and calibration model are stored as the diagonal terms in \mathcal{R} and \mathcal{M} , which correspond to the autocorrelated values. In the case of LS, the autocorrelations are ignored, thus giving the gains more freedom to fit the mean amplitude. When the autocorrelations are considered (as in ALS), the gain solutions must compensate for the

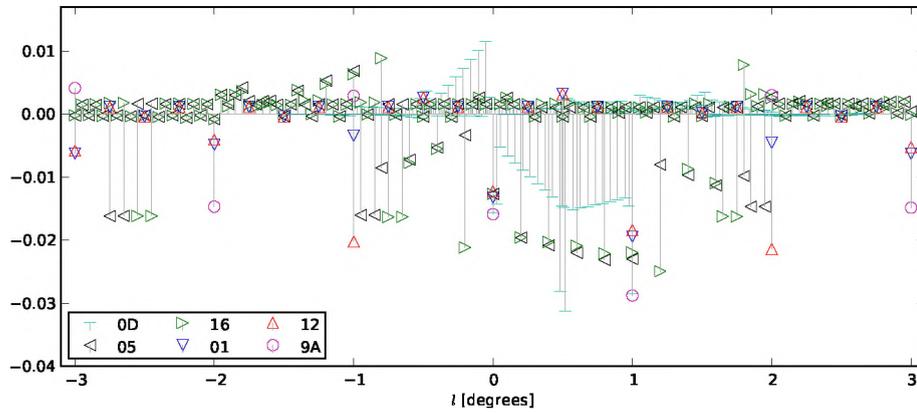


FIGURE 3.7: Theoretical ghost pattern for baselines 9A (36 m), 01 and 12 (144 m), 05 and 16 (720 m), 0D (2.7 km).

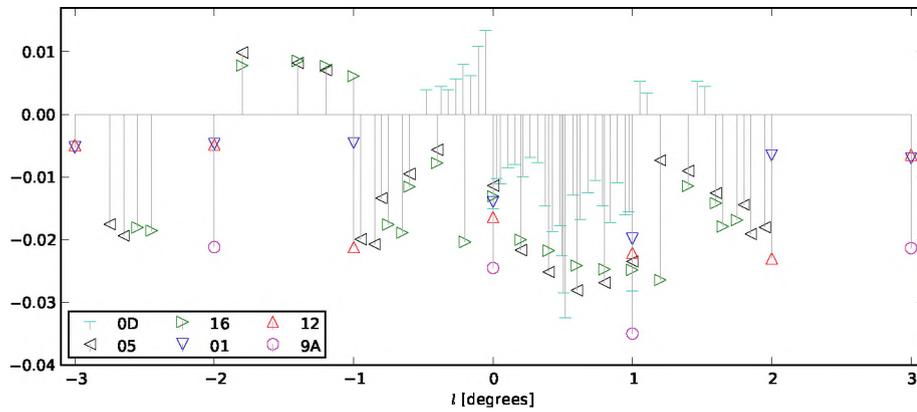


FIGURE 3.8: ALS ghost pattern for baselines 9A (36 m), 01 and 12 (144 m), 05 and 16 (720 m), 0D (2.7 km).

total flux ($A_1 + A_2$) using a model of flux A_1 . The restriction imposed on the gain solutions eventually results in mean amplitudes slightly above unity in \mathcal{G} and below unity in \mathcal{G}^\top . This explains the strong negative ghost at 0° in Fig. 3.8.

Comparing Fig. 3.7 to Fig. 3.8 and Fig. 3.9:

- (i) The ghost density in Fig. 3.7 is much higher than that in Fig.3.8 and 3.9. This can be explained by the following: the manual flux extraction was limited, as we were only able to select ghost sources visible to the naked eye.
- (ii) The ghost sources in Fig.3.8 and 3.9 coincide with the bright ones in Fig. 3.7.
- (iii) The amplitudes in Fig.3.8 and 3.9 are somewhat less than the theoretically predicted ones, since the dirty maps are dominated by sidelobes, which influence the flux measurements.

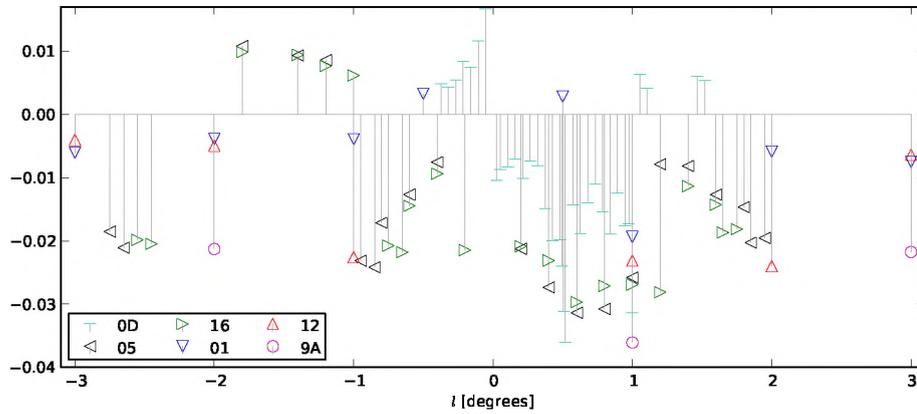


FIGURE 3.9: LS ghost pattern for baselines 9A (36 m), 01 and 12 (144 m), 05 and 16 (720 m), 0D (2.7 km).

All WSRT antennas were then considered. The ghost patterns resulting from ALS and LS calibration can be seen in Fig. 3.11. Fig. 3.11 exhibits both “inner” and “outer” ghosts. Inner ghosts refer to ghost sources that are formed between the dominator and the contaminator (in our case it is $[0^\circ, 1^\circ]$), while outer ghosts refer to ghost sources found outside the inner ghost interval ($[0^\circ, 1^\circ]$). The contaminator was shifted to 20° , that is, $l_0 = 20^\circ$. ALS and LS were performed on the new set of visibilities using all 14 antennas of WSRT. The resulting ghost distributions are displayed in Fig. 3.12.

Fig. 3.13 presents the theoretical ghost distributions exhibited as percentage of the unmodelled flux A_2 . Most of the strong ghost sources have negative amplitudes. The strongest ghost with an amplitude of around 13% of A_2 is the one sitting at l_0 . This particular ghost is referred to as the “flux suppression” ghost, as it suppresses the intensity of the unmodelled source by approximately 13%. This implies that the flux lost during calibration is equal to the amplitude of the “flux suppression” ghost.

Moreover, the ghost sources at $\{k = n\phi_0\}$ are stronger as these positions are favourable by all baselines. However, there are some exceptions where the ghost amplitudes are weaker than expected. For instance, the ghost amplitudes at -2° , -1° and 2° seem to be weaker than the inner ghosts. This behaviour can be explained by the mix of baselines, which have both positive and negative amplitudes. The addition of these positive and negative amplitudes results in a much smaller amplitude value. Furthermore, the amplitudes of the ghosts decrease as they move away from the phase centre and the pattern seems to extend infinitely within a percentage range of $0.1 \sim 0.2\%$ of A_2 . Fig. 3.13 shows that the theoretically predicted results are consistent with the simulated ones.

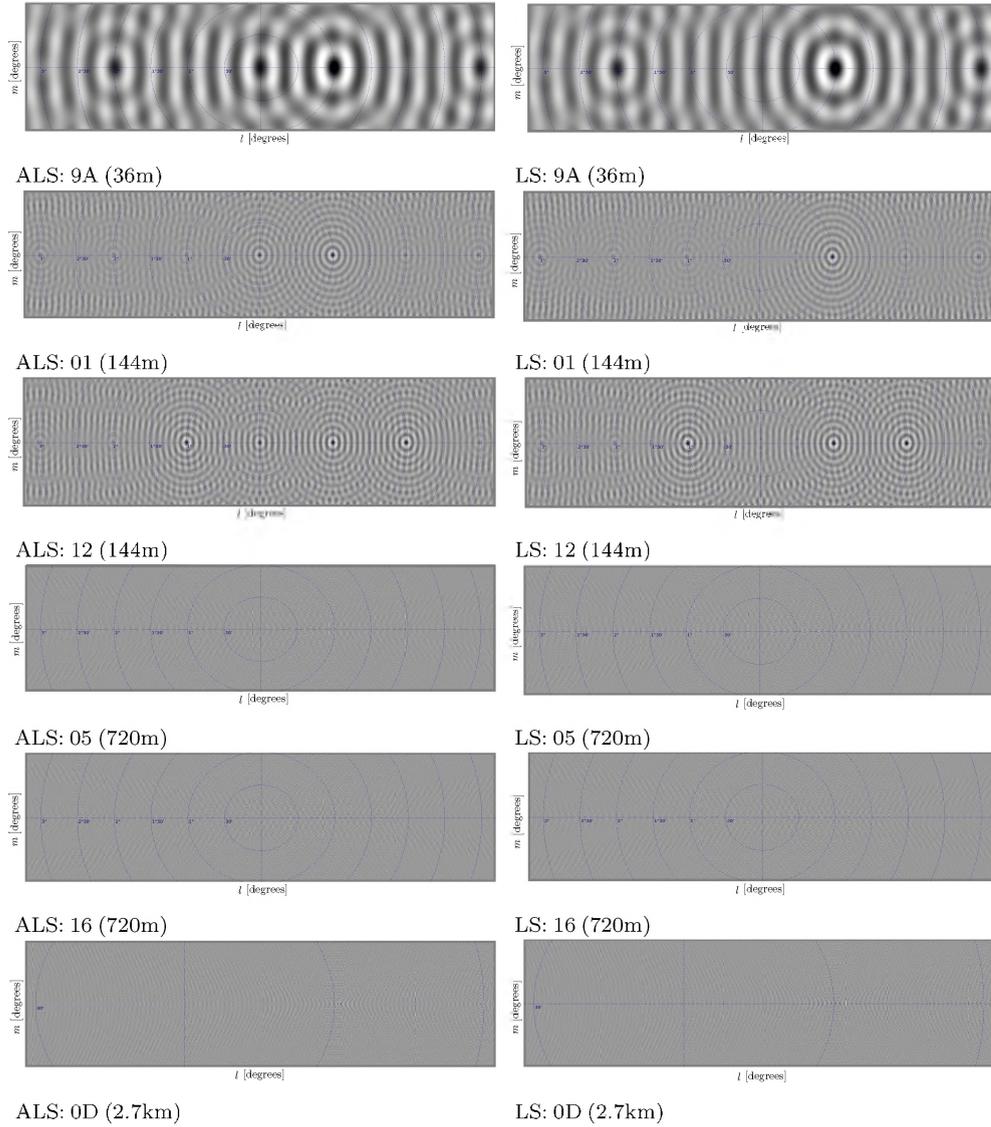


FIGURE 3.10: Dirty images of ALS and LS ghost pattern for baselines 9A (36 m), 01 and 12 (144 m), 05 and 16 (720 m), 0D (2.7 km).

3.5.2 Dependence on flux ratio

Another useful quantity to study is the influence of the fluxes A_1 and A_2 on the amplitudes of the ghost sources. The derivation of \mathcal{G} in Section 3.3 reveals that ghost amplitudes do depend on A_1 and A_2 . Since $A_1 = 1$ throughout this entire experiment, they would only be influenced by A_2 . The ghost amplitudes are outcomes of eigenvalue decomposition, therefore it is fairly difficult to find an analytical expression for them. If they are therefore assumed to behave linearly with A_2 ,

$$\zeta_k \approx K_k A_2, \quad (3.43)$$

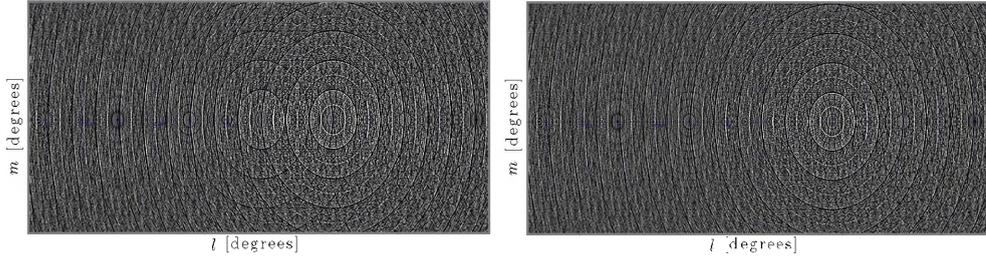


FIGURE 3.11: ALS (left) and LS(right) distilled ghost pattern for the full WSRT array with $l_0 = 1^\circ$.

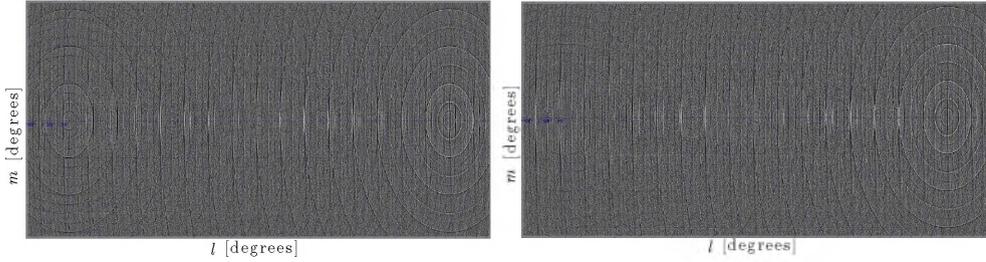


FIGURE 3.12: ALS (left) and LS(right) distilled ghost pattern for the full WSRT array with $l_0 = 20^\circ$.

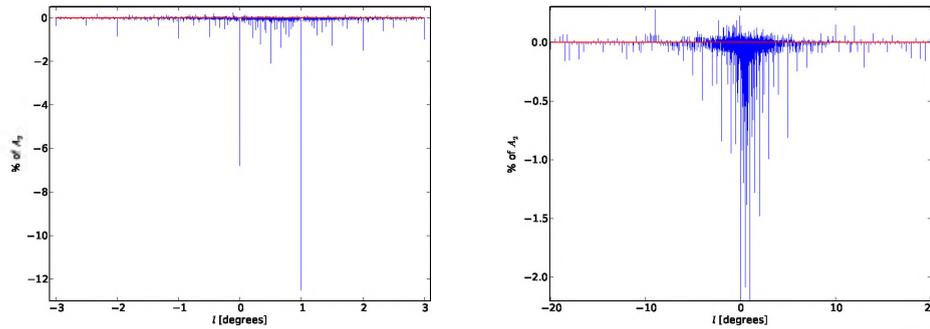


FIGURE 3.13: Theoretical residual ghost pattern within 3° (left plot) and 20° (right plot) of phase centre, for the full WSRT array. The secondary source was positioned at $l_0 = 1^\circ$. Amplitudes are given as a percentage of A_2 flux. The 3° plot shows all the ghosts, while in the 20° plot, only ghosts with amplitudes in excess $\approx 0.009\%$ of A_2 are shown, and the y axis is cut off just below the $\frac{1}{2}^\circ$ ghost - the 0° and 1° ghost response thus extends well below the plot limits.

where the constant K_k is estimated using LS. According to Fig. 3.14, the relative error between the ghost amplitudes and the estimated ones is negligible for the 13 brightest ghosts. The magnitude of the relative error is calculated using

$$\epsilon_k = \left| \frac{\zeta_k - K_k A_2}{\zeta_k} \right|, \quad (3.44)$$

Most of the ghost sources seem to vary linearly within 10% of A_2 , except the flux suppression ghost at 1° , which is linear to within 1%. The ghosts that seem to be least linear are at 2° and 3° . Thus, the linear relationship defined in Eq. 3.43 can be used to

estimate ghost amplitudes.

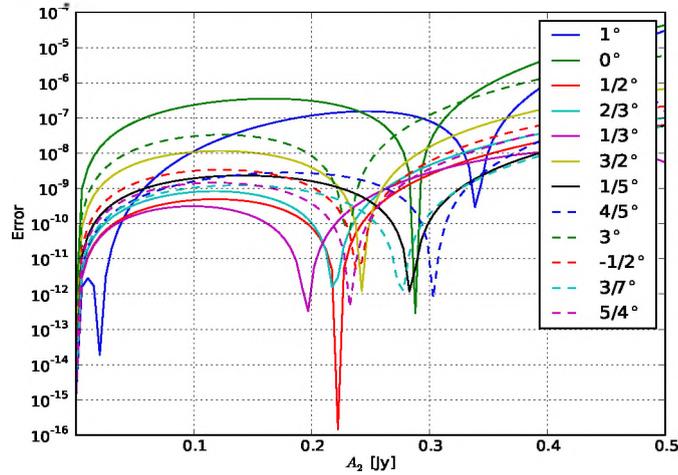


FIGURE 3.14: The relative error magnitude of the linear fit (Eq. 3.44) for the first 13 brightest ghost sources. The legends are the ghost locations, with $l_0 = 1^\circ$.

It is obvious from the calibration equation defined in Eq. 3.9 that scaling the true and model skies by a common factor does not affect the gain solutions, from which the distilled ghost amplitudes ζ_k are attained. Consider a calibration problem with $A_1 = A'_1 \neq 1$ and $A_2 = A'_2$. The resulting Fourier coefficients are equal to the coefficients generated when $A_1 = A'_1 \neq 1$ and $A_2 = A'_2$. This implies that the distilled ghost amplitudes depend on the flux ratio of the two sources and not their absolute flux values, whereas, the ghost amplitudes in the corrected or residual visibilities scale with the absolute flux as a result of the distilled ghost pattern being convolved with the sky. For fluxes A'_1 and A'_2 , the residual ghost amplitudes are given by

$$\begin{aligned} \zeta_k^\Delta &\approx A'_1 K_k \frac{A'_2}{A'_1} + A'_2 K_{k-\phi_0} \frac{A'_2}{A'_1} \\ &= K_k A'_2 + K_{k-\phi_0} \frac{A'_2}{A'_1}. \end{aligned} \quad (3.45)$$

When $A'_2 \ll A'_1$, the first term of Eq. 3.45 dominates the second term, making ζ_k^Δ nearly independent of A'_1 . This explains the observation in Smirnov (2010).

Fig. 3.15 shows the 13 brightest theoretically derived distilled ghost amplitudes ζ_k/A_2 as a function of A_2 . Straight lines are expected as per Eq. 3.43. The more the deviation from the horizontal line, the less linear these ghost sources behave with respect to A_2 . From Fig. 3.15, the ghosts at 2° and 3° appear to be the least linear.

Fig. 3.16 shows the theoretically derived relative ghost amplitudes ζ_k^Δ/A_2 as a function of A_2 for the 13 brightest ghosts in the residual ghost pattern. The linear dependence

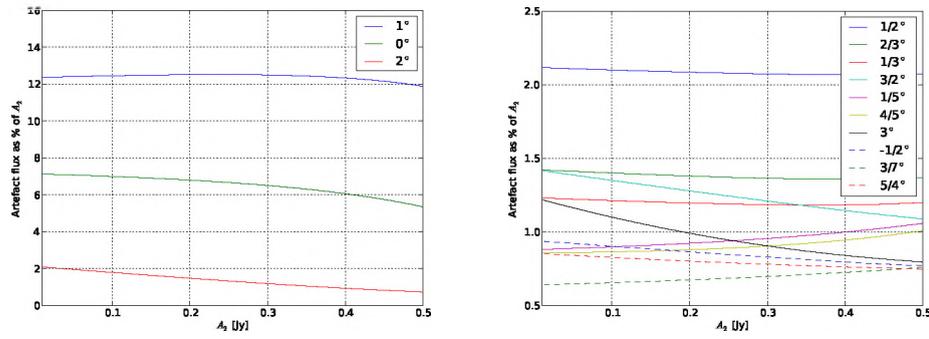


FIGURE 3.15: The relative amplitude $|\zeta_k/A_2|$ of the top 13 (top 3 on the left and 4-13 on the right) ghosts in the distilled ghost pattern, as a function of A_2 . Ranking is by ghost amplitude at $A_2 = 0.5$. The ghost positions are indicated by the legend, with source A_2 being at 1° .

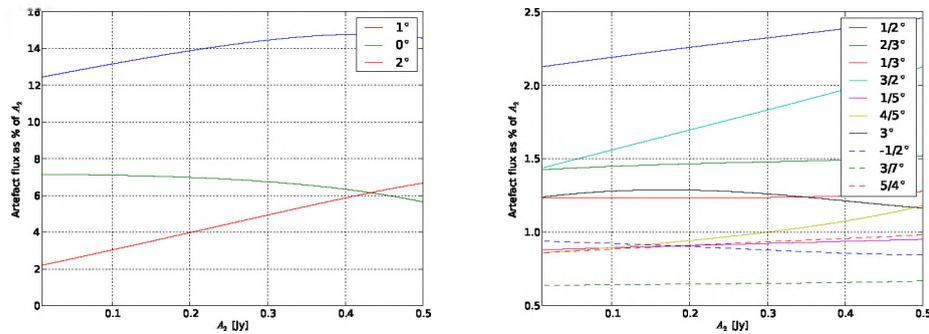


FIGURE 3.16: The relative amplitude $|\zeta_k^\Delta/A_2|$ of the top 13 (top 3 on the left and 4-13 on the right) ghosts in the residual ghost pattern, as a function of A_2 . Ranking is by ghost amplitude at $A_2 = 0.5$. The ghost positions are indicated by the legend, with source A_2 being at 1° .

of the residual pattern is different because of the presence of the additional term in Eq. 3.45. To obtain the relative ghost amplitudes, the terms in Eq. 3.45 are divided by A_2 . For larger values of A_2 , the second linear term begins to dominate, as shown in Fig. 3.16. The ghost at 2° becomes stronger than that at 1° as it contains the distilled component at 1° , scaled by A_2 .

Useful information can be gathered by investigating the influence of the source positions and the number of antennas on ghost amplitudes.

3.5.3 Dependence on the source position and antennas

From Proposition A.6, it is clear that l_0 and m_0 have no effect on the ghost amplitudes. They either shrink or stretch the ghost patterns. We therefore study the effect of the number of antennas on source suppression. The fluxes are set to $A_1 = 1$ Jy and $A_2 = 0.2$ Jy. Keeping m_0 constant at zero, l_0 is varied within a FoV of 0.5° . Fig. 3.17 displays the

theoretically derived amplitudes ζ_1^Δ of the suppression ghost at 1° in the residual ghost pattern, which is equal to the the source suppression of the contaminator mentioned in Subsection 3.5.2 (the loss in the contaminator's flux, A_2). Theoretically, the residual ghost amplitude ζ_1^Δ is not affected by the position of the contaminator. This result is consistent with Proposition A.6 that states that l_0 and m_0 do not influence the ghost amplitudes.

We imaged the corrected visibilities yielded from ALS and LS using *lwimager* and extracted the fluxes from the dirty images. Fig. 3.18 presents the source suppression of the contaminator, acquired with ALS and LS as a function of l_0 . A large degree of fluctuation can be observed. Grating lobes associated with the ghost sources due to regular layout of WSRT might account for this fluctuating behaviour.

The number of antennas being used appears to be the most influential factor, as shown in Fig. 3.17 and 3.18. The degrees of freedom increases with the number of antennas, thus enabling selfcal to model the missing flux better. This explains the decrease in the flux suppression ghost amplitudes as the number of antennas increases.

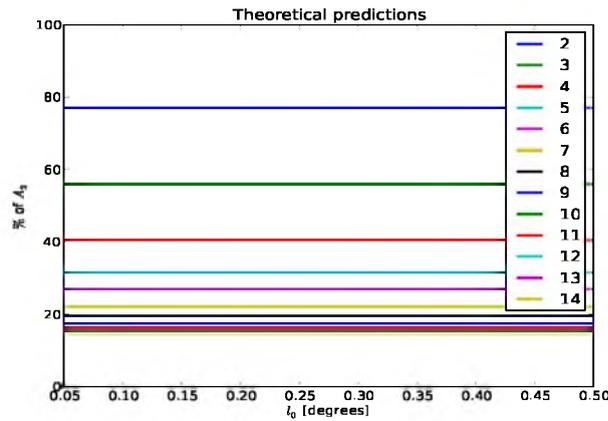


FIGURE 3.17: The theoretical amplitude $|\zeta_1^\Delta|$ (ghost amplitude at 1°) as a function of l_0 . The plot also exhibits the influence of the number of antennas on $|\zeta_1^\Delta|$.

3.5.4 The role of array redundancy

The array geometry of WSRT is an important factor in determining the structure of the ghost distribution as portrayed in Section 3.3. The set of discrete ghost positions can be represented as $\{\frac{k s_0}{\phi_0}\}$, where k is an integer value. Each baseline produces a subset of the above set. As mentioned earlier and displayed by Fig. 3.7-3.9, the ghosts for shorter baselines are sparsely spread compared to longer baselines. The shortest baseline (9A:36 m) produces ghost sources at intervals of 1° (represented by circle symbol). The ghosts that are formed along baselines of length 144 m (01 and 12) are 0.25° apart (represented

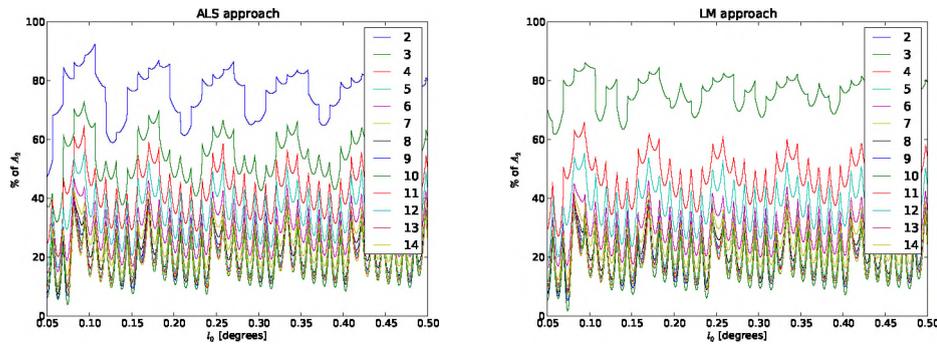


FIGURE 3.18: The relative amplitude of $|\zeta_1^\Delta|$ (ghost amplitude at 1°) as a function of l_0 . The plot also exhibits the influence of the number of antennas on $|\zeta_1^\Delta|$. The plot on left side displays the ghost amplitude obtained with the ALS approach, while the right-hand plot represents the outcomes of the LM approach.

by up/down triangles) while those along baselines of length 720 m (05 and 16) are 0.05° apart (represented by left/right triangles). Lastly, the ghost sources produced by the longest baseline (0D: 2.7 km, represented by the horizontal tick marks) are finely spaced.

When the full array is used, the ghosts are still formed at the same positions but with different amplitudes (sometimes with a different sign). Different calibration solutions are obtained for specific baselines, which account for the variation in amplitudes. The positions $k = n\phi_0$; $n \in \mathbb{Z}$ contribute to all baselines. Indeed, we have seen that the strongest ghosts occur at 1° , 0° and -2° in Fig. 3.15. The next strongest one seems to form at $\frac{1}{2}^\circ$, that is, at $k = \frac{\phi_0}{2}$, since most of the ϕ'_{pq} s are even in the (36, 108, 1332, 1404 m) WSRT configuration.

Moreover, Eq. 3.26 provides us with qualitative understanding of the spatial ghost distribution for a less regular-spaced array. Assume \mathbf{b}_0 to be of length one and that the antennas are configured in such a way that the spacings between them are prime integer multiples of \mathbf{b}_0 . Then this configuration will result in a large Φ matrix, the least common multiple being the product of all spacings. Each baseline will result in a unique ghost pattern, that is, no two baselines will produce a ghost source at the same position, except at $k = n\mathbf{s}_0$. The resulting map will hence have a large number of weak sources and their respective GSF will contribute to the noise. Also, decreasing \mathbf{b}_0 would result in large ϕ_0 , thus generating finely spaced ghost sources. The unforeseen disadvantage of highly redundant arrays being prone to strong ghosts should be taken care of when designing future arrays.

3.6 Conclusion

From this study, we reach the following conclusions:

- The formation of the string point-source like objects observed in Fig. 1.3 and 1.4 is caused by the selfcal algorithm, which uses incomplete sky models. This implies that these ghost sources will always appear while calibrating with an incomplete sky model.
- The ghost sources occur at integer multiples of $\frac{s_0}{\phi_0}$, given any regularly spaced array. The ghost positions are therefore determined based on the source positions and the array configuration.
- The amplitudes of the ghost sources behave linearly with the flux of the contaminator, A_2 .
- Theoretically, the source positions do not have any effect on the ghost amplitudes, but in real life they do. The images are dominated by grating lobes associated with the ghost sources, which are produced in response to the regular configuration of WSRT. These grating lobes influence the flux measurements, thus causing large degrees of fluctuation in the ghost amplitudes.
- The number of antennas being used has a great effect on the amplitudes of the ghost sources. The more antennas involved, the lower is the suppression rate (or the flux suppression ghost amplitudes).
- Highly redundant arrays are much likely to be prone to strong ghost sources.

3.7 Summary

This chapter analysed the underlying cause of one class of calibration artefacts known as ghosts by implementing a theoretical framework, whereby the spatial ghost distribution was predicted. The predicted outcomes were then compared with simulations. In the next chapter, we shall study how the degree of incompleteness of a calibration sky model affects the recovered fluxes or source suppression.

Chapter 4

Source Suppression

“This sight... is by far the noblest
astronomy affords”

–Edmond Halley

Acquiring fascinating radio maps is wonderful, but it is not the final step. The desired quantity in producing radio maps is the spectral flux density or the spectral power flux density, measured in Watt per square metre per Hertz per steradian ($\text{Wm}^{-2}\text{Hz}^{-1}\text{sr}^{-1}$), which predicts the strength of the electromagnetic radiation from the observed source. The loss in spectral flux density during calibration is termed as source suppression. This chapter investigates whether a link exists between the use of incomplete calibration sky models and source suppression. Alternately, this chapter can be regarded as an extended study of the “flux suppression” ghost portrayed in Chapter 3. It also focuses on analysing source suppression as functions of the solution intervals and the number of DDEs. We first examine the concept of spectral flux density. Then we construct our calibration pipeline and design an appropriate experimental setup. The point estimation for source suppression is decided upon prior to the aforementioned investigations. Conclusions are drawn based on the results.

4.1 Spectral Flux Density

The surface brightness or intensity I , measured in units of $\text{Wm}^{-2}\text{Hz sr}^{-1}$ is defined as the radiation measured per unit area per unit time per unit solid angle within a specific frequency band. According to [Thompson et al. \(2001\)](#), for black-body radiation, the

intensity is related to the physical temperature T of the radiating body, that is,

$$I = \frac{2kT\nu^2}{c^2} \left[\frac{h\nu}{kT} \right], \quad (4.1)$$

where k is Boltzmann's constant, c is the speed of light and h is Planck's constant. Since radio photons have a very low frequency, $\frac{h\nu}{kT} \ll 1$, so we may approximate Eq. 4.1 with

$$I = \frac{2kT\nu^2}{c^2},$$

which is commonly referred to as the Rayleigh-Jeans law. The quantity of interest is the spectral flux density S , given by the integral of the surface brightness

$$S = \int_{\Omega} \int I d\Omega.$$

For convenience, spectral flux density is often called flux density by radio astronomers. The surface brightness may vary with direction, therefore it is expressed as a function of the angles θ and ϕ , as pointed out in Kraus (1996). The angles θ and ϕ are shown in Fig. 4.1. If the source is observed by an antenna with a beam pattern $P_n(\theta, \phi)$, then

$$S = \int_{\Omega} \int I(\theta, \phi) P_n(\theta, \phi) d\Omega.$$

The beam pattern is a measure of the antenna response in terms of θ and ϕ and it is distinct for each antenna. The surface brightness of radio maps is usually measured in units of “Jy per beam”, where $1 \text{ Jy} = 10^{-26} \text{ Wm}^{-2} \text{ Hz}^{-1}$.

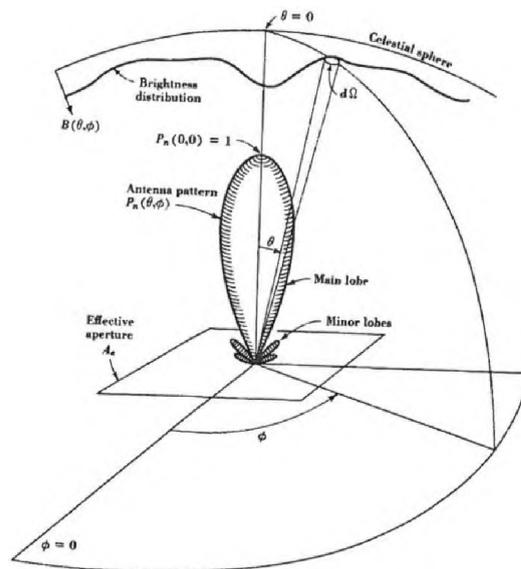


FIGURE 4.1: The radiation pattern with respect to the celestial sphere (Credit J.D. Krauss).

4.2 Implementation of calibration pipeline

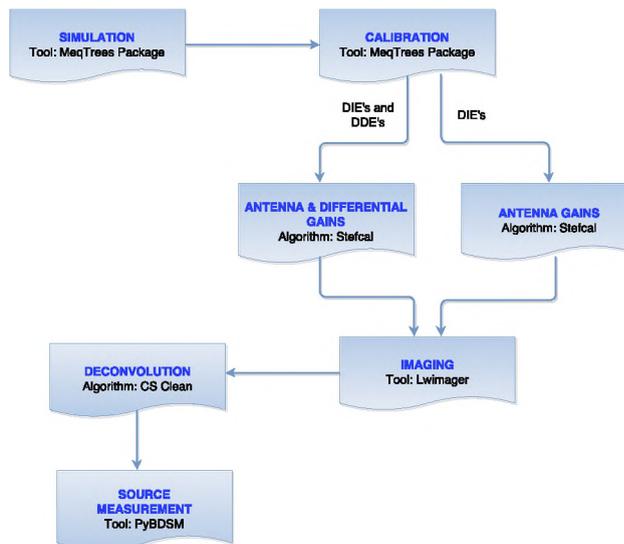


FIGURE 4.2: Flow chart of the LS calibration pipeline

The first thing that was done in this study was to develop a customised python pipeline¹, which could load a sky model within the FoV of WSRT. The key functional blocks of the pipeline are presented in Fig. 4.2. We briefly discuss each component below.

(1) Simulation

Given a sky model (nominal sky), generate per-baseline visibilities V_{pq} through a 12-hour synthesis simulation using a single channel of width 312 kHz at a frequency of 1460 MHz. Use an integration of 30 s. The MeqTrees package (Noordam and Smirnov (2010)) is used as the simulating tool.

(2) Calibration

- Construct calibration sky model, \mathbf{M} , by taking out a portion $P\%$ of the faintest sources from the nominal sky.
- Estimate the antenna-based gains using

$$\hat{\mathbf{G}} = \operatorname{argmin} \|\mathbf{V}^R - \mathbf{G}\mathbf{V}^M\mathbf{G}^H\|_F^2,$$

¹The python implementation of the pipeline discussed above can be found <https://github.com/Chuneeta/Calibration-Pipeline>

where \mathbf{V}^R is the observed visibilities and \mathbf{V}^M is the model visibilities. The minimisation is done via the linear least square solver, StEFCal, discussed in Chapter 2. A solution interval of 30 s was used, with one frequency channel.

- Solve for the differential gains ΔE_{sp} , where s is the source subjected to DDE, using StEFCal algorithm (in the absence of DDEs, solve only for antenna gains). The solution interval is set to one, per antenna, per source, per band, which is similar to the antenna gains solution interval.

(3) Imaging

Compute the corrected visibilities $V_{pq}^\Delta = \hat{g}_p^{-1} V_{pq} \hat{g}_q^{-H}$, where ‘ $-H$ ’ denotes the inverse transpose. Image V_{pq}^Δ over all baselines using *lwimager*. The visibilities are Fourier transformed using a gridding approach (Sramek and Schwab (1999)), which usually results into poor sampling of the uv plane. A radial weighting scheme is considered. The resulting “dirty image” is given by

$$I_D = B_D * I, \quad (4.2)$$

where the dirty beam

$$B_D = \sum_{pq} \mathcal{F}^{-1}\{S_{pq}\}. \quad (4.3)$$

In Eq. 4.3, S_{pq} denotes the weighted sampling function of baseline pq .

(4) Deconvolution

Define the region where the emission is expected. Deconvolve that region using the iterative Cotton-Schwab (CS) CLEAN algorithm ((Cornwell et al., 1999)). The CLEAN components obtained for that region are added into a Gaussian restoring beam. The cleaning threshold is set to one thousandth of the intensity of the brightest source in the sky model. The deconvolution process mitigates most sidelobes present in Eq.4.2, which removes the undesired noise from the real emission and allows for the fainter components (masked by these sidelobes) to be noticed.

$$I_D \sim I.$$

(5) Source Measurement

Extract the flux density, S , of the individual radio sources from the residual map, using Python Blob Detection and Source Measurement (PyBDSM). PyBDSM decomposes the interferometric maps into Gaussians as pointed out in (Rafferty and Mohan, 2013). The lower threshold is set to three sigma and the upper threshold to five sigma, where sigma denotes the noise level in the image. Based upon these

settings, the sum of the pixel values that lie within that region is the output flux density of the corresponding sources.

With a handy LS calibration pipeline, one can proceed with the experimental setup.

4.2.1 Experimental setup

A random sky distribution is chosen as a preliminary experimental setup. The field centre is taken to be $(l = 0^\circ, m = 0^\circ)$ and the sources are distributed within a field of view of 2° , ranging from -1° to 1° . The coordinates, l and m are uniformly distributed such that git clone <https://ridhima24@bitbucket.org/pmsutter/blackjack.git>

$$l(x) = m(x) = \begin{cases} \frac{1}{c-b} & \text{if } b \leq x \leq c; \\ 0 & \text{if } x < b \text{ or } x > c. \end{cases} \quad (4.4)$$

The corresponding flux values follow a translated pareto distribution

$$A(x) = \frac{a}{(x+1)^{a+1}}, \quad (4.5)$$

where a is the shape parameter. The normal pareto distribution is restricted to $x \geq 1$, but the need for a realistic dynamic range (DR) 1 : 3000 necessitates the usage of Eq. 4.5. The parameter ‘a’ is set to six. All sky distributions used throughout this study were generated using the same principle, unless specified otherwise.

4.3 Source Suppression Estimates

The question that arises next is how are we going to evaluate the loss in flux density with respect to the nominal sky distribution, given that the calibration pipeline yields the flux density r_i of individual unmodelled sources. The retrieved flux density r_i can be related to its corresponding nominal flux density (flux density of the source from the nominal sky) a_i via

$$r_i = \beta a_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\epsilon); i \in \{1, 2, \dots, n\}, \quad (4.6)$$

where the β is the slope. The number of unmodelled sources is indicated by n and ϵ_i denotes the error term. The error justifying the difference between the independent (a_i) and dependent (r_i) variables follows a zero mean normal distribution with a variance of

σ_ϵ^2 . The suppression factor can thus be defined as follows:

$$\alpha = 1 - \beta. \quad (4.7)$$

The parameter β is first estimated and later it is substituted in Eq. 4.7 to obtain the suppression factor estimates. We will be examining three different estimators of β ; a simple unweighted average, a weighted average and ordinary least squares (OLS).

4.3.1 Unweighted average

The unweighted estimator averages the data points over the unmodelled sources such that all data points make an equal contribution to the final average as shown below:

$$\hat{\beta}_1 = \frac{1}{n} \sum \left(\frac{r_i}{a_i} \right) = \frac{1}{n} \sum \left(\frac{r}{a} \right)_i. \quad (4.8)$$

The standard error (SE) associated with β_1 is given by

$$SE_{\hat{\beta}_1} = \left(\frac{\sigma_\epsilon^2}{n^2} \sum \frac{1}{a_i^2} \right)^{\frac{1}{2}}, \quad (4.9)$$

where $\sigma_\epsilon^2 = \text{var}(r_i)$. Since σ_ϵ is unknown, it is estimated with

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-1} \sum (\epsilon_i)^2. \quad (4.10)$$

Unweighted estimators are generally sensitive to outliers (data points that fall outside an expected boundary).

4.3.2 Weighted Average

A weighted mean is an arithmetic mean whereby some data points contribute more than other data points to the final mean. The second estimator that we are considering is calculated using

$$\hat{\beta}_2 = \frac{\sum w_i \left(\frac{r_i}{a_i} \right)}{\sum w_i}, \quad (4.11)$$

with w_i representing the relative contribution of the i^{th} data point. For this investigation, the weighting factor is taken to be the nominal flux a_i of the point source. Substituting a_i in Eq. 4.11 gives

$$\hat{\beta}_2 = \frac{\sum r_i}{\sum a_i}. \quad (4.12)$$

If the weight is unity throughout, that is, $w_i = 1$ for all i , then Eq. 4.11 reduces to Eq. 4.8. Similar to the unweighted average, the SE corresponding to $\hat{\beta}_2$ is expressed as

the square root of the variance

$$SE_{\hat{\beta}_2} = \left(n\sigma_\epsilon^2 \frac{1}{(\sum a_i)^2} \right)^{\frac{1}{2}}, \quad (4.13)$$

where the estimate of σ_ϵ^2 is defined in Eq. 4.10.

4.3.3 Ordinary least square (OLS) estimation

This approach minimizes the square vertical distances between the nominal fluxes and retrieved fluxes as depicted below:

$$\min_{\beta_3} \sum (r_i - \beta_3 a_i)^2. \quad (4.14)$$

The minimal estimate β_3 is obtained by equating the derivative of Eq. 4.14 to zero. Differentiating Eq. 4.14 with respect to β gives

$$\begin{aligned} \frac{d}{d\beta} \sum (r_i - \beta a_i)^2 &= \sum 2(r_i - \beta_3 a_i)(-a_i) \\ &= -2 \sum (r_i a_i - \beta_3 a_i^2). \end{aligned} \quad (4.15)$$

Equating Eq. 4.15 to zero results in

$$\hat{\beta}_3 = \frac{\sum r_i a_i}{\sum a_i^2}. \quad (4.16)$$

Note that if the weights in Eq. 4.11 are taken to be the square nominal fluxes, that is, $w_i = a_i^2$, then Eq. 4.11 transforms into Eq. 4.16. The SE corresponding to β_3 is once again given by the square root of its variance

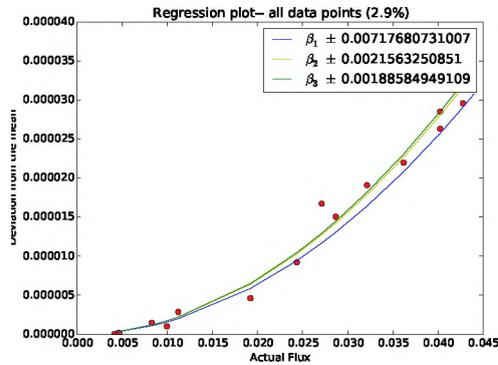
$$SE_{\hat{\beta}_3} = \sigma_\epsilon^2 \frac{1}{\sum a_i^2}, \quad (4.17)$$

where the estimated value of σ_ϵ^2 can be calculated with Eq. 4.10.

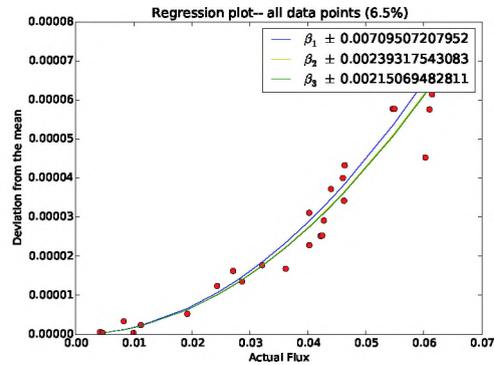
4.3.4 Optimal estimator

The optimal one out of the three estimators needs to be determined. Therefore, a random sky model with 100 sources is generated and fed to the python implemented pipeline described in Section 4.2. The portion $P\%$, of unmodelled sources was initially taken to be 20%. The output flux densities are graphically illustrated by the red dots in Fig. 4.4. The total recovered flux density is then estimated using the three estimators with the cyan line denoting the unweighted average, the yellow line representing the

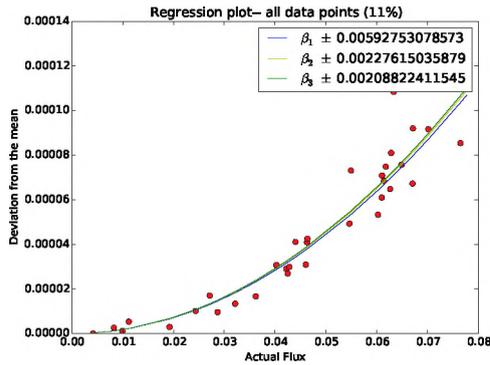
weighted mean and the green line representing the OLS approach. A proper conclusion cannot be drawn based upon a single experiment. The experiment is hence repeated for different P values; 20, 30, ..., 90 and the resulting plots can be found in Fig. 4.4. In most of the plots, the regression lines lie on top of each other, which makes it hard to determine the best estimator. The SEs associated with the estimates aid in choosing the ultimate estimator, which is OLS.



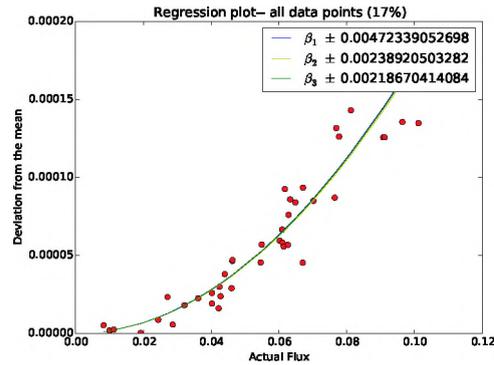
(A) 20% unmodelled sources



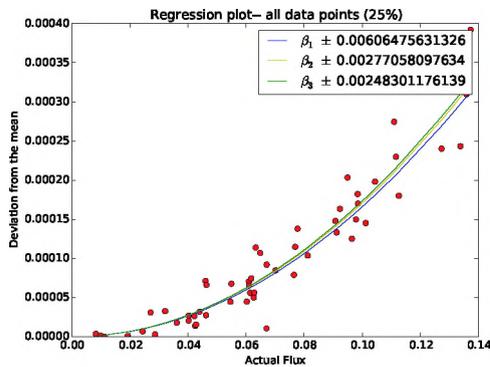
(B) 30% unmodelled sources



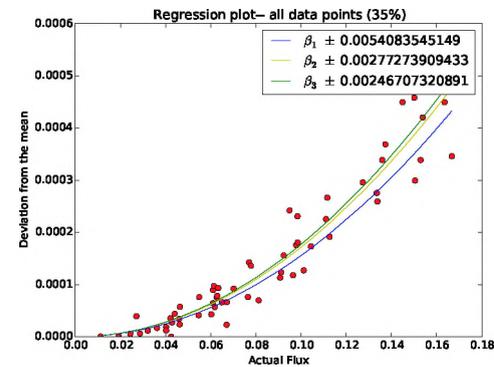
(C) 40% unmodelled sources



(D) 50% unmodelled sources



(E) 60% unmodelled sources



(F) 70% unmodelled sources

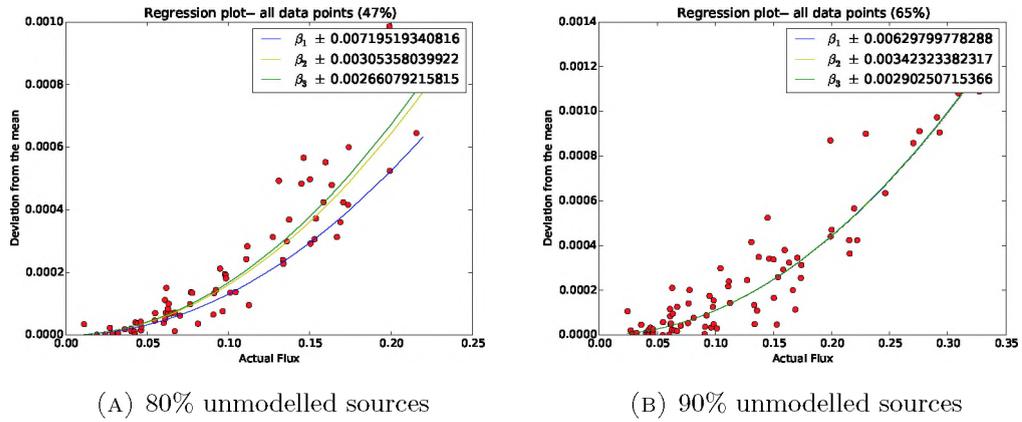
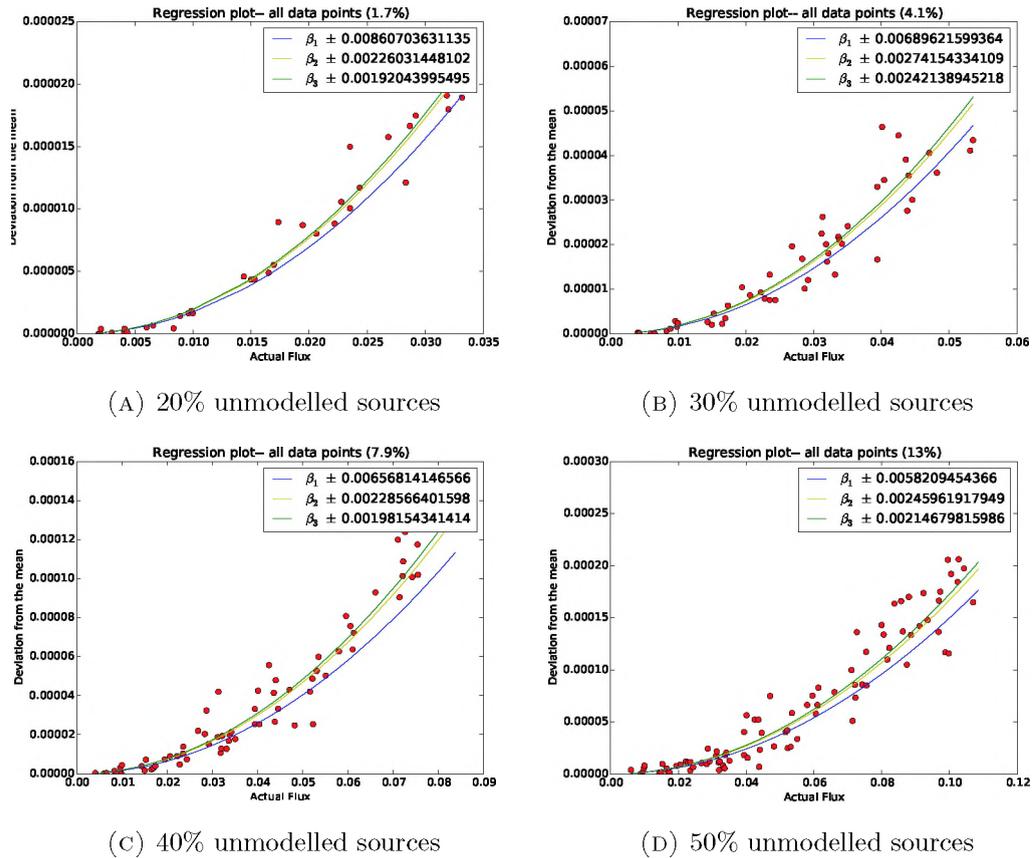


FIGURE 4.4: Regression lines acquired through unweighted averaging (cyan line), weighted averaging (yellow line) and OLS estimation (green line) with a 100-source random sky model. The legend at the bottom right-hand displays the SE associated with the different slopes, β_1 , β_2 , β_3 .

The same procedure is performed on a random sky of 200 sources. The resulting regression plots are displayed in Fig. 4.6. Fig. 4.6 also indicates that $\hat{\beta}_3$ seems to be the optimal estimate.



Intriguingly, Fig. 4.6b, 4.6c, and 4.6d exhibit two distant data points. To explore the origin of these data points, one has to take a closer look at the measurements found in

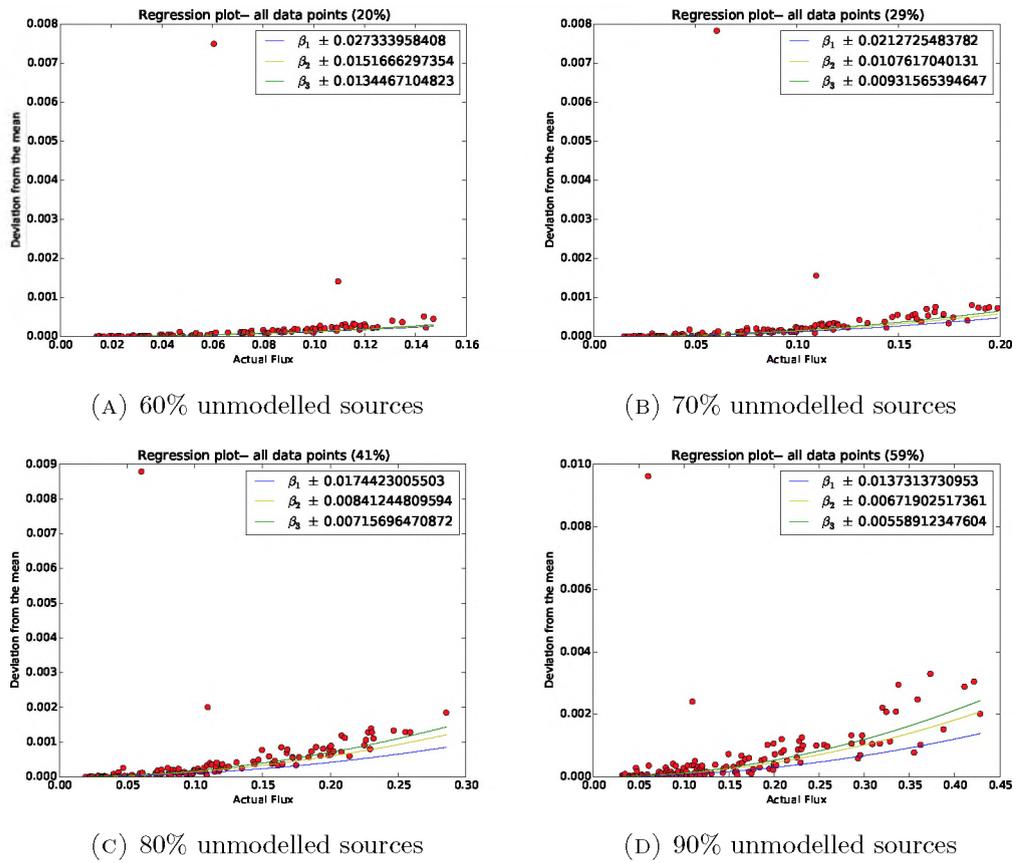


FIGURE 4.6: Regression lines acquired through unweighted averaging (cyan line), weighted averaging (yellow line) and OLS estimation (green line) with a 200-source random sky model. The legend at the bottom right-hand displays the SE associated with the different slopes, β_1 , β_2 , β_3 .

Fig. 4.7 with which the plot in Fig. 4.6b was created. Each column contains information of the individual unmodelled sources. The columns are the identity of the source (ID), RA in degrees (Ra_d), DEC in degrees (Dec_d), the nominal flux density (N_Flux), the retrieved flux density (R_Flux), and the error associated with the retrieved flux density (E_R_Flux). The sources highlighted in blue are the sources whose fluxes are overestimated. A122 and A129 can be associated with the two distant points that are spotted in Fig. 4.6b and 4.6c. It is quite worrisome that both A122 and A129 share a common value for R_Flux. These two sources are indicated by the black square in Fig. 4.8. These observations indicate that PyBDSM interprets adjacent sources as a single point source.

```

***** 200 Sources in skymodel *****
***** 140 Sources in calibration model *****
***** 29% unmodelled flux *****

```

ID	Ra_d	Dec_d	N_Flux	R_Flux	E_R_Flux
A176	86.9021	50.0365	0.01499	0.01621	0.003848
A97	86.5185	50.2561	0.01652	0.01389	0.004115
A193	86.6268	49.1178	0.01696	0.01611	0.003796
A11	85.0331	50.2816	0.01736	0.01825	0.003912
A41	84.1295	50.3906	0.01945	0.01682	0.003996
A154	86.5958	50.2233	0.02063	0.01729	0.00408
A21	86.0892	50.5738	0.02221	0.01959	0.004129
A134	86.5765	50.5655	0.02277	0.01839	0.004027
A30	85.1146	50.5504	0.02353	0.0205	0.004042
A188	86.1155	49.818	0.02353	0.02027	0.004059
A16	85.9774	49.5548	0.02436	0.02458	0.00393
A6	85.2501	50.1815	0.02683	0.02779	0.00415
A159	86.2697	50.1468	0.02833	0.02487	0.004143
A86	85.9511	50.2569	0.02868	0.02238	0.004203
A161	84.6032	50.5806	0.02919	0.02507	0.00404
A63	84.605	50.5581	0.03119	0.02813	0.003986
A91	85.6572	49.4515	0.03134	0.02855	0.004011
A60	86.8163	49.4428	0.03186	0.02855	0.003842
A163	84.7031	49.5604	0.032	0.02644	0.004007
A119	86.7428	49.6091	0.03217	0.02901	0.003914
A79	86.2326	50.0407	0.03317	0.02969	0.004084
A18	85.5825	49.3796	0.03365	0.03384	0.004012
A130	87.0769	50.8488	0.03381	0.02857	0.00386
A74	84.5883	49.255	0.0342	0.03089	0.003964
A84	85.4779	49.5616	0.03506	0.03289	0.004041
A124	84.5498	49.9457	0.03941	0.03552	0.004015
A43	84.2985	49.8989	0.03941	0.03531	0.004017
A56	84.3001	50.3391	0.04012	0.03047	0.004027
A77	85.767	49.4199	0.04042	0.0375	0.004018
A52	84.8234	49.3441	0.04253	0.03588	0.004063
A1	86.3688	49.3032	0.04361	0.0353	0.003953
A61	84.2221	48.897	0.04384	0.04024	0.003924
A67	86.8348	50.2414	0.04404	0.04109	0.003939
A23	84.4414	49.8524	0.04457	0.03855	0.004014
A151	85.5349	50.3369	0.04708	0.03412	0.004195
A156	84.1661	50.1281	0.04821	0.04176	0.004027
A132	85.286	50.1355	0.05161	0.04501	0.004199
A27	84.3795	49.8195	0.05214	0.0473	0.004
A39	85.4109	49.9594	0.05229	0.0469	0.004183
A92	85.033	50.6659	0.05312	0.04801	0.004072
A103	84.3327	50.2539	0.05353	0.04344	0.004042
A110	85.1624	49.0927	0.05509	0.05056	0.003925
A114	86.8731	50.7543	0.05802	0.05273	0.003891
A189	84.4052	50.0295	0.05959	0.0504	0.004012
A129	84.2366	49.48	0.06055	0.1511	0.003932
A185	84.6025	50.447	0.06106	0.05214	0.004042
A25	84.6577	48.9562	0.06131	0.05163	0.003955
A37	86.4197	49.061	0.06661	0.05924	0.00389
A169	85.8906	50.1201	0.07108	0.06087	0.004149
A157	86.3428	49.8476	0.07143	0.0652	0.004078
A104	86.4761	49.2015	0.07213	0.06542	0.003894
A150	86.235	49.9822	0.07225	0.06152	0.004136
A184	84.2018	50.424	0.07268	0.06143	0.004017
A178	87.0412	50.7777	0.07424	0.06488	0.003879
A24	85.8127	50.4807	0.07541	0.06122	0.004138
A44	86.3517	50.6044	0.07551	0.06409	0.003994
A109	84.4113	49.2633	0.08	0.06637	0.003984
A136	85.2853	50.485	0.0806	0.07068	0.004125
A171	84.5179	50.8299	0.08165	0.07359	0.003938
A68	85.0947	50.7538	0.08222	0.07147	0.004034
A140	85.5647	50.0088	0.08377	0.06854	0.004145
A93	85.3369	49.9342	0.08574	0.07391	0.00422
A162	86.5612	49.3412	0.08628	0.07698	0.003896
A183	85.1942	49.3352	0.08744	0.07692	0.004065
A172	85.7119	48.9964	0.08804	0.07558	0.003932
A125	86.4383	50.3018	0.08862	0.07766	0.004115
A191	84.149	50.8065	0.09119	0.0791	0.003832
A190	84.4241	49.7451	0.09234	0.07855	0.004033
A59	84.3296	49.6657	0.09357	0.08069	0.003982
A19	84.101	49.7914	0.09679	0.08387	0.003986
A54	86.5138	49.9684	0.09693	0.08554	0.004027
A127	86.4602	50.2037	0.0973	0.08603	0.004097
A160	84.2723	49.3109	0.09875	0.08843	0.003997
A5	87.0296	50.7239	0.09955	0.08452	0.003915
A175	85.4656	49.9967	0.09978	0.08781	0.004164
A89	84.899	50.7705	0.1005	0.08654	0.004032
A69	84.1417	49.7209	0.1024	0.0873	0.003988
A90	86.3257	50.1769	0.1027	0.08882	0.004129
A155	85.1446	50.378	0.1043	0.09065	0.004166
A73	84.8405	50.8143	0.107	0.09359	0.00399
A182	86.0126	50.5273	0.1085	0.09376	0.004103
A181	86.1283	49.5953	0.1089	0.1002	0.003965
A72	86.861	49.2715	0.1091	0.09229	0.003851
A122	84.2366	49.4721	0.1095	0.1511	0.003932
A38	84.9974	50.3222	0.1111	0.09416	0.004109

A102	85.4369	50.086	0.1123	0.09439	0.004209
A76	85.6628	50.7522	0.1125	0.09496	0.004074
A3	86.7928	49.143	0.1146	0.1007	0.003829
A102	85.4369	50.086	0.1123	0.09439	0.004209
A76	85.6628	50.7522	0.1125	0.09496	0.004074
A3	86.7928	49.143	0.1146	0.1007	0.003829
A139	84.582	49.791	0.1155	0.09915	0.004063
A166	84.0939	49.4124	0.1161	0.09796	0.003961
A144	86.1282	50.6708	0.1179	0.105	0.004059
A148	86.4595	50.7657	0.1183	0.09964	0.003997
A65	84.6457	50.6138	0.125	0.1098	0.00399
A167	84.4751	50.5774	0.1308	0.1095	0.004056
A83	86.11	49.3757	0.1348	0.1162	0.003954
A14	84.9052	50.2064	0.1433	0.1226	0.004119
A42	84.4505	49.6479	0.1441	0.1291	0.003999
A26	84.5953	50.0676	0.147	0.1259	0.004053
A94	85.7815	49.9172	0.1496	0.1234	0.004165
A4	87.0049	48.9266	0.1549	0.1321	0.003734
A194	84.587	50.1705	0.157	0.1343	0.004021
A29	84.1101	50.6416	0.1583	0.1339	0.003865
A87	86.9254	50.5026	0.1591	0.1387	0.003936
A116	84.3289	49.2837	0.1613	0.1417	0.004
A145	85.6043	49.0883	0.1639	0.137	0.003997
A95	85.2894	50.176	0.1647	0.1408	0.004245
A10	85.6143	49.8046	0.1679	0.1432	0.004147
A82	84.3834	50.148	0.1683	0.1413	0.00403
A50	84.3526	50.5442	0.1722	0.1489	0.004004
A105	86.4085	49.3073	0.1723	0.1487	0.00395
A32	85.5975	48.9525	0.1748	0.1559	0.003958
A177	85.8771	50.2404	0.177	0.1562	0.00417
A107	84.7232	50.4522	0.184	0.1621	0.004079
A196	84.5943	48.9904	0.1862	0.1581	0.00396
A46	86.2421	50.1214	0.1895	0.1618	0.004145
A49	86.3728	49.765	0.1928	0.1653	0.004033
A22	86.8876	50.7012	0.1948	0.1668	0.003954
A100	86.7578	50.8298	0.1989	0.1717	0.003936

FIGURE 4.7: Output text file displaying information regarding individual unmodelled sources extracted by PyBDSM when 70% are removed from a 200-source sky model. The sources highlighted in blue are the sources that are overestimated.

We are now well-equipped for analysing the link between the use of incomplete calibration sky models and source suppression.

4.4 Incomplete Calibration Sky Models and Source Suppression

It was feasible at this stage to proceed with the analysis of the link between the use of incomplete calibration sky models and source suppression. One hundred sky models², each containing 100 randomly distributed sources were, therefore generated and $P\%$ of the faintest sources were removed from individual sky samples. Note that the percentage of unmodelled sources $P\%$ is not equal to the unmodelled flux percentage, thus, the unmodelled flux percentages associated with each sample do not correspond with each other. B -spline or basis spline interpolation (Chunshi et al. (2012)) was hence used as an approach to mitigate this obstacle.

Though the sky models were generated using the same parameters, the difference in their suppression curves could be sufficiently large to lead to an incorrect or biased average.

²The number of samples was chosen based on the processing time of each sample.

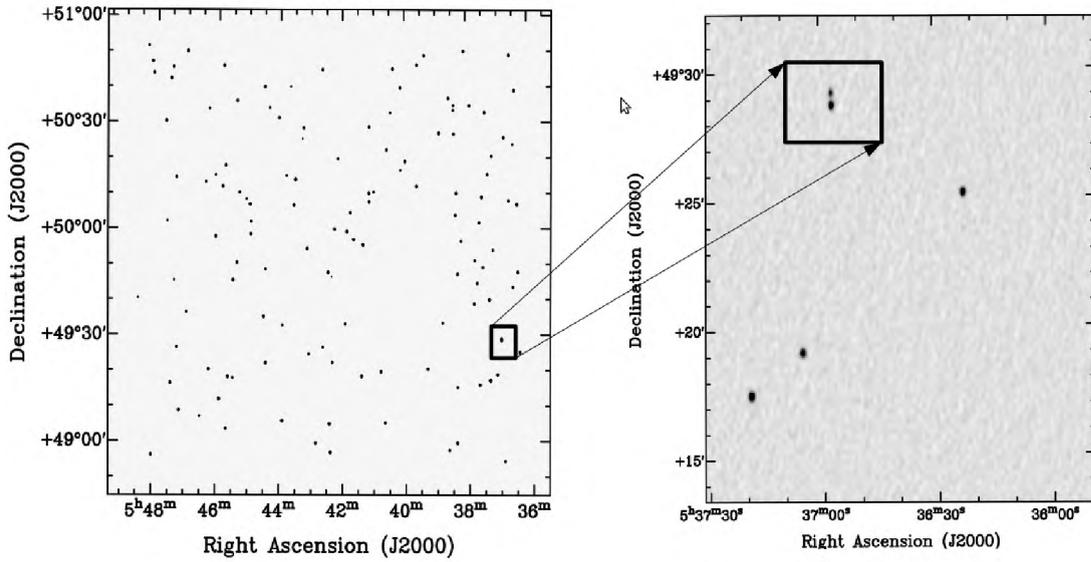


FIGURE 4.8: Clean radio map exhibiting 70% of unmodelled sources out of a 200-source sky model. The left image presents a full map of the sky with the red circle denoting the two overestimated sources. A magnified version of the squared part is found on the right-hand side.

Weighted averaging was therefore preferred, whereby the inverse of the interpolated variance σ_{int}^2 was used as the weighting factor. The interpolated standard deviation σ_{int} was acquired by performing B -spline interpolation on the SE associated with the suppression estimates $\hat{\beta}_3$. The weighted mean $\bar{\beta}_{\text{int}}$ over the 100 sky samples was then equal to

$$\bar{\beta}_{\text{int}} = \frac{\sum \gamma_i (\hat{\beta}_{\text{int}})_i}{\sum \gamma_i}, \quad (4.18)$$

where $\gamma_i = \frac{1}{(\sigma_{\text{int}}^2)_i}$ and $(\hat{\beta}_{\text{int}})_i$ was the interpolated source suppression estimate. This weighted averaging was robust against estimates with high SE.

The RMSE associated with $\bar{\beta}_{\text{int}}$ is given by

$$\sqrt{\frac{\sum^n ((\hat{\beta}_{\text{int}})_i - \bar{\beta}_{\text{int}})^2}{n}},$$

with n denoting the number of samples.

The next section discusses the effect of unmodelled flux on the B -spline interpolated suppression estimates.

4.4.1 Suppression with unmodelled flux

In order to investigate how flux density in the calibration sky model affects source suppression, we define the set

$$S_p = \{P_i\}_{i \in \mathbb{Z}}; \quad P_i = 5i; \quad 1 \leq i \leq 18. \quad (4.19)$$

A randomly generated sky model was introduced to the calibration pipeline, described in Section 4.2, with P taken to be the first element of S_p . The visibilities were therefore calibrated with a model containing 95% of the sources. StEFCal tries to compensate for the missing flux that can be associated with the unmodelled sources, which is 95% of the total amount of source. This experiment was repeated for all elements of S_p . The resulting 18 OLS suppression estimates were then interpolated using the B -spline approach (Chunshi et al. (2012)).

The procedure was carried out on the 100 random sky models, each generating 18 OLS suppression estimates. The interpolated values from the suppression estimates were then averaged over the 100 samples using Eq. 4.18.

The entire process discussed above was individually done for two different conditions namely:

1. The signal was assumed to be solely contaminated by the antenna-based gains.
2. The brightest source in each sky model was deliberately subjected to DDEs by adding a dE tag to it. Differential gains were therefore solved for that particular source.

A similar analysis was performed on 200-source sky models. The suppression outcomes for both 100-source and 200-source sky models are presented in Fig. 4.9. The suppression seems to be steady with increasing unmodelled flux in both cases stated above. However, the introduction of DDEs on a single source makes the suppression worse. The fact that some of the sources are overestimated causes instability in the suppression estimates.

For a quick investigation, the overestimated sources or outliers were removed from the set of source measurements. In this context, outliers refer to sources with greater recovered flux densities than their respective nominal flux densities. Source suppression was then estimated from the new set of measurements and the new estimates were interpolated using the B -spline technique. The mean suppression values evaluated using Eq. 4.18 can be observed from Fig. 4.10. The plots displayed in Fig. 4.10 suggest that it was the outliers that were primarily responsible for causing the fluctuations observed in Fig. 4.9.

The cause of the large variance associated with the differential gains is still unclear to us.

Moreover, the suppression values become less trustworthy as a function of unmodelled flux. The error associated with the antenna-based gains is much smaller than that associated with the differential gains. Surprisingly, Fig. 4.10 shows a decrease for the antenna gains. The increase in degrees of freedom due to the high amount of missing source components from the calibration model accounts for this behaviour.

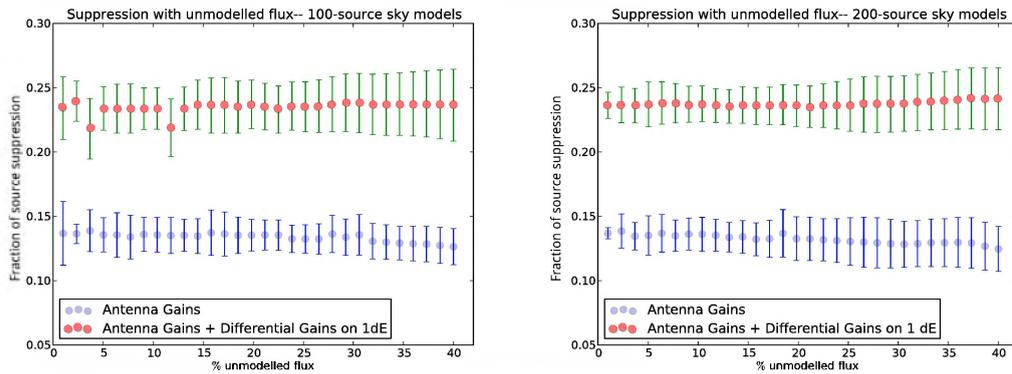


FIGURE 4.9: Source suppression as a function of unmodelled flux. The plot on the left-hand side displays the variation for 100-source sky models while the plot on the right-hand side was created by employing the 200-source sky models. All data points were considered.

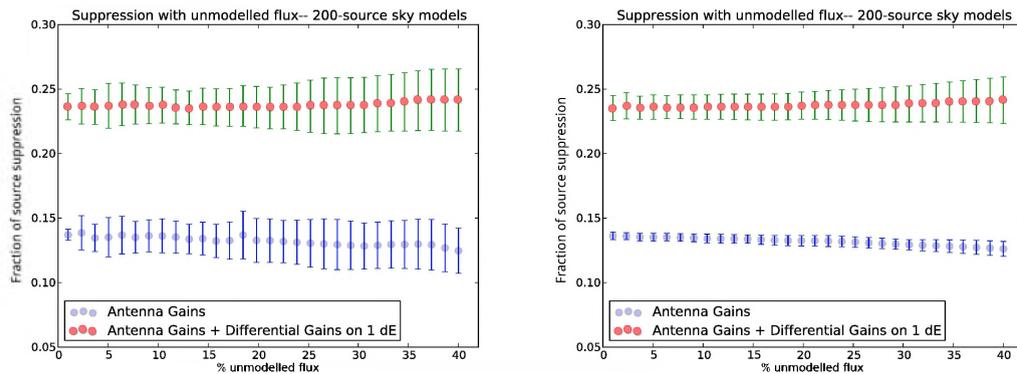


FIGURE 4.10: Source suppression as a function of unmodelled flux. The plot on the left-hand side displays the variation for 100-source sky models while the plot on the right-hand side was created by employing the 200-source sky models. Outliers were excluded.

Another observation of Fig. 4.9 and Fig. 4.10 is the negligible difference in source suppression between the 100-source and 200-source sky models. Nevertheless, the suppression values for 200-source sky models look more reliable than those of the 100-source sky models.

The drastic increase in source suppression with the inclusion of DDEs is discussed next.

4.4.2 Analysis of differential gains, ΔE

The proliferation in degrees of freedom increases source suppression. We investigate how source suppression scales with growth in the direction-dependent solutions. A fixed calibration sky model is decided upon, which does not include the 10 faintest sources present in the nominal sky. We define a solution set

$$N = \{N_i\}_{i \in \mathbb{Z}}; \quad N_i = 5i + 1; \quad 0 \leq i \leq 11. \quad (4.20)$$

For the first index $i = 1$, $N_i = 6$, which implies that the solution was set to six. In the MeqTrees package, the solution is set by using the options ‘`stefcal_diffgain.timesmooth`’. The differential gains, ΔE were thus solved for a time interval of $6 \times 30s = 180s$. If the direction-dependent solutions (dE solutions) were set to five, then ΔE were applied to the first five brightest sources for the different time intervals in N .

The experiment described above was carried out on the 100 samples of 100-source and 200-source sky models. The resulting fluxes were estimated by the OLS estimator. The estimates were then B -spline interpolated followed, by a weighted averaging over the 100 samples using Eq. 4.18. Fig. 4.11 shows the scaling of estimated weighted source suppression with the number of direction-dependent solutions. It shows an increase in the suppression rate with increasing direction-dependent solutions, as predicted, but for larger solution intervals, the suppression rate decreases.

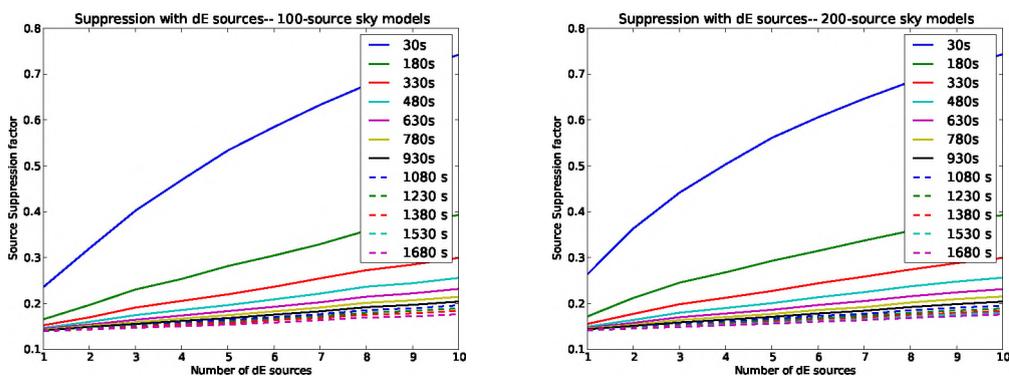


FIGURE 4.11: Source suppression as function of the number of direction-dependent solutions. The legend in the upper right-hand corner exhibits the solution interval in seconds (s).

In general, direction-dependent solutions are solved for longer time intervals than direction-independent gains. Larger solution intervals imply an extra summation over the discrete

time intervals, while solving for ΔE . The additional summation averages the solutions over discrete time intervals. Fig. 4.12 portrays the suppression rate as a function of the ΔE solution interval. The source suppression seems to behave asymptotically with respect to the solution interval, approaching the suppression factor yielded only with gains solutions.

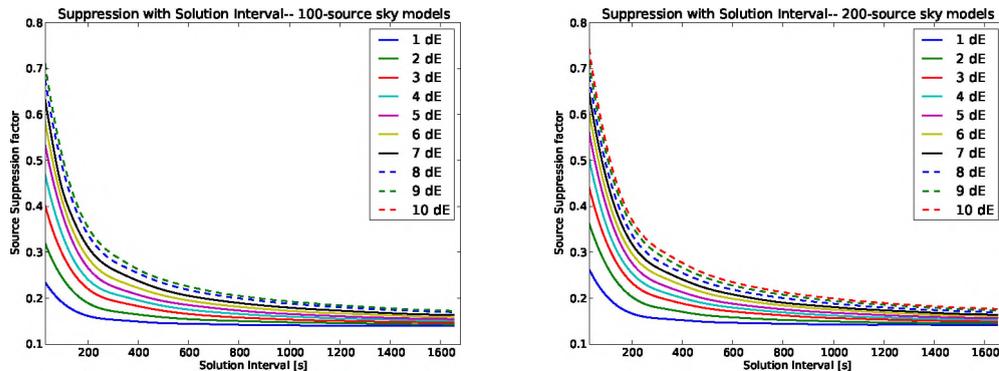


FIGURE 4.12: Source suppression as function of the solution interval. The legend in the upper right-hand corner exhibits the number of direction-dependent solutions.

We are now equipped with the following empirical models:

1. Source suppression as a function of unmodelled flux (flux density model),
2. Source suppression as a function of ΔE solution time interval (solution interval model),
3. Source suppression as a function of direction-dependent solutions (dE solution model).

The accuracy of the empirical models listed above is verified using the realistic 3C 147 field.

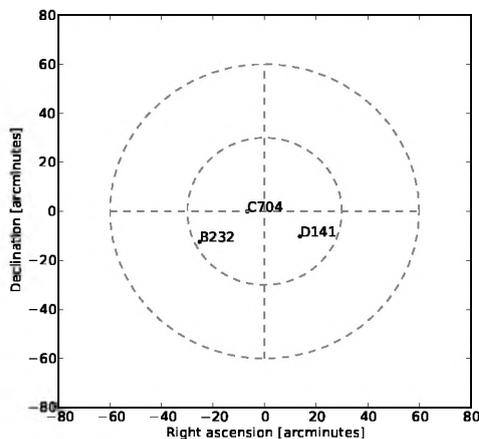
4.5 Observations with 3C 147

The 3C 147 field has a quasar³ of around 22 Jy, with equatorial coordinates equal to $(5^h 42^m 36^s, 49^\circ 51' 07'')$. The light it emits takes around 5.1 billion years to reach the earth (a redshift of 0.545). Its world record DR 1,600,000:1, with 300 sources in place, provides a favourable platform for DIEs and DDEs investigations (Smirnov (2011d)).

³Quasars are the most distant, energetic and luminous active galactic nuclei. They are radio sources that appear like point sources.

4.5.1 Analysis of source suppression

For this investigation, the python calibration pipeline is initialised with the field 3C 147. Originally, the sky model has 300 sources as stated above, but the source count is scaled down to 173 through clustering. The clusters are regarded as point sources throughout this analysis. There are three clusters, situated in the proximity of the phase centre that



ID	RA	DEC	Flux
C270	5h41m57.34s	+49°54'04.32''	51.8 mJy
B232	5h40m43.39s	+49°41'43.08''	42.3 mJy
D141	5h43m18.96s	+49°43'55.56''	21.8 mJy

FIGURE 4.13: Position of sources (relative to the phase centre) with DDEs. The table displays the right ascension (RA) and the declination (DEC) of the sources along with corresponding flux.

are more likely prone to DDEs: C270, B232 and D141 of 51.8 mJy, 42.3 mJy and 21.8 mJy respectively, portrayed in Fig. 4.13.

The change in source suppression with unmodelled flux, the ΔE solution time interval and the number of DDE with the 3C 147 are then studied.

1. Source suppression as a function of unmodelled flux

The pipeline was executed using calibration models constructed by taking out different source percentages P from S_p , defined in Subsection 4.4.1. StEFCal was applied to solve for both antenna-based gains and differential gains. Both the antenna and differential gains were solved for a time interval of 30 s. Differential gains were only applied to C270. OLS estimation was performed on the flux densities resulting from the individual unmodelled flux values.

From the plot on the left-hand side of Fig. 4.14 the suppression seems to wobble around specific suppression factors (0.13-10.14 for antenna gains and 0.24-0.25

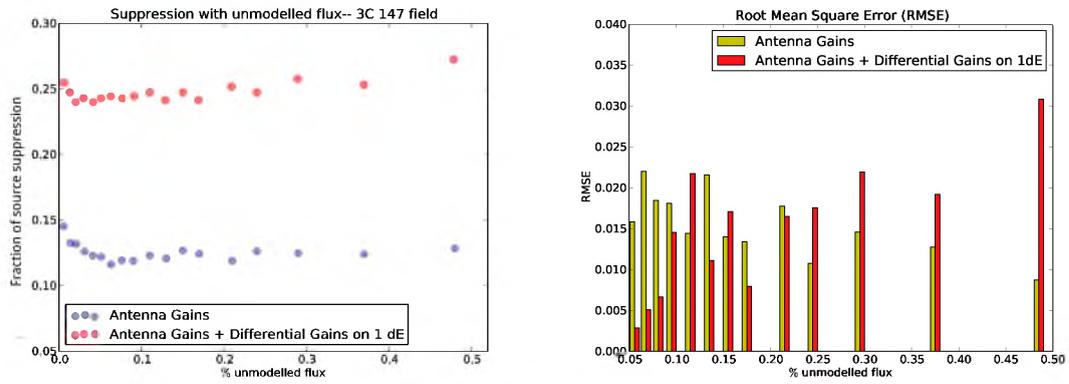
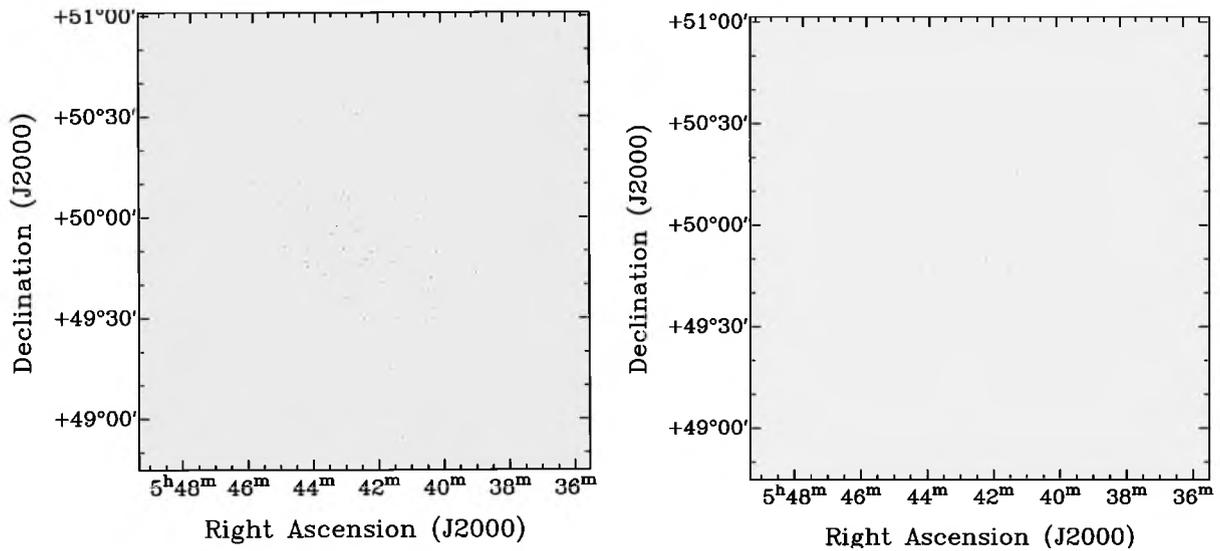


FIGURE 4.14: The plot on the left-hand side shows the source suppression as a function of unmodelled source for the 3C 147 field. The bar chart to the right represents the RMSE involved in the flux density model with respect to the 3C 147 field.



(A) Residual clean map after correcting for (B) Residual clean map after correcting for the antenna-based gains additional DDEs.

FIGURE 4.15: Residual images after calibrating for DDEs and DDEs, using a calibration model containing 50% of the nominal sources

for antenna and differential gains). The drastic increase in suppression with the addition of a DDE can be seen in Fig. 4.15b. The sidelobes exhibited by Fig. 4.15a are also suppressed in Fig. 4.15b. The results extracted from the 3C 147 field are then compared with the 200-source flux density results in Subsection 4.4.1 with respect to unmodelled flux. The corresponding RMSE is calculated using

$$\text{RMSE}_i = \sqrt{(\hat{\beta}_i - (\hat{\beta}_{\text{int}})_i)^2},$$

where $\hat{\beta}_i$ is the suppression estimate obtained with the 3C 147 field and $(\hat{\beta}_{\text{int}})_i$ is the corresponding suppression average from the derived flux density model. The

RMSE displayed on the right-hand side of the plot in Fig. 4.14 is relatively small.

2. Source suppression as a function of ΔE solution time interval and number of direction-dependent solutions.

First, we agreed upon a P value which corresponds to the unmodelled flux considered in Subsection 4.4.2 while designing the solution interval and dE solution models. We defined the solution set

$$N = \{N_i\}_{i \in \mathbb{Z}}; \quad N_i = 5i + 1; \quad 0 \leq i \leq 49. \quad (4.21)$$

Direction-dependent solutions were only applied to C270 for the various solution intervals in N . The same process was then followed by applying differential solutions to C270 and B232. Finally ΔE was solved for all three dE sources illustrated in Fig. 4.13.

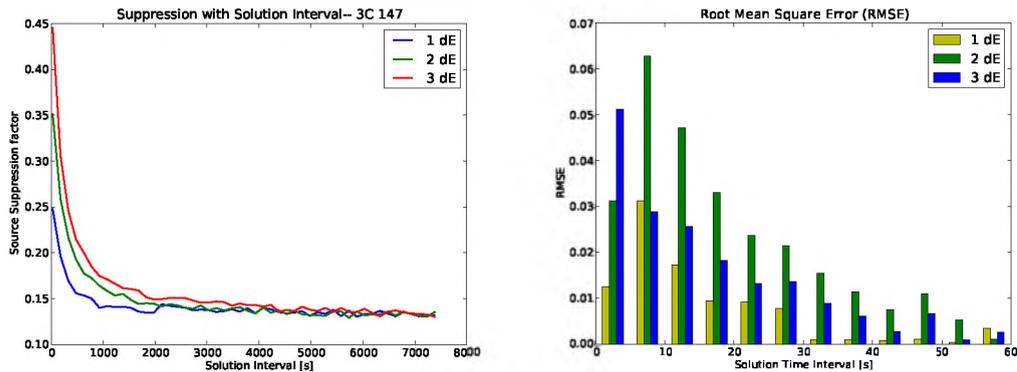


FIGURE 4.16: The plot on the left hand side shows the source suppression as functions of ΔE solution time interval and the number of sources subjected to DDEs for the 3C 147 field. The bar chart to the right represents the root mean square error involved in the flux density model with respect to the 3C 147 field.

The plot to the left in Fig. 4.16 pictures the same asymptotic behaviour as that of source suppression with respect to ΔE solution time interval, and the increasing source suppression with respect to the direction-dependent solutions, seen in Fig. 4.12. The RMSE associated with the 200-source solution interval and dE models derived in Subsection. 4.4.2 is found on the right-hand side plot in Fig. 4.16. Note that the unmodelled flux percentage used for this experiment is 0.11.

4.6 Conclusion

The main outcome of this study is that no matter the amount of flux StEFCal needs to resolve for, the suppression rate is almost fixed (the changes are negligible) for a given interferometer. However, the number of sources subjected to DDEs greatly influences source suppression because of the increase in the degrees of freedom. The degrees of freedom can be restricted by augmenting the number of solution intervals. Increasing the solution interval does help in reducing source suppression to a great extent, but it reaches an asymptote. The asymptote is approximately equal to the suppression factor obtained while calibrating only for the antenna-based gains.

4.7 Summary

The chapter models the variation of the source suppression with

- (i) unmodelled flux,
- (ii) ΔE solution time interval,
- (iii) the number of direction-dependent solutions,

where source suppression refers to the loss in spectral flux density while solving for DDEs and DDEs via StEFCal and imaging.

Chapter 5

General Conclusion

“Astronomy – Exploring the vastness
of the Universe”

–Anonymous

The goal of this thesis is threefold. The first objective was to construct a two-source theoretical framework based on ALS calibration that predicts the ghost distribution, illustrated in Fig. 1.3 and Fig. 1.4, marking the positions and the relative intensities of the ghost sources. The ghost distribution was found to be a fundamental aspect of selfcal and the ghost sources will unfailingly appear when calibrating with an incomplete sky model. Based upon the regularly spaced antenna configuration of WSRT, it was found that ghosts occur at integer multiples of $\frac{s_0}{\phi_0}$, where s_0 denotes the l and m coordinates of the unmodelled source and ϕ_0 represents the common linear factor associated with all baselines.

The predicted ghost amplitudes are assumed to be directly proportional to the flux A_2 of the contaminator for comparatively smaller A_2 values. When A_2 is significantly high, the dominator comes into play, that is, the flux A_1 of the dominator affects the ghost amplitudes. Theoretically, the position of the contaminator does not influence the ghost intensities, but in practice they do. During an interferometric observation, sources (both real and fake) in the dirty images are accompanied by sidelobes caused by the improper sampling of uv points. This alters the flux measurements, which explains the mismatch between the theoretical and simulated results.

The spacing between the ghost sources ($\frac{s_0}{\phi_0}$) decreases with smaller s_0 , but the size of ghost sources remains the same, since they are similar to the PSF size. There exists a critical s_0 beyond which the ghost sources coalesce into a line connecting the sources. Ghosts are generally visible in poorly calibrated maps. With a more accurate calibration, they disappear, which explains why radio astronomers often overlook this issue. However, unaware of the driving factor, the flux suppression ghost (the ghost sitting on top of the unmodelled source) has been indirectly studied as source suppression. Source suppression refers to the loss in flux density of a radio map.

The second objective of this study was to link the degree of incompleteness of a calibration sky model with the amount of flux lost during calibration. It was observed that source suppression has almost no dependence on the unmodelled flux. In hindsight, the suppression rate of a radio interferometer is nearly constant with respect to the calibration sky model, but it differs depending on the array being used. It is worth mentioning that, as the source components in the calibration sky model decreases, the estimated suppression values can no longer be trusted because of large errors associated with them.

While calibrating for direction-independent gains using a calibration sky model whereby the required amount of source components is not provided, the calibration algorithm loses its capability to fit in correctly for the missing flux. This accounts for the bend observed when a large amount of sources was removed from the calibration sky model. Many sources were overestimated, which led to a decrease in source suppression.

The final objective of this thesis was to explore the effect of differential gains on source suppression. Source suppression was therefore investigated as functions of the number of sources subjected to DDEs and the solution time interval. The loss in flux density increases significantly as the number of sources subjected to DDEs increases. Suppression can be reduced by extending the solution intervals. However, a threshold exists beyond which the suppression remains constant. The suppression rate of the direction-independent gain calibration can be approached

with longer solution intervals.

A future study could focus on finding an analytic expression that predicts the solution interval threshold to reach the suppression rate of the direction-independent gain calibration. Different weighting schemes could also be applied and the suppression rates can be compared. This experiment has been carried out on a noise-free data. Another interesting investigation would be incorporating thermal noise into the data.

Appendix A

APPENDIX A: Lemmas and Propositions

Definition A.1 (Regularly-spaced array). Let us pick a coordinate system with origin at the first antenna position $\mathbf{u}_1 = 0$. We shall call a set of antenna positions $\{\mathbf{u}_p\}$ *regularly-spaced* if there exists a *common quotient baseline* (CQB) \mathbf{b}_0 such that each antenna position is an integer multiple of \mathbf{b}_0 , i.e. that $\mathbf{u}_p = \phi_p \mathbf{b}_0$, with ϕ_p being a whole number. We will also require that \mathbf{b}_0 is the largest such baseline (equivalently, the greatest common divisor of $\{\phi_p\}$ is 1).

Definition A.2 (Array geometry matrix). The array geometry matrix Φ is an $n \times n$ integer matrix with elements $\phi_{pq} = \phi_q - \phi_p$.

Definition A.3 (Extrapolated visibility matrix). Let $(\mathbf{b}) : \mathbb{R}^2 \rightarrow \mathbb{C}^{n \times n}$ be an $n \times n$ Hermitian function-valued matrix with entries

$$r_{pq}(\mathbf{b}) = r(\phi_{pq}\mathbf{b}), \quad (\text{A.1})$$

where $r(\mathbf{b})$ is given by Eq. 3.7, ϕ_{pq} is given by the array geometry matrix Φ , $\mathbf{b} = (u, v)$ and $\mathbf{s}_0 = (l_0, m_0) \neq 0$ are real two-vectors, $A_1 = 1$, and $0 < A_2 < 1$.

Proposition A.4. *Let $\mathcal{R}(\mathbf{b}) : \mathbb{R}^2 \rightarrow \mathbb{C}^{n \times n}$ be an $n \times n$ Hermitian function-valued matrix with entries*

$$r_{pq}(\mathbf{b}) = r(\phi_{pq}\mathbf{b}).$$

Then

$$\text{Rank}(\mathcal{R}(\mathbf{b})) \leq 2,$$

and its eigenvalues are given by

$$\frac{n(A_1 + A_2)}{2} \pm h, \quad (\text{A.2})$$

where $h = \frac{1}{2} \sqrt{[n^2 - 4 \binom{n}{2}] |A_1 + A_2|^2 + \kappa}$ and $\kappa = 4 \sum_{p < q} \mathcal{R}(A_1^2 + A_2^2 + 2A_1 A_2 \cos(2\pi \phi_{pq} \mathbf{b} \cdot \mathbf{s}_0))$.

Proof. The rank of $\mathcal{R}(\mathbf{b})$ is less than two as per Lemma A.8–A.10 or it follows trivially from the property that given any two matrices \mathbf{A} and \mathbf{B} , $\text{Rank}(\mathbf{A} + \mathbf{B}) < \text{Rank}(\mathbf{A}) + \text{Rank}(\mathbf{B})$ since

$$\begin{aligned} \text{Rank}(\mathcal{R}) &= \text{Rank}(\{A_1\}_{pq} + \{A_2 e^{2\pi i \mathbf{b}_{pq} \cdot \mathbf{s}_0}\}_{pq}) \\ &< \text{Rank}(\{A_1\}_{pq}) + \text{Rank}(\{A_2 e^{2\pi i \mathbf{b}_{pq} \cdot \mathbf{s}_0}\}_{pq}) \\ &< 1 + 1 = 2 \end{aligned}$$

Since the rank of $\mathcal{R}(\mathbf{b})$ does not exceed two, as stated in Ikramov (2009), its characteristics equation is

$$\left(\lambda^2 - \text{tr}(\mathcal{R}(\mathbf{b}))\lambda + \sum_{p < q} \begin{vmatrix} r_{pp} & r_{pq} \\ r_{qp} & r_{qq} \end{vmatrix} \right) \lambda^{n-2} = 0. \quad (\text{A.3})$$

Solving for λ in Eq. A.3 gives Eq. A.2. □

Proposition A.4 states that the rank of $\mathcal{R}(\mathbf{b})$ is two and gives an analytic expression of its largest eigenvalue, $\lambda(\mathbf{b})$ (Eq. A.2). The expression of $\lambda(\mathbf{b})$ is also used in Proposition A.5 and Lemma A.16. Proposition A.4 has three direct dependencies, namely Lemma A.8, Lemma A.9 and Lemma A.10. Lemma A.8 gives the properties that a matrix must have so that its rank does not exceed $k \in \mathbb{N}$. Lemma A.9 and Lemma A.10 show that $\mathcal{R}(\mathbf{b})$ has the required properties so that its rank does not exceed two. The validity of Lemma A.9 and Lemma A.10 follows from Lemma A.13, which gives the mathematical properties of Φ .

Proposition A.5. *The entries $g_{pq}(\mathbf{b})$ of $\mathcal{G}(\mathbf{b})$ are differentiable Hermitian functions given by*

$$g_{pq}(u, v) = g_{pq}\left(u + \frac{j}{l_0}, v + \frac{k}{l_0}\right),$$

and

$$g_{pq}\left(u, -\frac{l_0}{m_0}u + c\right) = g_{pq}(0, c),$$

for all $j, k \in \mathbb{Z}$ and $u, v, c \in \mathbb{R}$.

Proof. By Lemma A.14 and Eq. A.2

$$\begin{aligned} \mathcal{G}(\mathbf{b})\left(u + \frac{j}{l_0}, v + \frac{k}{l_0}\right) &= \lambda(u, v) \mathbf{x}(u, v) \mathbf{x}^H(u, v) \\ &= \mathcal{G}(\mathbf{b})(u, v), \end{aligned} \quad (\text{A.4})$$

for all $j, k \in \mathbb{Z}$. This implies $g_{pq}(u, v) = g_{pq}\left(u + \frac{j}{l_0}, v + \frac{k}{l_0}\right)$, $\forall j, k \in \mathbb{Z}$. $g_{pq}\left(u, -\frac{l_0}{m_0}u + c\right) = g_{pq}(0, c)$, $\forall u, c \in \mathbb{R}$ (by Lemma A.15). The differentiability of $g_{pq}(\mathbf{b})$ is deduced by Lemma A.16.

From Eq. 3.9, we know that $g_{pq}(u, v)$ is the best possible fit of $r_{pq}(u, v)$ (using LS). Given that \mathcal{R} and \mathcal{G} are Hermitian matrices and $\mathcal{R}(-\mathbf{b}) = \overline{\mathcal{R}(\mathbf{b})} \Rightarrow \mathcal{G}(-\mathbf{b}) = \overline{\mathcal{G}(\mathbf{b})}$, the listed statements trivially follow:

- (a) The best possible fit of $r_{pq}(-u, -v) = \bar{r}_{pq}(u, v)$ is $g_{pq}(-u, -v)$.
- (b) The best possible fit of $r_{qp}(u, v) = r_{pq}(-u, -v) = \bar{r}_{pq}(u, v)$ is $g_{qp}(u, v) = \bar{g}_{pq}(u, v)$.

The above statements imply that $g_{pq}(-u, -v) = \bar{g}_{pq}(u, v)$. □

Proposition A.5 shows that the elements $g_{pq}(\mathbf{b})$ of $\mathcal{G}(\mathbf{b})$ are periodic, effectively one-dimensional, differentiable, Hermitian functions. The properties of $g_{pq}(\mathbf{b})$ follow from Lemma A.14 (periodicity), Lemma A.15 (one-dimensionality) and Lemma A.16 (differentiability). Lemma A.16 is a consequence of Rellich's theorem.

Proposition A.6. *Each element $g_{pq}(\mathbf{b})$ of $\mathcal{G}(\mathbf{b})$ can be written as the following sum*

$$g_{pq}(\mathbf{b}) = \sum_{j=-\infty}^{\infty} c_j e^{2\pi i j \mathbf{b} \cdot \mathbf{s}_0}, \quad (\text{A.5})$$

that is

$$g_{pq}(u, v) = \sum_{j=-\infty}^{\infty} c_j e^{2\pi i j (ul_0 + vm_0)}, \quad (\text{A.6})$$

where

$$c_j = \mu \int_{-\frac{1}{2|m_0|}}^{\frac{1}{2|m_0|}} \int_{-\frac{1}{2|l_0|}}^{\frac{1}{2|l_0|}} g_{pq}(u, v) e^{-2\pi i j (ul_0 + vm_0)} du dv, \quad (\text{A.7})$$

with $\mu = |l_0||m_0|$ and $c_j \in \mathbb{R}$.

Proof. Since g_{pq} is a differentiable periodic function in \mathbb{R}^2 (Proposition A.5), consider the standard Fourier series expansion

$$g_{pq}(u, v) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{jk} e^{2\pi i (jl_0 u + km_0 v)}, \quad (\text{A.8})$$

with

$$c_{jk} = \mu \int_{-\frac{1}{2|m_0|}}^{\frac{1}{2|m_0|}} \int_{-\frac{1}{2|l_0|}}^{\frac{1}{2|l_0|}} g_{pq}(u, v) e^{-2\pi i(jul_0 + kvm_0)} dudv \quad (\text{A.9})$$

and $\mu = |l_0||m_0|$.

Since g_{pq} is Hermitian (Proposition A.5), the coefficients (c_{jk}) are real numbers. Fix $c \in \mathbb{R}$. Note that $g_{pq}(u, -\frac{l_0}{m_0}u + c) = g_{pq}(0, c)$, $\forall u \in \mathbb{R}$ (Proposition A.5). Denote this constant $g_{pq}(0, c)$ by $\alpha \in \mathbb{C}$.

Evaluating the ‘‘diagonal’’ of the series in Eq. A.8, at $(u, v) = (u, -\frac{l_0}{m_0}u + c)$, results in another constant (i.e. independent of u), say $\beta \in \mathbb{C}$.

Thus

$$h(u) \equiv \sum_{j=-\infty}^{\infty} \sum_{\substack{k=-\infty \\ k \neq j}}^{\infty} c_{jk} e^{2\pi i(j-k)l_0u + km_0c} = \alpha - \beta. \quad (\text{A.10})$$

So, setting $n = j - k$, we get

$$\alpha - \beta = \sum_{j=-\infty}^{\infty} \sum_{\substack{k=-\infty \\ k \neq j}}^{\infty} c_{jk} e^{2\pi ijk m_0c} \cdot e^{2\pi i(j-k)l_0u}, \quad (\text{A.11})$$

$$= \sum_{j=-\infty}^{\infty} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} c_{j, j-n} e^{2\pi i(j-n)m_0c} \cdot e^{2\pi inl_0u}, \quad (\text{A.12})$$

$$= \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} d_n e^{2\pi inl_0u}, \quad (\text{A.13})$$

where $d_n = \sum_{j=-\infty}^{\infty} c_{j, j-n} e^{2\pi i(j-n)m_0c}$. Thus $h(u)$, which is a one-dimensional Fourier series without a constant term, is a constant $\alpha - \beta$. This is only possible if $\alpha - \beta = 0$ and $d_n = 0$ whenever $n \neq 0$. This is again only possible if each $c_{j, j-n} = 0$ whenever $n \neq 0$. Thus $c_{jk} = 0$ whenever $j \neq k$.

Therefore, in the two-dimensional Fourier series expansion of g_{pq} , only the terms with $j = k$ contribute. \square

Proposition A.6 states that $g_{pq}(\mathbf{b})$ can be expressed as an effectively one-dimensional Fourier series and follows from Proposition A.5.

Note that Proposition A.6 can also be stated using $e^{-2\pi i j \mathbf{b} \cdot \mathbf{s}_0}$ instead of $e^{2\pi i j \mathbf{b} \cdot \mathbf{s}_0}$ in which case Eq. A.7 becomes

$$c_j = \mu \int_{-\frac{1}{2|m_0|}}^{\frac{1}{2|m_0|}} \int_{-\frac{1}{2|l_0|}}^{\frac{1}{2|l_0|}} g_{pq}(u, v) e^{2\pi i j (ul_0 + vm_0)} dudv, \quad (\text{A.14})$$

with $\mu = |l_0||m_0|$ and $c_j \in \mathbb{R}$. It is also important to note that Proposition A.6 assumes that $l_0 \neq 0$ and $m_0 \neq 0$. When either l_0 or m_0 is zero the derivation

simplifies and becomes one-dimensional. To avoid clutter the derivation of the one-dimensional case is not repeated here.

Proposition A.7. *Let $h(u, v) = \frac{1}{g_{pq}(u, v)}$, then $h(u, v)$ will be a differentiable Hermitian function if $g_{pq}(u, v) \neq 0, \forall u, v \in \mathbb{R}$. Moreover, $h(u + \frac{j}{l_0}, v + \frac{k}{l_0}) = h(u, v)$ and $h(u, -\frac{l_0}{m_0}u + c) = h(0, c), \forall j, k \in \mathbb{Z}$ and $u, v, c \in \mathbb{R}$.*

Proof. To see that h has the same period as g_{pq} , notice that for any $j, k \in \mathbb{Z}$ we have $h(u + j, v + k) = (g_{pq}(u + j\frac{1}{l_0}, v + k))^{-1} = g_{pq}(u, v)^{-1} = h(u, v)$. Similarly $h(u, u - \frac{l_0}{m_0}u + c) = h(0, c) \forall u, c \in \mathbb{R}$. To see that h is Hermitian, recall that complex conjugation satisfies $\overline{\frac{1}{z}} = \frac{1}{\overline{z}}$. Thus one computes $h(-u, -v) = \frac{1}{g_{pq}(-u, -v)} = \frac{1}{\overline{g_{pq}(u, v)}} = \overline{\left(\frac{1}{g_{pq}(u, v)}\right)} = \overline{h(u, v)}$. Finally, $h(u, v)$ is also differentiable since

$$\frac{\partial h(u, v)}{\partial u} = -\frac{\frac{\partial g_{pq}(u, v)}{\partial u}}{g_{pq}^2(u, v)}, \quad (\text{A.15})$$

$$\frac{\partial h(u, v)}{\partial v} = -\frac{\frac{\partial g_{pq}(u, v)}{\partial v}}{g_{pq}^2(u, v)}, \quad (\text{A.16})$$

$$(\text{A.17})$$

exist ($g_{pq}(u, v) \neq 0$ by assumption). \square

Proposition A.7 shows that the elements of $\mathcal{G}^\top(\mathbf{b})$ are also periodic, effectively one-dimensional, differentiable, Hermitian functions. Proposition A.6 and Proposition A.7 therefore imply that the elements of $\mathcal{G}^\top(\mathbf{b})$ also have a one-dimensional Fourier-series representation.

Lemma A.8. *Let \mathbf{A} be symmetric or Hermitian. If all principal submatrices having $k + 1$ rows or $k + 2$ rows are singular, the rank of \mathbf{A} does not exceed k (Perlis, 1952).*

Lemma A.9. *All 3×3 function-valued principal submatrices of $\mathcal{R}(\mathbf{b})$ are singular.*

Proof. Because of the construction of $\mathcal{R}(\mathbf{b})$ all 3×3 function-valued principal submatrices of $\mathcal{R}(\mathbf{b})$ have the following form (see Lemma A.13).

$$\begin{bmatrix} A_1 + A_2 & A_1 + A_2 e^{-2\pi i a b \cdot \mathbf{s}_0} & A_1 + A_2 e^{-2\pi i A b \cdot \mathbf{s}_0} \\ A_1 + A_2 e^{2\pi i a b \cdot \mathbf{s}_0} & A_1 + A_2 & A_1 + A_2 e^{-2\pi i b b \cdot \mathbf{s}_0} \\ A_1 + A_2 e^{2\pi i A b \cdot \mathbf{s}_0} & A_1 + A_2 e^{2\pi i b b \cdot \mathbf{s}_0} & A_1 + A_2 \end{bmatrix}, \quad (\text{A.18})$$

where $A = a + b$ and $a, b \in \mathbb{N}$. The determinant of the matrix in Eq. A.18 is equal to zero (Kopp, 2008). \square

Lemma A.10. *All 4×4 function-valued principal submatrices of $\mathcal{R}(\mathbf{b})$ are singular.*

Proof. Because of the construction of $\mathcal{R}(\mathbf{b})$ all 4×4 function-valued principal sub-matrices of $\mathcal{R}(\mathbf{b})$ have the following form (see Lemma A.13):

$$\begin{bmatrix} A_1 + A_2 & A_1 + A_2 e^{-ka} & A_1 + A_2 e^{-kA} & A_1 + A_2 e^{-kC} \\ A_1 + A_2 e^{ka} & A_1 + A_2 & A_1 + A_2 e^{-kb} & A_1 + A_2 e^{-kB} \\ A_1 + A_2 e^{kA} & A_1 + A_2 e^{kb} & A_1 + A_2 & A_1 + A_2 e^{-kc} \\ A_1 + A_2 e^{kC} & A_1 + A_2 e^{kB} & A_1 + A_2 e^{kc} & A_1 + A_2 \end{bmatrix}, \quad (\text{A.19})$$

where $k = 2\pi i \mathbf{b} \cdot \mathbf{s}_0$, $A = a + b$, $B = b + c$, $C = a + b + c$ and $a, b, c \in \mathbb{N}$. The determinant of the matrix in Eq. A.19 is equal to zero. \square

Definition A.11. Let $|\mathbf{A}|_{d+1}$, where $\mathbf{A} \in \mathbb{Z}^{k \times k}$, be defined as $\sum_{p=1}^{n-1} a_{pp+1}$.

Definition A.12. Let \mathcal{A} denote the set of all $m \times m$ principal sub-matrices of Φ , with $m = n - 1$. Let \mathcal{B} denote the set of all $k \times k$ principal sub-matrices of Φ , with $2 \leq k \leq n$.

Lemma A.13. *The array geometry matrix Φ has the following properties:*

- (i) $\phi_{pp} = 0$ (diagonal),
- (ii) $\phi_{pq} \neq 0$ (non-diagonal),
- (iii) $\phi_{pq} > 0$; $\forall q > p$,
- (iv) $\phi_{pq} = -\phi_{qp}$,
- (v) $\gcd(\{\phi_{pq}\}_{q>p}) = 1$.
- (vi) $|\Phi|_{d+1} = \phi_{1n}$.
- (vii) $|\mathbf{B}|_{d+1} = b_{1k}$, $\forall \mathbf{B} \in \mathcal{B}$.

Proof. Property (i) is true since $\phi_{pp} = \phi_p - \phi_p = 0$. Properties (ii)–(iv) follow trivially from the assumption that the antenna positions $\{\mathbf{u}_p\}$ satisfy $\|\mathbf{u}_q\|_2 > \|\mathbf{u}_p\|_2 \forall q > p$. Property (v) is true since $\gcd(\{\phi_{pq}\}_{q>p}) = \gcd(\gcd(\{\phi_{1q}\}), \{\phi_{dq}\}_{q>d, d>1}) = \gcd(\gcd(\{\phi_q\}), \{\phi_{dq}\}_{q>d, d>1}) = \gcd(1, \{\phi_{dq}\}_{q>d, d>1}) = 1$. Property (vi) is true since $|\Phi|_{d+1} = \sum_{p=1}^{n-1} \phi_{pp+1} = \sum_{p=1}^{n-1} \phi_{p+1} - \phi_p = \phi_{1n}$.

Property (vii) can be proven using the following argument. Assume that $\mathbf{A}^j \in \mathcal{A}$ is obtained from Φ by deleting the j -th row and column from Φ (where j was chosen arbitrarily). When calculating $|\mathbf{A}^j|_{d+1}$ three separate cases arise,

- $1 < j < n$: $|\mathbf{A}^j|_{d+1} = \sum_{i=1}^{m-1} a_{i+1}^j = \sum_{\substack{p=1 \\ p \neq j, j+1}}^{n-1} \phi_{pp+1} + \phi_{j, j+2} = \sum_{\substack{p=1 \\ p \neq j, j+1}}^{n-1} \phi_{pp+1} + \phi_{jj+1} + \phi_{j+1, j+2} = \phi_{1n} = a_{1m}$,
- $j = n$: $|\mathbf{A}^j|_{d+1} = \phi_{1n-1} = a_{1m}$,
- $j = 1$: $|\mathbf{A}^j|_{d+1} = \phi_{2n} = a_{1m}$.

The above shows that $|\mathbf{A}|_{d+1} = a_{1m} \forall \mathbf{A} \in \mathcal{A}$ (since j was chosen arbitrarily). Expanding the above derivation by using $1 < t \leq n - 2$ arbitrary deletions yields the required result. \square

Lemma A.14. *Let $\lambda(u, v)$ denote the largest eigenvalue of $\mathcal{R}(u, v)$ and $\mathbf{x}(u, v)$ its associated normalised eigenvector, then $\mathbf{x}(u, v) = \mathbf{x}\left(u + \frac{j}{l_0}, v + \frac{k}{m_0}\right), \forall j, k \in \mathbb{Z}$.*

Proof. Notice that $\forall j, k \in \mathbb{Z}$

$$\mathcal{R}\left(u + \frac{j}{l_0}, v + \frac{k}{m_0}\right) = \mathcal{R}(u, v), \quad (\text{A.20})$$

$$\lambda\left(u + \frac{j}{l_0}, v + \frac{k}{m_0}\right) = \lambda(u, v); \quad (\text{A.21})$$

implying that $\mathbf{x}(u, v) = \mathbf{x}\left(u + \frac{j}{m_0}, v + \frac{k}{l_0}\right), \forall j, k \in \mathbb{Z}$. \square

Lemma A.15. *Let $\lambda(u, v)$ denote the largest eigenvalue of $\mathcal{R}(u, v)$ and $\mathbf{x}(u, v)$ its associated normalised eigenvector, then $\mathbf{x}\left(u, -\frac{l_0}{m_0}u + c\right) = \mathbf{x}(0, c), \forall u, c \in \mathbb{R}$.*

Proof. Notice that $\forall u, c \in \mathbb{R}$

$$\mathcal{R}\left(u, -\frac{l_0}{m_0}u + c\right) = \mathcal{R}(0, c), \quad (\text{A.22})$$

$$\lambda\left(u, -\frac{l_0}{m_0}u + c\right) = \lambda(0, c), \quad (\text{A.23})$$

implying that $\mathbf{x}\left(u, -\frac{l_0}{m_0}u + c\right) = \mathbf{x}(0, c), \forall u, c \in \mathbb{R}$. \square

Lemma A.16. *Let $\lambda(u, v)$ denote the largest eigenvalue of $\mathcal{R}(u, v)$ and $\mathbf{x}(u, v)$ its associated normalised eigenvector. The real function $\lambda(u, v)$ and the function-valued vector $\mathbf{x}(u, v)$ are differentiable.*

Proof. The parameter dimension of $\mathcal{R}(\mathbf{b})$ is effectively one, i.e. $t(t) := \mathcal{R}(u, v), \lambda_t(t) := \lambda(u, v)$ and $\mathbf{x}_t(t) := \mathbf{x}(u, v)$ with $t(u, v) := \mathbf{b} \cdot \mathbf{s}_0$. The entries of $\mathcal{R}_t(t)$ are analytic functions depending on $t \in \mathbb{R}$. Therefore, Rellich's theorem (Lemma A.17) implies that $\lambda(t)$ and $\mathbf{x}(t)$ are analytic (the largest eigenvalue is simple - Eq. A.2) and therefore also differentiable. We can therefore calculate

$$\frac{\partial \lambda(u, v)}{\partial u} = \frac{d\lambda_t(t(u, v))}{dt} \frac{\partial t(u, v)}{\partial u}, \quad (\text{A.24})$$

$$\frac{\partial \lambda(u, v)}{\partial v} = \frac{d\lambda_t(t(u, v))}{dt} \frac{\partial t(u, v)}{\partial v}. \quad (\text{A.25})$$

The above equations imply that the real function $\lambda(u, v)$ is differentiable. A similar argument can be used to prove that $\mathbf{x}(u, v)$ is also differentiable. \square

Lemma A.17 (Rellich's Theorem). *Let $\mathbf{A}(t) : \mathbb{R} \rightarrow \mathbb{C}^{n \times n}$ be a Hermitian function-valued matrix that depends on t analytically.*

(i) *The n roots of the characteristic polynomial of $\mathbf{A}(t)$ can be arranged so that each root $\lambda_j(t)$ for $j = 1, \dots, n$ is an analytic function of t .*

(ii) *There exists an eigenvector $\mathbf{x}_j(t)$ associated with $\lambda_j(t)$ for $j = 1, \dots, n$ satisfying*

(a) $\|\mathbf{x}_j(t)\|_2 = 1 \quad \forall t \in \mathbb{R},$

(b) $\mathbf{x}_j(t)$ *is an analytic function-valued vector of t (Reed and Simon, 1978; Lax, 1996; Kılıç et al., 2011).*

Bibliography

- A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr. *Interferometry and Synthesis in Radio Astronomy*. Wiley, New York, 2 edition, 2001.
- T. J. Pearson and A. C. S. Readhead. Image formation by self-calibration in radio astronomy. *Annual Review of Astronomy and Astrophysics*, 22:97–130, 1984.
- A.J. Boonstra and A.J. van der Veen. Gain calibration methods for radio telescope arrays. *IEEE Trans. Sig. Proc.*, 51(1):25–38, 2003.
- S. J. Wijnholds and A. J. Van der Veen. Self-Calibration of radio astronomical arrays with non-diagonal noise covariance matrix. *Proceedings of the 17th European Signal Processing Conference(EuSipCo)*, pages 24–28, 2009.
- S. van der Tol, B. I. Jeffs, and A. J. van der Veen. Self-calibration for the LOFAR radio astronomical array. *IEEE Trans. Signal Process*, 55:4497 – 4510, 2007.
- H. T. Intema, S. van der Tol, W. D. Cotton, A. S. Cohen, I. M. van Bemmel, and H. J. A. Röttgering. Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme. I. Method description and first results. *Astronomy and Astrophysics*, 501:1185–1205, July 2009. doi: 10.1051/0004-6361.
- T. J. Cornwell, R. Braun, and D. S. Briggs. Deconvolution. In *Synthesis Imaging in Radio Astronomy II*, volume 180, page 151, 1999.
- R. P. Linfield. A modified hybrid mapping technique for VLBI data. *Astronomical Journal*, 92:213–218, 1986.
- T. J. Cornwell and P. N. Wilkinson. A new method for making maps with unstable radio interferometers. *Monthly Notices of the Royal Astronomical Society*, 196: 1067–1086, September 1981.
- A. H. Bridle and F. R. Schwab. Bandwidth and Time-Average Smearing. In G. B. Taylor and C. L. Carilli, and R. A. Perley, editor, *Synthesis Imaging in*

- Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, pages 371–+, 1999.
- I. Martí-Vidal and J.M. Marcaide. Spurious source generation in mapping from noisy phase-self-calibrated data. *Astronomy & Astrophysics*, 480(1):289–295, 2008.
- I. Martí-Vidal, J. C. Guirado, S. Jiménez-Monferrer, and J. M. Marcaide. Atmospheric turbulence in phase-referenced and wide-field interferometric images. Application to the Square Kilometre Array. *Astronomy & Astrophysics*, 517, 2010.
- S. Kazemi and S. Yatawatta. Robust radio interferometric calibration using the t-distribution. *preprint (arXiv:1307.5040)*, 2013.
- T. L. Grobler, C. D. Nunhokee, O. M. Smirnov, A. J. van Zyl, and A. G. de Bruyn. Calibration artefact in radio astronomy.I.Ghost sources in WSRT data. *Monthly Notices of the Royal Astronomical Society*, 439:4030–4047, 2014.
- O. M Smirnov. Ghostbusters: The Unknown Unknowns Of Selfcal. presentation at CALIM2010 conference (Dwingeloo 2010), 2010.
- O. M Smirnov. Solving For Primary Beams, Pointing Errors And The Westerbork Wobble. presentation at CALIM2011 conference (Manchester 2011), 2011a.
- S. Zaroubi. The Epoch of Reionization. *preprint (arXiv:1206.0267v1)*, 2012.
- F. Villaescusa-Navarro, M. Viel, K. K. Datta, and T. R. Choudhury. Modelling the neutral hydrogen distribution in the post-reionization Universe: intensity mapping. 2014.
- J. Delhaize, M. J. Meyer, L. Staveley-Smith, and B. J. Boyle. Detection of HI in distant galaxies using spectral stacking. *Monthly Notices of the Royal Astronomical Society*, 433:1398–1410, August 2013.
- R. J. Sault, J. P. Hamaker, and J. D. Bregman. Understanding radio polarimetry. II. Instrumental calibration of an interferometer array. *Astronomy and Astrophysics Supplement Series*, 117:149–159, May 1996.
- S. Salvini and S. J. Wijnholds. Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications. *Astronomy & Astrophysics*, 571, 2014.
- J. E. Noordam and O. M. Smirnov. The MeqTrees software system and its use for third-generation calibration of radio interferometers. *Astronomy & Astrophysics*, 524:A61, December 2010. doi: 10.1051/0004-6361.

- O. M. Smirnov. Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism. *Astronomy & Astrophysics*, 527:A106, 2011b.
- F. R. Schwab. Adaptive calibration of radio interferometer data. In *1980 Technical Symposium East*, pages 18–25. International Society for Optics and Photonics, 1980.
- S. Yatawatta, S. Kazemi, and S. Zaroubi. GPU accelerated nonlinear optimization in radio interferometric calibration. In *Innovative Parallel Computing (InPar), 2012*, pages 1–6. IEEE, 2012.
- K. Madsen, H. B. Nielsen, and O. Tingleff. *Methods for Non-Linear Least Squares Problems* (2nd ed.), 2004.
- S. Kazemi, S. Yatawatta, S. Zaroubi, P. Lampropoulos, A. G. de Bruyn, L. V. E. Koopmans, J. Noordam, et al. Radio interferometric calibration using the SAGE algorithm. *Monthly Notices of the Royal Astronomical Society*, 414(2):1656–1666, 2011.
- S. Kazemi, S. Yatawatta, and S. Zaroubi. Clustered calibration: an improvement to radio interferometric direction-dependent self-calibration. *Monthly Notices of the Royal Astronomical Society*, 430(2):1457–1472, 2013.
- S. Bhatnagar, T. J. Cornwell, and K. Golap. Solving for the antenna based pointing errors. Memo, National Radio Astronomy Observatory, 2004.
- O. M. Smirnov. Revisiting the radio interferometer measurement equation. II. Calibration and direction-dependent effects. *Astronomy & Astrophysics*, 527: A107, 2011c.
- J. E. Noordam. LOFAR calibration challenges. In Jr. J. M. Oschmann, editor, *Ground-based Telescopes*, volume 5489 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 817, October 2004. doi: 10.1117/12.544262.
- K. Levenberg. A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.
- D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. of App. Math.*, 11(2):431–441, 1963.
- C. J. Lonsdale. Configuration Considerations for Low Frequency Arrays. In N. Kasim, M. Perez, W. Junor, & P. Henning, editor, *From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science*, volume 345 of *Astronomical Society of the Pacific Conference Series*, pages 399–+, December 2005.

- R. A. Sramek and F. R. Schwab. Imaging. In *Synthesis Imaging in Radio Astronomy II*, volume 180, page 151, 1999.
- J.D. Kraus. *Radio Astronomy*. New York: McGraw-Hill, 1996.
- M. Rafferty and N Mohan. PyBDDSM Documentation, 2013.
- L. Chunshi, L. Yu, and Z. Yimin. Study of B-spline Interpolation, Correlation and Inverse Algorithm. *Advances in Automation and Robotics*, 2, 2012.
- O. M. Smirnov. Revisiting the radio interferometer measurement equation. III. Addressing direction-dependent effects in 21cm WSRT observations of 3C 147. *Astronomy & Astrophysics*, 527:A108, 2011d.
- K. D. Ikramov. On the ranks of principal submatrices of diagonalizable matrices. *J. of Math. Sci.*, 157(5):695–696, 2009.
- S. Perlis. *Theory of matrices*. Dover Publications, 1952.
- J. Kopp. Efficient numerical diagonalization of hermitian 3×3 matrices. *Int. J. of Modern Phys. C*, 19(03):523–548, 2008.
- Michael Reed and Barry Simon. *Analysis of operators. Methods of modern mathematical physics IV*. New York: Academic, 1978.
- P.D. Lax. *Linear Algebra and its applications*. John Wiley & sons, 1996.
- Mustafa Kılıç, Emre Mengi, and E Alper Yıldırım. Numerical Optimization of Eigenvalues of Hermitian Matrix Functions. *preprint (arXiv:1109.2080)*, 2011.