Probabilistic Methods for Radio Interferometry Data Analysis

A thesis presented
by
Iniyan Natarajan
to
The Department of Astronomy
for the degree of
Doctor of Philosophy

University of Cape Town
May 2017

Advisors

Dr. Kurt van der Heyden (University of Cape Town)
Prof. Oleg M. Smirnov (Rhodes University / Square Kilometre Array Africa)
Dr. Jonathan Zwart (University of Cape Town / University of the Western Cape)

Declaration of Authorship

I, Iniyan Natarajan, know the meaning of plagiarism, and declare that all of the work in the thesis, save for that which is properly acknowledged, is my own. The thesis has been submitted to the Turnitin module and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.

Signed: Signed by candidate

Date: 8 May 2017

ABSTRACT

Probability theory provides a uniquely valid set of rules for plausible reasoning. This enables us to apply this mathematical formalism of probability, also known as *Bayesian*, with greater flexibility to problems of scientific inference. In this thesis, we are concerned with applying this method to the analysis of visibility data from radio interferometers. Any radio interferometry observation can be described using the Radio Interferometry Measurement Equation (RIME). Throughout the thesis, we use the RIME to model the visibilities in performing the probabilistic analysis.

We first develop the theory for employing the RIME in performing Bayesian analysis of interferometric data. We then apply this to the problem of super-resolution with radio interferometers by performing model selection successfully between different source structures, all smaller in scale than the size of the point spread function (PSF) of the interferometer, on Westerbork Synthesis Radio Telescope (WSRT) simulations at a frequency of 1.4 GHz. We also quantify the change in the scale of the sources that can be resolved by WSRT at this frequency, with changing signal-to-noise (SNR) of the data, using simulations.

Following this, we apply this method to a 5 GHz European VLBI Network (EVN) observation of the flaring blazar CGRaBS J0809+5341, to ascertain the presence of a jet emanating from its core, taking into account the imperfections in the station gain calibration performed on the data, especially on the longest baselines, prior to our analysis. We find that the extended source model is preferred over the point source model with an odds ratio of $10^9 : 1$. Using the flux-density and shape parameter estimates of this model, we also derive the brightness temperature of the blazar (10^{11}–10^{12} K), which confirms the presence of a relativistically boosted jet with an intrinsic brightness temperature lower than the apparent brightness temperature, consistent with the literature. We also develop a Bayesian criterion for super-resolution in the presence of baseline-dependent noise and calibration errors and find that these errors play an important role in determining how close one can get to the theoretical super-resolution limit.

We then proceed to include fringe-fitting, the process of solving for the time and frequency dependent phase variations introduced by the interstellar medium and the Earth's atmosphere, in our probabilistic approach. Fringe-fitting is one of the first corrections made to Very Long Baseline Interferometry (VLBI) observations, and, by extending our method to include simultaneous fringe-fitting and source structure estimation, we will be able to perform end-to-end VLBI analysis using our method. To this end, we estimate source amplitude and fringe-fitting phase terms (phase offsets and delays) on 43 GHz Very Long Baseline Array and 230 GHz Event Horizon Telescope (EHT) simulations of point sources. We then perform model selection on a 5 μas extended Gaussian source (one-fourth the size of the PSF) on a synthetic 230 GHz EHT observation. Finally we incorporate turbulent time-varying phase offsets and delays in our model selection and show that the delays can be estimated to within 10–16 per cent error (often better than contemporary software packages) while simultaneously estimating the extended source structure.

Contents

Acknowledgements

There are many whose guidance, inspiration, and support I have to thank for the knowledge and the experience I have gained over the past few years leading up to and culminating in my PhD.

My supervisors have always been supportive of me and I can only hope that I have done justice to the job they entrusted me with. Kurt van der Heyden has always given me excellent advice, both professional and personal, at critical junctures throughout my PhD. This would not have been possible without him. I am grateful to Kurt for the freedom he gave me in choosing and defining my research topic for the thesis, while helping me navigate through the process.

Likewise, I am lucky to have had access to the expertise of Oleg Smirnov, always quick and insightful in discussions and with his responses, regardless of the time of the day or day of the week, or year, for that matter. I am thankful to Oleg for being the architect of the various projects and collaborations I have had the good fortune to be part of.

I am indebted to Jonathan Zwart for introducing me to the wonderful world of Bayesian probability and ensuring that I had useful contributions to make in applying probability theory to radio astronomy. Jon has consistently been the one person over the years whom I could approach with questions about probability theory, which he would patiently assist me with. Moreover, if the thesis is readable at all, it is due largely to his meticulous dissection of the manuscript.

Two projects, published in peer-reviewed journals, led by Michelle Lochner and Simon Perkins, form part of my research and I am grateful for the experience of collaborating with the respective teams. A third project, leading to another peer-reviewed article in collaboration with Zsolt Paragi, also forms part of the thesis. Working with Zsolt was a pleasure and I hope this is just the beginning. For the final project of the thesis, I worked with Roger Deane, whose scientific interests have opened up new horizons for the probabilistic approach in radio astronomy to conquer. It was absolute joy to make rapid progress from theory to experiments to inferences, and I look forward to many more such successful ventures. I thank him also for his punctilious proofreading of the corresponding chapter.

Ian Stewart, Jonathan Zwart, Trienko Grobler, Simon Perkins, Cyril Tasse, Zsolt Paragi, Sebastian van der Tol, Roger Deane, Martin Kunz, Luke Chamandy, Henrik Junklewitz, and many others have made numerous intangible but essential contributions towards improving my understanding of the subject. Often these took the form of recurring discussions, but in some cases, they were crucial insights or expositions that helped me crystallise amorphous ideas.

The research was sponsored by Patrick Marais of the Department of Computer Science at UCT via the MeerKAT High Performance Computing for Radio Astronomy Programme. The Square Kilometre Array South Africa also provided welcome financial support during the final year. The

computations were carried out either on the UCT HPC Cluster or on the servers located at Rhodes University. The administrative staff of the Departments of Astronomy and Computer Science at UCT and the Department of Physics & Electronics at Rhodes University have been of great help during my stay.

I learnt much of the foundations of radio astronomy during my years working with Ramesh Balasubramanyam and his team at the Raman Research Institute, who introduced me to radio astronomy and set me on the path to a career in science. His support could not have come at a better time for a young computer science undergraduate and I am beholden to him for that.

I am fortunate to have made many friends who ensured that the past two olympiads were fun and oftentimes informative. These include, but are not limited to, Bradley and Chanelle Frank, who, in both Groningen and Cape Town, saw to it that I had no reason to complain, Albert Cañellas i Pagès, Prajwal Kafle, Vinesh Rajpaul, Kosma von Maltitz, Zolile Mguda, Thuso Simon, Millie Hilgart, Solveig Gleser, Derek Chipp, Moses Mogotsi, Rajin Ramphul, Brendon Nayanah, Christina Magoulas, Zara Randriamanakoto, Bishop Mongwane, Abhik Ghosh, Khan Asad, Preshin Moodley, Sam Legodi, Nikki Peceur, and Narges Hatamkhani. I have intellectual debts to many of the above.

Finally, friends back in India who have been my well-wishers from afar include Ganesh Kumar Shanmugam, Kaarthick Subramanian, Bala Viknesh, Karthick Balasundaram, and Rekha Mallikarjuna. Without their constant companionship, the entire PhD experience would have been far less enjoyable, not to mention the visits to India.

My family has provided unwavering support and encouragement throughout this journey. My father, Natarajan Kandasamy, has always led by example how to live a life free of hypocrisy, while my mother, Manimozhi Thiruvenkadam taught me how to ensure objectivity in one's beliefs and actions. My mother, in applying her editorial pencil to the zeroth draft of the thesis, played more than the clichéd role of mother-as-a-layperson, and refined my technical writing, making it palatable to others. My brother, Poovendhan Natarajan, has been a discerning sounding board and has helped me keep my feet on the ground. This work is dedicated to them.

Preface

The universe as we know it, originated 13.8 billion years ago in a cosmic expansion which created all of space and time; the planet we call home, 4.5 billion years ago from accreting gas and dust — leftovers from the formation of the Sun; the *we* I speak of, a mere 200,000 years ago, from primates that parted ways with their arboreal cousins and migrated to grasslands. *Mere?* Can we truly comprehend any time span that is longer than a few thousand years at the most? What about the size of the universe? Travelling at a speed of 300,000 kilometres per second, the speed at which light travels in vacuum, it would take about 93 billion years to go from one end of the observable universe to the other. Again, we hit the limits of human perception; limits that were defined by natural selection that shaped the evolution of our ancestors in the African savannah. That the immense expanse of the cosmos obeys any set of laws at all, that is at least partially expressible and understandable by us, is a stroke of luck for any species that endeavours to discover its origins.

In our efforts to understand Nature, we have so far managed to describe the interactions between all of observable matter (or energy, which is just another manifestation of matter) in the universe in terms of four fundamental fields or forces: the strong force, the weak force, electromagnetism, and gravity. The strong and weak forces govern the interactions between the elementary particles of matter at the scale of the atomic nucleus. At any scale larger than this, our only guiding forces are electromagnetism (or sometimes, *electroweak*, a unification of electromagnetism and the weak force) and gravity. The effects of both these forces can be felt over large distances; infinite, as far as we know. Almost all of what we know about celestial objects comes from observing electromagnetic radiation from space.

Since the time of Newton's unweaving of the rainbow in the 17th century, we have understood much about the nature of electromagnetism. The classical concepts of both *particle* and *wave* must be invoked to get a complete picture of any elementary "particle", and the quantum of electromagnetism, the photon, is no exception to that. Every photon has a wavelength associated with it and hence can be represented as a wave, and this wavelength determines what techniques we employ in intercepting and interpreting the photon. The visible light, which gives rise to the magnificent hues that we see around us, occupies but a tiny fraction of the electromagnetic spectrum. The short wavelength (or high frequency) part of the spectrum comprising gamma rays, X-rays, and ultraviolet rays from outer space is blocked by the upper layers of the atmosphere and hence is observed by telescopes stationed in space. A significant part of the spectrum with wavelengths longer than those of visible light and infrared radiation, consisting of short and long wavelength radio waves, is observable from the surface of the earth and has provided us with evidence of a variety of new celestial phenomena — from high energy objects such as active galaxies, to a window into the childhood of the universe in the form of the cosmic microwave background.

The wave theory of light adequately explains the propagation of electromagnetic waves in the radio frequency range. The phenomenon of interference, in which the observed intensity of waves at a particular point is understood as a superposition of interfering wavefronts, has been put to good use in building networks of individual telescopes, called interferometers. These interferometers combine the wavefronts received by different telescopes in a way that enables us to resolve finer structures in the radio sky, than is possible with a single radio telescope. In this thesis, we are concerned only with the methods for analysing and understanding the nature of radio observations made using interferometers.

Modelling Physical Phenomena

The idea of logic, of reason, of science, has existed long before humans found it necessary to invent those words. *What kind of idea is it?* The kind that refuses to go away if ignored; the kind that rallies and comes back unchanged if lost or suppressed; the kind that, the hundredth time, changes the world. Not revelation, but reason. Having arrived late on the scene, we are forced to reason backwards from what little we can observe of the universe, to answer the question of how it came to be. The information we possess is almost always incomplete, limited by our senses and our location in space and time.

To overcome some of these limitations, we have evolved or fashioned tools that aid and extend our senses. Part of our mental toolkit is our ability to understand the world in terms of *models*. The myriad cultural myths that exist are humanity's attempts to make sense of the world it finds itself in. It is perhaps not surprising that, for instance, what we perceive as optical illusions are subversions of the assumptions built into our brains by evolution, as has been proven by numerous experiments in cognitive science. As faith in inadequate models of the world is eroded, not by disbelief but by doubt, better models spring into existence, illuminating more and more of the secrets Nature has hidden in its darkest recesses.

In his *Ars Conjectandi* (The Art of Conjecturing) published in 1713, Jacob Bernoulli distinguished between deductive logic and inductive logic, with the latter being our only way forward in the face of incomplete information. This is a situation we are often confronted with in everyday life, and for our purposes here, in the natural sciences. Bernoulli, followed by Bayes and Laplace, pioneered the concept of probability as a measurable degree of certainty or plausibility. In this view, commonly termed *Bayesian*, probability represents how much we believe something to be true, having taken into account all available information that is relevant. In 1946, Richard Cox formulated the theory of probability as the basic calculus for logically consistent plausible reasoning; in other words, for *scientific inference*. As Edwin T. Jaynes noted in response to criticism that this view of probability is "subjective", objectivity is achieved when two people having the same information assign the same probability to an event. Objectivity demands that we consider all the pertinent information available, not an arbitrarily chosen subset of it. Any such choice would warrant the accusation of us being subjective, since we would be either ignoring available information or presuming information we do not possess, both of which would entail further scrutiny.

The scientific method is a way of reasoning from *premises* and available *data* to arrive at *inferences* about the problem in question. To draw inferences using probability theory, we formulate models of the scientific phenomenon and test their predictions against the data that have been obtained. Based on how much a theory predicts the observed data, our belief (or, more accurately, that of a rational agent) in the theory is modified accordingly. A practical point of concern is that, more often than not, the mathematical functions that describe the probability distributions corresponding to the model parameters can be quite complicated and difficult to solve using standard analytical methods. Hence we resort to methods such as *numerical sampling*, which involve repeated computations of simpler mathematical equations and provide approximate solutions to analytically intractable problems at a fraction of the time. The number of computations and hence the time required to obtain a numerical solution depend on how accurate we need the results to be. This necessitates the use of computer algorithms designed to perform computations iteratively on the data, to obtain solutions in a reasonable amount of time.

The Digital Revolution

In 1936, Alan Mathison Turing published "On Computable Numbers, with an Application to the *Entscheidungsproblem*"[*], culminating the works of David Hilbert and Kurt Gödel on the nature of logical and mathematical reasoning, and laying the theoretical foundations for the invention of the electronic computer. In this pioneering paper, he showed that any conceivable algorithm can be executed by a machine (now known as the "Turing machine"). Insofar as a process can be solved algorithmically using mathematical equations, a machine that carries out the process can be built. Insofar as rational thought is consonant with the rules of logic, a machine that *thinks* rationally can be built. Today he is rightfully considered the father of computer science and the founder of the field of artificial intelligence. John von Neumann and various others built on Turing's work, eventually creating machines that were the precursors of the modern digital electronic computer.

Computers today have come a long way from being the hulking behemoths of the mid-twentieth century. With the invention of semiconducting transistors, and their subsequent integration on silicon chips known as integrated circuits (IC) on a very large scale (ICs today can accommodate billions of transistors), computing power has increased enormously. Algorithms that were considered prohibitively costly in terms of execution time three decades ago, have become feasible today. In the past decade-and-a-half, the development of high performance Graphics Processing Units (GPUs), mainly

---

[*] German for "decision problem".

driven by the computer gaming industry, has made possible the numerical simulations necessary for tackling many scientific problems for which analytical solutions are intractable. It is in this technological backdrop that a full probability-theoretic approach to solving radio interferometry equations by exploiting the power of numerical samplers has become possible.

Where We Go From Here

The thesis is divided into six chapters. Chapter 1 provides an introduction to radio astronomy, with a specific focus on interferometers and the measurement equation that formalises the technique of interferometry. Chapter 2 serves as an introduction to probability theory, including a short discussion of numerical sampling algorithms used in the thesis.

Chapter 3 is adapted from my contribution to Lochner et al. (2015). It illustrates the application of probability theory to analyse radio interferometry data directly without computing images of the sky. The project described here investigates how this approach can be used to discriminate between source structures that lie beyond the diffraction limit of interferometers in the presence of random noise, using simulations of Westerbork Synthesis Radio Telescope observations. We accomplish this by observing three different simulated skies and performing model selection between them and presenting the evidences for each model tested, along with the parameter estimates for the correct models in each case.

Chapter 4 is adapted mostly from Natarajan et al. (2017). We apply the probabilistic method for achieving resolution beyond the diffraction limit discussed in Chapter 3 to Very Long Baseline Interferometry (VLBI) observations of the blazar CGRaBS J0809+5341 using the European VLBI Network. We also take a brief detour to discuss the development of Montblanc, a GPU-accelerated implementation of the radio interferometry measurement equation that facilitates fast model computation for the probabilistic analysis, published as Perkins et al. (2015). Integrating Montblanc with our software, we perform model selection between a point source and an extended Gaussian source model of the blazar, incorporating the individual station gains and baseline-dependent noise in the models. We then proceed to derive a probabilistic criterion for the resolution limit of interferometers.

In Chapter 5, we extend our probabilistic approach to include the vital process of fringe-fitting performed at the beginning of VLBI data analysis. Fringe-fitting is the process of minimising the net loss of amplitude that plagues VLBI observations due to the phase fluctuations that remain after the correlator model has been applied to the data. We do this by modelling the phase offsets and the phase fluctuations with respect to frequency, known as *delays*. We then proceed to incorporate time-variable delays in our models, while simultaneously performing model selection between point and extended source structures.

Chapter 6 summarises the conclusions of the thesis and explores some possible directions in which the work done so far can be extended in the future.

எப்பொருள் யார்யார்வாய்க் கேட்பினும் அப்பொருள்
மெய்ப்பொருள் காண்ப தறிவு.

திருக்குறள் (423)

Though things diverse from divers sages' lips we learn,
'Tis wisdom's part in each the true thing to discern.

Thirukkuṛaḷ (423)

1

Introduction to Radio Astronomy

In December 1932, Karl G. Jansky, a radio engineer commissioned by Bell Telephone Laboratories to study the origin of thunderstorm static that interfered with their radio transmissions, reported the detection of a steady *hiss type static* of unknown origin at a frequency of 20.5 MHz (Jansky, 1932). The following year, he confirmed that this radiation originated from the centre of the Milky Way galaxy (Jansky, 1933), making this the first ever observation of radio waves of extra-terrestrial origin. However, his proposal to investigate this radiation further was rejected by Bell Laboratories since they deemed it too weak to affect their communication service. Radio astronomy lay dormant until Grote Reber, another radio engineer and an admirer of Jansky's work, succeeded in mapping the Milky Way at 160 MHz (Reber, 1940) using the parabolic radio antenna he had built in his back yard. Reber continued to refine and publish his observations during the difficult years of World War II. Although the war hampered the progress of radio astronomy until 1945, the advancements made in developing sensitive radio equipment meant that scientists could now employ them to observe celestial radio sources. They promptly did, throwing open a low-frequency window onto the universe.

The Earth's atmosphere reflects a significant portion of the incoming electromagnetic radiation back into space, and, apart from visible light, is transparent only to radio waves of frequency roughly between 15 MHz and 1.5 THz. The ionosphere either absorbs or fully reflects the radio waves below 15 MHz, and the resonant vibration and rotation modes of the O_2 and H_2O molecules in the troposphere contribute to the cut-off above 1.5 THz (Wilson et al., 2009). Diverse astronomical sources such as supernova remnants, pulsars, black hole accretion disks, active galaxies, clusters of galaxies, and the cosmic microwave background (CMB) radiate in this large frequency range, with different astrophysical phenomena resulting in different emission mechanisms that are responsible for their radiation.

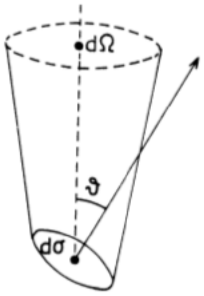


Figure 1.1: The relation between $d\Omega$, $d\sigma$ and the angle between the normal to $d\sigma$ and the direction to $d\Omega$ in the definition of brightness (equation 1.1) (Wilson et al., 2009).

1.1 Basic Relations

Consider the scenario in which an observer intercepts the radiation from an astronomical source. When the scale of the system under study is much larger than the wavelength of the radiation emitted, which is the case with astronomical sources, we may consider the waves to travel in straight lines called *rays*. The corresponding branch of study is called *geometrical optics* (Born & Wolf, 1999). Under this assumption, the infinitesimal power dP from a source that subtends an infinitesimal solid angle $d\Omega$ on the sky, intercepted by an infinitesimal surface $d\sigma$ (Figure 1.1) in an infinitesimal bandwidth $d\nu$ is given by (e.g. Kraus, 1986)

$$dP = I_\nu \cos\vartheta \, d\sigma \, d\Omega \, d\nu. \tag{1.1}$$

This equation defines a fundamental quantity of measurement in radio astronomy called *brightness* or *specific intensity*[*], denoted by I_ν. It is a measure of the power received per unit area per unit solid angle per unit bandwidth and is measured in units of $\mathrm{W\,m^{-2}\,Hz^{-1}\,sr^{-1}}$. A related quantity of measurement called the *flux-density*, is used to describe relatively compact sources, i.e. sources that appear as points according to the *Rayleigh criterion* (Rayleigh, 1879). The flux-density is obtained by integrating the brightness over the total solid angle Ω_s subtended by the source at the location of the observer,

$$S_\nu = \int_{\Omega_s} I_\nu \cos\vartheta \, d\Omega. \tag{1.2}$$

S_ν is measured in units of $\mathrm{W\,m^{-2}\,Hz^{-1}}$. Since the flux-density of radio sources is usually very small, the more convenient unit of *Jansky* (Jy) has been introduced:

$$1\,\mathrm{Jy} := 10^{-26}\,\mathrm{W\,m^{-2}\,Hz^{-1}}. \tag{1.3}$$

Brightness is an intrinsic property of a source, and, in the absence of extinction or emission along the line-of-sight, is independent of the distance between the source and the observer. In contrast,

---

[*] Sometimes also referred to as *intensity* in radio astronomy.

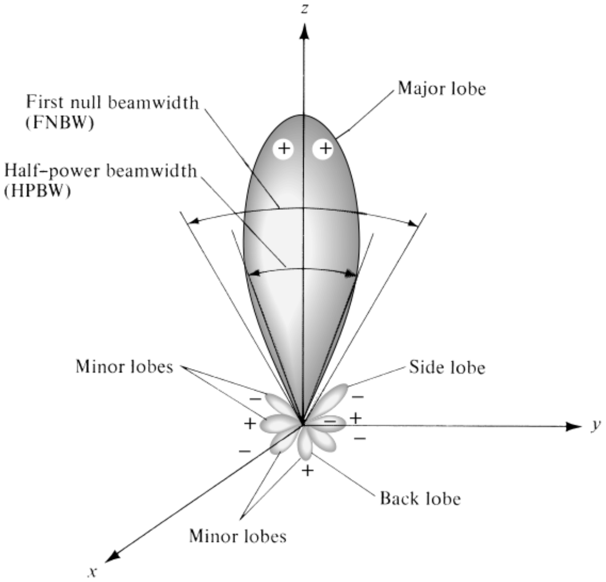


Figure 1.2: The *major*, *minor*, *side*, and *back lobes* of a radiation pattern, with the antenna pointing along the z-axis (Balanis, 2005).

the flux-density falls with the distance, since the solid angle subtended by the source depends on the observer's distance from it.

1.2 FUNDAMENTAL PARAMETERS OF ANTENNAS

An antenna is the transitional structure between free-space and a transmission line (Balanis, 2005). The mathematical or graphical representation of the response of an antenna to electromagnetic waves as a function of direction is called the *radiation pattern* of the antenna. Typically, the *field pattern* represents the amplitude of the electric or the magnetic field in two or three dimensional space and the *power pattern* represents the square of the field pattern. For a hypothetical isotropic antenna, the radiation pattern is the same in all directions; in practice, any real antenna radiates (or receives) power more effectively in some directions than in others. Figure 1.2, for instance, shows a schematic representation of the different components of the radiation pattern of an antenna with a circular aperture.

An antenna can function as both a receiver and a transmitter. For an antenna made of linear components[*], the *reciprocity theorem* (e.g. Wilson et al., 2009) states that the radiation pattern is the same in both the transmit and receive modes as long as they are polarisation-matched. This enables us to discuss the properties of antennas in transmission mode with the understanding that the same arguments apply for reception.

An important parameter associated with an antenna is its *beamwidth* , defined as the angular separation between two identical points on either side of the maximum of the radiation pattern. The

---

[*] Any electrical element in which the relationship between the current and the voltage is a linear function. Resistors, capacitors, and inductors are some common examples of linear elements.

most common criterion used to describe the resolving power of an antenna is the *full width at half maximum* (FWHM) or the *half-power beam width* (HPBW), given by the angle between the points at which the radiation intensity falls to half of its maximum value[*] (Figure 1.2). The *Rayleigh criterion*, which can be derived from diffraction theory (e.g. Feynman et al., 1977), states that if λ is the wavelength of the radiation received, the angular resolution ϑ of a telescope of diameter D is given by (Rayleigh, 1879)

$$\vartheta \propto \frac{\lambda}{D}. \tag{1.4}$$

The power pattern may be normalised to unity:

$$P_n = \frac{P(\vartheta, \varphi)}{P_{max}}. \tag{1.5}$$

Integrating this normalised power pattern over the full sphere 4π, we obtain the *beam solid angle,* Ω_A:

$$\Omega_A = \iint_{4\pi} P_n(\vartheta, \varphi)\, d\Omega. \tag{1.6}$$

This is the angle through which all the power transmitted by an antenna would stream if the power per unit solid angle were constant and maximum over this angle (Kraus, 1986). If this integral range is restricted to the main lobe MB, we obtain the *main-beam solid angle,* Ω_M:

$$\Omega_M = \iint_{MB} P_n(\vartheta, \varphi)\, d\Omega. \tag{1.7}$$

This quantity is useful in estimating the ratio $\frac{\Omega_M}{\Omega_A}$, known as the *beam efficiency*.

Another useful performance parameter is the *effective aperture* A_e of the antenna (Wilson et al., 2009). This is defined as the ratio between the amount of power extracted by an antenna P_e and the power per unit area (*power density*) $\langle S \rangle$ of a plane wave incident on it. A_e has dimensions of area:

$$A_e \equiv \frac{P_e}{\langle S \rangle}. \tag{1.8}$$

If the physical or geometrical area of the antenna is denoted by A_g, we define an *aperture efficiency* parameter η_A by

$$\eta_A \equiv \frac{A_e}{A_g}. \tag{1.9}$$

The effective area is always less than or equal to the physical area and hence η_A is equal to 1 at the most. The relation between aperture efficiency and the beam solid angle is a fundamental relationship in

---

[*] Or where the amplitude of the field falls to $\frac{1}{\sqrt{2}}$ of the maximum.

antenna theory; if λ is the wavelength corresponding to the frequency of radiation, then

$$A_e \Omega_A = \lambda^2. \tag{1.10}$$

A measure of the directionality of the radiation pattern of an antenna is given by its *directivity* D. This is the ratio between the power radiated by the antenna per unit solid angle in the direction of its strongest emission (also known as the *radiation intensity,* U) and the *average* radiation intensity ($P_{tot}/4\pi$) which is equal to the radiation intensity of an isotropic radiator (Balanis, 2005).

$$D = \frac{U(\vartheta, \varphi)}{P_{tot}/4\pi}. \tag{1.11}$$

For any directional antenna, the directivity is always greater than unity. If the beam solid angle is known, then the directivity may be obtained by calculating the ratio between the solid angle of a sphere and the beam solid angle, or, from equation (1.10), using the aperture efficiency (e.g. Condon & Ransom, 2016):

$$D = \frac{4\pi}{\Omega_A} = \frac{4\pi}{\lambda^2} A_e. \tag{1.12}$$

Finally, we introduce the notion of the *antenna temperature,* T_A. The *Johnson-Nyquist theorem* states that, in thermal equilibrium, a resistor R connected to a matched load delivers a power per unit bandwidth that is proportional to its temperature (e.g. Wilson et al., 2009). If the resistor is matched to an antenna, the power per unit bandwidth delivered to the antenna is given by

$$\mathcal{P}_\nu = k_B T_A, \tag{1.13}$$

where the proportionality constant k_B is the Boltzmann constant. The antenna temperature is a convenient way to express the power delivered to the antenna. It is useful when one needs to compare the power delivered to quantities such as the receiver noise expressed in temperature units. For a compact source that subtends a solid angle Ω_s of the sky, the antenna temperature is

$$T_A \approx T_s \frac{\Omega_s}{\Omega_A}. \tag{1.14}$$

The quantity T_s is called the *brightness temperature* of the source which, in the Rayleigh-Jeans limit (e.g. Wilson et al., 2009), is related to the specific intensity as

$$T_s(\lambda) = \frac{I_\nu \lambda^2}{2k_B}. \tag{1.15}$$

For thermal emission, this is the same as the thermodynamic temperature of the source. For non-thermal emission, it is the equivalent temperature of a blackbody that radiates the same amount of

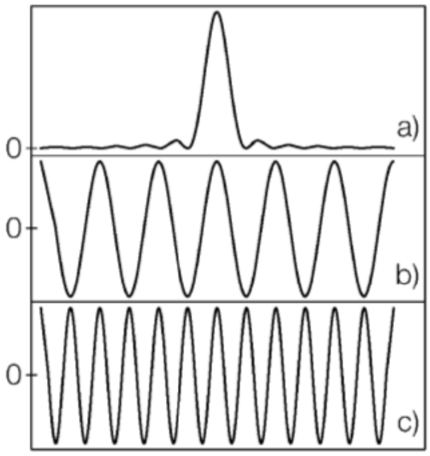


Figure 1.3: Antenna power patterns, with the horizontal axis representing the angle which increases from 0° away from the centre. (a) uniformly illuminated aperture with a diameter D, with a FWHM of about $\frac{\lambda}{D}$; (b) two-element multiplying interferometer with two antennas of diameter d separated by a distance D, with $d \ll D$; (c) the interferometer system described in (b) but spaced by $2D$ (Wilson et al., 2009).

power at the same wavelength λ. Hence, it is also known as the *equivalent blackbody temperature* (see e.g. Kraus, 1986).

1.3 RADIO INTERFEROMETRY

Equation (1.4) states that the angular resolution of a telescope depends on the wavelength of reception and the diameter of the aperture. Radio telescopes cannot achieve the angular resolutions of optical telescopes with a single-dish antenna owing to the long wavelengths involved. We may increase the resolution of a radio telescope by coherently combining the outputs of two or more antennas of diameter d each, separated by a maximum distance of D, with $d \ll D$ (Wilson et al., 2009). Such an array of telescopes is called an *interferometer*. An interferometer combines the outputs from two separate telescope elements (antennas or *stations*[*]) by applying the principle of interference of waves. Where the waves are *in phase*, there is constructive interference and the power is maximum and where they are *out of phase*, the interference is destructive and the power falls to zero (Feynman et al., 1977).

Figure 1.3 compares the power patterns of a single-dish antenna and a two-element interferometer. Panel (a) shows the power pattern of a single-dish antenna, with a well-defined main lobe and other minor lobes. In panels (b) and (c), we see the patterns of two interfering waves, with the two antennas that make up the interferometer separated by distances of D and $2D$ respectively. The alternating maxima and minima with no appreciable distinction between the major and the minor lobes is called the *fringe pattern*. The sidelobes can be minimised by adding more antennas and correlating their outputs in such a way that the interference is constructive at the location of the central fringe and destructive at the locations of the adjacent fringes.

---

[*] Henceforth we use the terms *antenna* and *station* interchangeably.

1.3.1 APERTURE SYNTHESIS

Aperture synthesis is the process of synthesising a large aperture using two or more smaller antennas (Ryle, 1955). Imagine a single-dish antenna as a composition of N elementary segments. The output voltage V measured at the antenna terminals may then be represented as the sum of contributions ΔV_i from each individual segment i (e.g. Christiansen & Högbom, 1985):

$$V(t) = \sum_i \Delta V_i(t). \tag{1.16}$$

The power measured by the antenna receiver is proportional to the time-average (denoted by angle brackets) of the square of the output voltage:

$$\langle P \rangle \propto \langle (\sum_i \Delta V_i(t))^2 \rangle = \sum_i \sum_j \langle (\Delta V_i . \Delta V_j) \rangle \,. \tag{1.17}$$

We may consider each elementary segment to represent an individual antenna of an interferometer. Since the interferometer consists of a finite number of discrete antennas distributed sparsely[*], only a handful of these elements have measurements associated with them at any given time. As long as the brightness distribution of the sky does not vary much, which is true of many radio sources[†], measurements for the missing orientations and spacings can be obtained over the course of minutes or hours. This can be achieved either by changing the antenna configuration or, more commonly, by using the rotation of the Earth to collect measurements at different relative antenna orientations. Figure 1.4 illustrates how this is accomplished for an interferometer with a baseline running East-West. The plane in which the ellipse is traced out is called the *uv-plane*, with the *u*-axis along the East-West direction and the *v*-axis running perpendicular to it. Measurements for various antenna orientations and spacings fill out a three-dimensional projection plane called the *uvw-plane*. The beam 'synthesised' by this method is the *point spread function* (PSF) of the interferometer and the individual antenna beams are called the *primary beams*. The PSF determines the resolution of the interferometer and the primary beam determines its field-of-view (FoV).

1.3.2 THE RADIO INTERFEROMETER MEASUREMENT EQUATION (RIME)

Originally developed for radio polarimetry by Hamaker, Bregman, and Sault (Hamaker et al., 1996), the Radio Interferometer Measurement Equation (henceforth referred to as the RIME) provides a solid theoretical foundation for radio interferometry. Smirnov (2011a) extended the formalism to the full-sky case and incorporated direction-dependent observation effects. We follow this derivation in our description of the RIME. A generalised tensor formalism is presented in the final paper of the

---

[*] Also known as an *unfilled aperture* antenna.

[†] Notable exceptions are *transient* radio sources which are visible only for short (of the order of seconds or minutes) bursts of time.

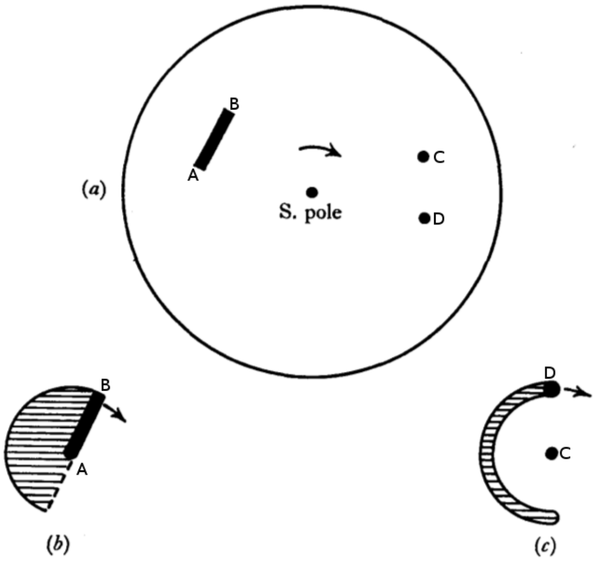


Figure 1.4: (a) View of an East-West line antenna A-B and a two-station interferometer C-D from a point above the South Pole. (b) If the source is observed for 12 hours, exactly half of a filled-circle (or, more generally, filled-ellipse) is traced out by the line A-B. (c) For the same 12 hours, the range of apparent spacings and directions produced by the interferometer C-D takes the form of a half-circle (or a half-an-ellipse). After each complete rotation of the Earth, the spacing between C-D can be changed to fill out the ellipse as shown in (b). The plane in which (b) and (c) are drawn is called the *uv*-plane. The measurements corresponding to the remaining half provide no new information and may be deduced from the first half. For an interferometer with N stations, $\frac{N(N-1)}{2}$ combinations are enough to provide all possible measurements. Figure adapted from Christiansen & Högbom (1985).

RIME series (Smirnov, 2011d).

1.3.2.1 A Single Point Source

Consider quasi-monochromatic[*] radiation received from a single 'point' source in the sky. Using an orthonormal co-ordinate system *xyz* in which the *z*-axis lies along the direction of propagation, this signal may be described by the complex vector $\vec{e}$:

$$\vec{e} := \begin{pmatrix} e_x \\ e_y \end{pmatrix}. \tag{1.18}$$

We make a fundamental assumption here that all transformations that the signal undergoes along its path are *linear* with respect to $\vec{e}$. This enables us to represent these transformations using a matrix multiplication (e.g. Lang, 1986):

$$\vec{e'} = J\vec{e}. \tag{1.19}$$

---

[*] An electromagnetic wave whose frequency is between ν and $\delta\nu$, with $\delta\nu \to 0$.

Here, J is a 2×2 complex matrix known as the *Jones* matrix (Jones, 1941). Multiple effects along the signal path can be written out separately in a *Jones chain*[*] as

$$\vec{e'} = J_n J_{n-1} \dots J_1 \vec{e}. \tag{1.20}$$

A complex voltage pattern is generated when the signal hits the antenna. If the two orthogonal polarisations received by the antenna feeds are denoted by a and b, the voltage V generated can be written out as a 2-vector analogous to $\vec{e}$:

$$\vec{v} \equiv \begin{pmatrix} v_a \\ v_b \end{pmatrix} = J\vec{e}, \tag{1.21}$$

where J represents the cumulative Jones matrix of all propagation effects along the signal path. This equation establishes a basic linear relationship between the voltage vector $\vec{v}$ and the signal vector $\vec{e}$ representing the incoming electromagnetic wave.

In an interferometer, the voltage outputs from two spatially separated antennas p and q are correlated to obtain 4 pairwise correlations (known as *visibilities*) that can be arranged into a matrix called the *visibility matrix*[†]:

$$\mathrm{V}_{pq} = 2\begin{pmatrix} \langle v_{pa} v_{qa}^* \rangle & \langle v_{pa} v_{qb}^* \rangle \\ \langle v_{pb} v_{qa}^* \rangle & \langle v_{pb} v_{qb}^* \rangle \end{pmatrix}, \tag{1.22}$$

where the angle brackets denote averaging over time and frequency[‡] and $*$ denotes the complex conjugate. This 2×2 matrix can be understood as the matrix product of v_p and v_q^H, where H denotes the *Hermitian* or conjugate transpose.

$$\mathrm{V}_{pq} = 2\left\langle \begin{pmatrix} v_{pa} \\ v_{pb} \end{pmatrix} \begin{pmatrix} v_{qa}^* & v_{qb}^* \end{pmatrix} \right\rangle = 2\left\langle \vec{v}_p \vec{v}_q^H \right\rangle. \tag{1.23}$$

Combining equations (1.21) and (1.23), we obtain

$$\mathrm{V}_{pq} = 2\left\langle J_p \vec{e} (J_q \vec{e})^H \right\rangle = 2\left\langle J_p (\vec{e}\,\vec{e}^H) J_q^H \right\rangle. \tag{1.24}$$

Now if we assume[§] that the Jones matrices are constant over the time- and frequency- averaging in-

---

[*] Because matrix multiplication is not always commutative, the order in which the effects occur must, in general, be preserved while writing out the corresponding Jones matrices.

[†] The factor of 2 is introduced in this equation to ensure that the brightness matrix (introduced shortly) becomes unity for a 1 Jy unpolarised source (Smirnov, 2011a).

[‡] The sizes of the averaging bins depend on considerations of *smearing* and *decoherence* (e.g. Thompson et al., 2017; Smirnov, 2011a). See also section 5.1.

[§] A detailed account of the resulting effects for Very Long Baseline Interferometry (VLBI, introduced in Chapter 4) observations is given in section 5.1.

tervals (Smirnov, 2011a), then we can move them outside the averaging operator:

$$\mathrm{V}_{pq} = 2 J_p \begin{pmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{pmatrix} J_q^H. \tag{1.25}$$

There exists a fundamental relationship between the 2 × 2 matrix shown above and the four Stokes parameters (I, Q, U, and V) used to describe the state of polarisation of electromagnetic radiation (Hamaker & Bregman, 1996):

$$2 \begin{pmatrix} \langle e_x e_x^* \rangle & \langle e_x e_y^* \rangle \\ \langle e_y e_x^* \rangle & \langle e_y e_y^* \rangle \end{pmatrix} = \begin{pmatrix} I + Q & U + iV \\ U - iV & I - Q \end{pmatrix} \equiv \mathrm{B}. \tag{1.26}$$

The matrix B is called the *brightness matrix*. Equation (1.26) relates the brightness, which is an intrinsic property of the source, to the output visibilities V_{pq} (equation 1.25) obtained using an interferometer:

$$\mathrm{V}_{pq} = J_p \mathrm{B} J_q^H. \tag{1.27}$$

The various propagation effects associated with each signal path can be written out explicitly, resulting in the *onion* form of the RIME:

$$\mathrm{V}_{pq} = J_{pn} \left(\dots \left(J_{p2} (J_{p1} \mathrm{B} J_{q1}^H) J_{q2}^H \right) \dots \right) J_{qm}^H. \tag{1.28}$$

Different propagation effects are described by Jones matrices of different mathematical properties. A propagation effect that affects both the components of $\vec{e}$ equally can be represented by a *scalar* matrix. Scalar matrices are immune to coordinate transformations and commute with any matrix[*]. Some effects can be described using diagonal matrices, which commute among themselves. Any rotation of the electric field vector can be represented by a rotation matrix, which, for the 2 × 2 case, commutes with other rotation matrices. The actual values of the diagonal and the rotation matrices vary based on the coordinate system chosen.

In the ideal case in which no corrupting effects act on the signal, there still remains the phase delay associated with the difference between the geometric path lengths from the source to the antennas p and q. The correlator introduces additional delay terms to compensate for this difference in the chosen direction called the *phase centre*. Adopting the conventional coordinate system (Thompson et al., 2017), with the z-axis pointing at the phase centre and the antenna located at $\vec{u}_p = (u_p, v_p, w_p)$, the phase difference κ_p at $\vec{u}_p$ relative to $\vec{u} = \vec{0}$ for a signal arriving from direction $\vec{\sigma}$ is given by

$$\kappa_p = 2\pi\lambda^{-1}\left(u_p l + v_p m + w_p (n - 1) \right), \tag{1.29}$$

where l, m, and $n \equiv \sqrt{1 - l^2 - m^2}$ are the direction cosines of $\vec{\sigma}$ and λ is the wavelength of the signal.

---

[*] For matrix properties, refer to any introductory textbook on linear algebra (e.g. Lang, 1986).

If we define $\vec{u}$ in units of wavelength, then we may drop the λ^{-1} term and introduce a scalar *K-Jones* matrix defined as

$$K_p = \mathrm{e}^{-i\kappa_p} = \mathrm{e}^{-2\pi i(\,u_p l + v_p m + w_p (n-1))}. \tag{1.30}$$

The RIME for a single uncorrupted point source then becomes

$$\mathrm{X}_{pq} = K_p \mathrm{B} K_q^H. \tag{1.31}$$

This is the visibility that would be measured by an interferometer if there were no corrupting effects along the signal path. We will call this the *source coherency* matrix, X_{pq}. Equation (1.31) may also be represented in terms of the baseline coordinates $\vec{u}_{pq}$ by substituting equation (1.30) into it and commuting the scalar matrices K_p and K_q:

$$\mathrm{X}_{pq} = \mathrm{V}_{pq} = \mathrm{B}\mathrm{e}^{-2\pi i(\,u_{pq} l + v_{pq} m + w_{pq} (n-1))}. \tag{1.32}$$

Any real-world observation will have some effects acting on the signal along the propagation path. Since the phase delay matrix K_p is scalar, we can commute it towards the brightness matrix B and express the RIME for a corrupted point source in terms of the source coherency, X_{pq}:

$$\mathrm{V}_{pq} = G_p\, K_p \mathrm{B} K_q^H G_q^H = G_p\, \mathrm{X}_{pq} G_q^H. \tag{1.33}$$

where G_p represents all other corrupting effects in the propagation path such as instrumental gains or polarisation effects.

1.3.2.2 Multiple Discrete Sources

The extension of this formulation to multiple point sources is straightforward. Consider N point sources in the sky contributing to the measured visibilities. The contribution from each source adds up linearly and each path from the source s to the antenna p can be described by its own Jones matrix J_{sp}:

$$\mathrm{V}_{pq} = \sum_s J_{sp}\, \mathrm{B}_s\, J_{sq}^{\;H}. \tag{1.34}$$

Now, if we denote the *direction-independent* effects (DIEs), or the *uv-Jones* terms, by G_p and the *direction-dependent* effects (DDEs), or the *sky-Jones* terms, by E_{sp}, then we can commute the K-Jones matrices as we did earlier and arrive at the RIME formalism for the case where the sky is composed of N discrete sources:

$$\mathrm{V}_{pq} = G_p \left(\sum_s E_{sp} K_{sp}\, \mathrm{B}_s\, K_{sq}^H E_{sq}^H \right) G_q^H = G_p \left(\sum_s E_{sp}\, \mathrm{X}_{spq}\, E_{sq}^H \right) G_q^H. \tag{1.35}$$

1.3.2.3 THE FULL-SKY RIME

In general, the sky is not a collection of discrete sources, but consists of a continuous brightness distribution $\mathrm{B}(\vec{\sigma})$, where $\vec{\sigma}$ is the unit direction vector. The individual Jones terms can be expressed as functions of $\vec{\sigma}$ and the total visibility measured by the interferometer is obtained by integrating over all possible directions:

$$\mathrm{V}_{pq} = \int_{4\pi} J_p(\vec{\sigma})\,\mathrm{B}(\vec{\sigma})J_q^H(\vec{\sigma}) \quad d\Omega. \tag{1.36}$$

To make this spherical integral more tractable, we project the spherical surface onto a plane that is tangential to the surface at the phase centre*. The equation then reduces to the two-dimensional integral

$$\mathrm{V}_{pq} = \iint_{lm} J_p(l,m)\,\mathrm{B}(l,m)J_q^H(l,m) \quad \frac{dl\,dm}{n}. \tag{1.37}$$

Writing out the antenna-based DIEs and DDEs and the baseline-based K_{pq} terms explicitly, we get

$$\mathrm{V}_{pq} = G_p\left(\iint_{lm} \frac{1}{n}\,E_p\mathrm{B}E_q^H\mathrm{e}^{-2\pi i(u_{pq}l+v_{pq}m+w_{pq}(n-1))}\,dl\,dm\right)G_q^H. \tag{1.38}$$

Except in cases where all the baselines lie East-West or only *snapshot*† observations are made, the baselines do not all lie in the same plane. The term $w_{pq}(n-1)$ accounts for this *non-coplanarity* of the baselines. Since $w_{pq} \equiv w_p - w_q$, we can decompose this term into antenna-based terms,

$$W_p = \frac{1}{\sqrt{n}}\mathrm{e}^{-2\pi i w_p(n-1)}. \tag{1.39}$$

The W_p terms are DDEs in their own right and can be subsumed into the overall DDE terms E_p. With this change, equation (1.38) reduces to the standard 2D Fourier Transform‡ of the *apparent sky as seen by baseline pq*:

$$\begin{aligned}\mathrm{V}_{pq} &= G_p\left(\iint_{lm} \mathrm{B}_{pq}\mathrm{e}^{-2\pi i(u_{pq}l+v_{pq}m)}\,dl\,dm\right)G_q^H,\\ \mathrm{B}_{pq} &\equiv E_p\mathrm{B}E_q.\end{aligned} \tag{1.40}$$

This means that as long as there are DDEs, the sky seen by each baseline is different. Only if all DDEs are *identical* across all antennas, we may assume that all baselines see the same *apparent* sky i.e. $\mathrm{B}_{pq} \equiv \mathrm{B}_{app} = E\,\mathrm{B}\,E^H$ and rewrite the full-sky RIME as

$$\mathrm{V}_{pq} = G_p\,\mathrm{X}_{pq}\,G_q^H, \tag{1.41}$$

---

[*] This is analogous to the integral carried out in Thompson et al. (2017, section 3.1).

[†] Snapshots are interferometric observations carried out for a duration of a few minutes at the most.

[‡] An excellent introduction to Fourier Transforms can be found in Bracewell (2000) and in the course notes and online video lectures by Brad Osgood (Osgood, 2016).

where the source coherency $X_{pq} = X(u_{pq}, v_{pq})$ is the element-wise 2D Fourier transform of the brightness matrix B_{pq}. This is effectively the *van Cittert-Zernike theorem*, which is often treated as the point of departure by expositions on the theory of radio interferometry (e.g. Thompson et al., 2017; Condon & Ransom, 2016). By treating the phase delay matrix K as a Jones matrix (Noordam, 1996) we have arrived at the van Cittert-Zernike theorem as a consequence of the RIME.

1.3.3 CALIBRATION & IMAGING

Measuring visibilities using an interferometer is only the first step towards estimating the physical characteristics of the observed source. Errors in tracking, receiver malfunctions, interference from the local weather, and other terrestrial radio sources are but some of the factors that contribute to the signal corruption effects that affect the observed visibilities (Taylor et al., 1999). These discrepant and corrupted data are identified as best as possible and discarded by the process of *editing* or *flagging*. The remaining data are corrected by the process of *calibration*. Traditionally, this refers to the process of measuring the uncertainties in the data and estimating the true visibilities.

The first generation of calibration techniques involved (and still involve) observing a source of known flux-density and position and using these measurements to constrain the visibilities of the target source. The next generation of calibration methods, termed as *second generation calibration* (2GC), also called *self-calibration* (e.g. Readhead & Wilkinson, 1978), use some form of the RIME implicitly[*] (see also section 5.1.2.3). Typical 2GC software might implement this as (Smirnov, 2011b)

$$V_{pq} = G_p(M_{pq} * X_{pq})G_q^H, \tag{1.42}$$

where G_p denotes the antenna-based gain terms, M_{pq}, the multiplicative interferometer errors and X_{pq} denotes the *model visibilities*. The X_{pq} term includes the primary beam gain as a *trivial*[†] DDE multiplied by the source coherency for each source. In this approach the model visibilities are separated from antenna-based gains and the visibilities are corrected by applying the G_p term solutions. These concepts break down when DDEs are involved. *Explicit* RIMEs go a long way towards handling these limitations by implementing explicit Jones chains, with specific parametrisations for each Jones term. However, this approach is too inflexible to implement generic DDEs not anticipated at the time of software design (Smirnov, 2011b).

Calibration and imaging have always been treated as distinct problems in 2GC. In the RIME-based *third generation calibration* (3GC) methods, they are seen as two aspects of optimising the solution to the RIME (Rau et al., 2009; Smirnov, 2011b) and are therefore treated as parts of the same numerical optimisation problem. The MEQTREES[‡] software suite (Noordam & Smirnov, 2010) implements

---

[*] Some plausible assumptions about the source structure are made to correct the observed visibilities (Taylor et al., 1999, Chapter 10).

[†] A trivial DDE is one that is constant in time and identical across stations.

[‡] `http://ska-sa.github.io/meqtrees`

phenomenological RIMEs for calibration, wherein the RIME is constructed using a minimum number of solvable terms needed to represent all the effects in the Jones chain, with each term subsuming several physical effects. This approach also allows for a more explicit and generic description of polarisation effects.

1.3.3.1 CALIBRATION IN THE PRESENCE OF DIRECTION-DEPENDENT EFFECTS (DDES)

In section 1.3.2.3, we made the important assumption that the DDEs are trivial, so that we may consider the apparent sky B_{app} to be the same for all baselines and the sky coherency $\mathrm{X}(u_{pq}, v_{pq})$ to be the element-wise Fourier transform of B_{app}. In the presence of *non-trivial* DDEs, this simplifying assumption breaks down. To see how, let us take a look at the expression for B_{app} more closely.

$$\mathrm{B}_{pq}(\vec{l}) \equiv \mathrm{B}_{app}(\vec{l}) = E_p(\vec{l})\,\mathrm{B}(\vec{l})\,E_q^H(\vec{l}), \tag{1.43}$$

where $\vec{l} \equiv (l, m)$. This shows that the DDEs are multiplicative in the *lm* plane, which corresponds to a convolution in the *uv* plane. Assuming $G_p \equiv 1$ for the moment, and denoting convolution by the symbol '$\circ$', equation (1.41) may now be written as

$$\begin{aligned} \mathrm{V}_{pq} &= \mathrm{X}_{pq}(\vec{u}_{pq}), \\ \mathrm{X}_{pq} &= U_p \circ \mathrm{X} \circ U_q^H, \end{aligned} \tag{1.44}$$

where the convolution kernels U_p are the Fourier transforms of E_p. Remembering the time-dependence of non-trivial DDEs, we write

$$\begin{aligned} \mathrm{V}_{pq}(t) &= \mathrm{X}_{pq}[t](\vec{u}_{pq}(t)), \\ \mathrm{X}_{pq}[t] &= U_p[t] \circ \mathrm{X} \circ U_q^H[t], \\ \mathrm{X} &= \mathcal{F}\mathrm{B},\ U_p[t] = \mathcal{F}E_p[t]. \end{aligned} \tag{1.45}$$

We see that DDEs convolve the *ideal* visibilities with a different kernel per antenna and time sample and we end up sampling a different *uv*-plane per baseline *pq*, per time interval. This is why, in the presence of DDEs, recovering the true visibilities is an ill-posed problem.

CALIBRATING THE KNOWN DDES Several methods that blur the boundaries between calibration and imaging have been proposed for dealing with both the known and the unknown non-trivial DDEs. Some known DDEs (Smirnov, 2011b) are primary beam effects (E_p), ionospheric phase delay (Z_p, a scalar Jones term), Faraday rotation (F_p, a rotation matrix) and tropospheric phase delay T_p. The methods proposed to correct for known DDEs include facet imaging (Cornwell & Perley, 1992), which images many small *facets* with different phase centres and stitches the resulting images together; AW-projection (Bhatnagar et al., 2008), an extension of W-projection (Bhatnagar et al., 2008), which uses an FFT-based convolution function approach to generalise the W_p-term correction to arbitrary DDEs; and subtraction in the *uv*-plane, which uses a Discrete Fourier Transform (DFT) to model and subtract discrete bright sources in the *uv*-domain.

CALIBRATING THE UNKNOWN DDES As a simple case of this, let us consider a field with two discrete point sources. In the presence of DDEs, traditional self-calibration will tend to subsume those DDEs along the direction of the strongest source into its gain solutions; but unmodelled DDEs in the direction of the next brightest source will corrupt the fitted visibilities. Any deficiencies in the sky model, such as unmodelled sources, will give rise to similar contaminations which will show up as *image artefacts* (Smirnov, 2011b,c). *Peeling* (Noordam, 2004), a method which is computationally expensive but useful when there are a few bright sources with contaminating DDEs, calibrates and removes the DDEs from each source in the order of decreasing source strength. The *differential gains* approach (Smirnov, 2011d) is a generalised form of peeling that uses a phenomenological RIME of the form

$$\mathrm{V}_{pq} = G_p \left(\sum_s \Delta E_{sp}\, \mathrm{X}_{spq}\, \Delta E_{sq}^H \right) G_q^H . \tag{1.46}$$

Here the G_p and the ΔE_{ps} solutions are estimated simultaneously for appropriate time/frequency intervals (small for G_p and larger for ΔE_p) so that the G_p terms subsume most DDEs in the direction of the dominant source, while the ΔE_{ps} terms account for the *difference* towards the fainter sources. Any DDE that is unknown or unmodelled can be accounted for using this method. Both peeling and differential gains have the same weakness i.e. of errors from the current step being frozen in and proliferating down the line. These errors can be partially mitigated by using larger *solution intervals* i.e. intervals in time and frequency for which an independent solution is sought for a solvable term (Smirnov, 2011d).

1.3.3.2 IMAGING INTERFEROMETRIC DATA

The Fourier transform relationship between the observed visibilities and the sky brightness (equation 1.40) can be exploited to produce a crude 'image' of the sky by linear Fourier inversion, often referred to as the 'dirty' image (Taylor et al., 1999, Chapter 7). At points in the *uv*-domain where no measurements were made, the visibilities (i.e. the Fourier transform of the sky brightness model) are free to take on any value (Taylor et al., 1999, Chapter 8). Introducing a generalised *weighted sampling function*, $W(u, v)$, to describe the sampling process, the observed visibilities V' may be written as

$$\mathrm{V}'(u, v) = W(u, v)\mathrm{V}(u, v). \tag{1.47}$$

This multiplication becomes a convolution when Fourier-inverted and the resulting Fourier transform of the sampling function is known as the *dirty beam*.

Different visibility samples can be weighted according to different weighting schemes (Briggs, 1995).

For instance, the weighting scheme most suited for high resolution tends to also lower the *sensitivity* (section 1.4) of the final image. The most commonly used weighting schemes are *natural* and *uniform* weighting. Natural weighting weights all points in the *uv*-domain equally (the weights are inversely proportional to the variance of the noise on the given visibility) and hence provides the best sensitivity for detecting weak sources. Since there are more samples closer to the origin of the *uv*-domain, natural weighting emphasises data from short spacings, thereby picking out large-scale structure in the sky. On the contrary, uniform weights are inversely proportional to the sampling density function. This minimises the sidelobe levels but worsens the sensitivity since the data are degraded. *Robust* weighting provides a balance between the two, with the mechanism to choose between natural and uniform weighting made available through an adjustable parameter (Briggs, 1995) (Figure 1.5).

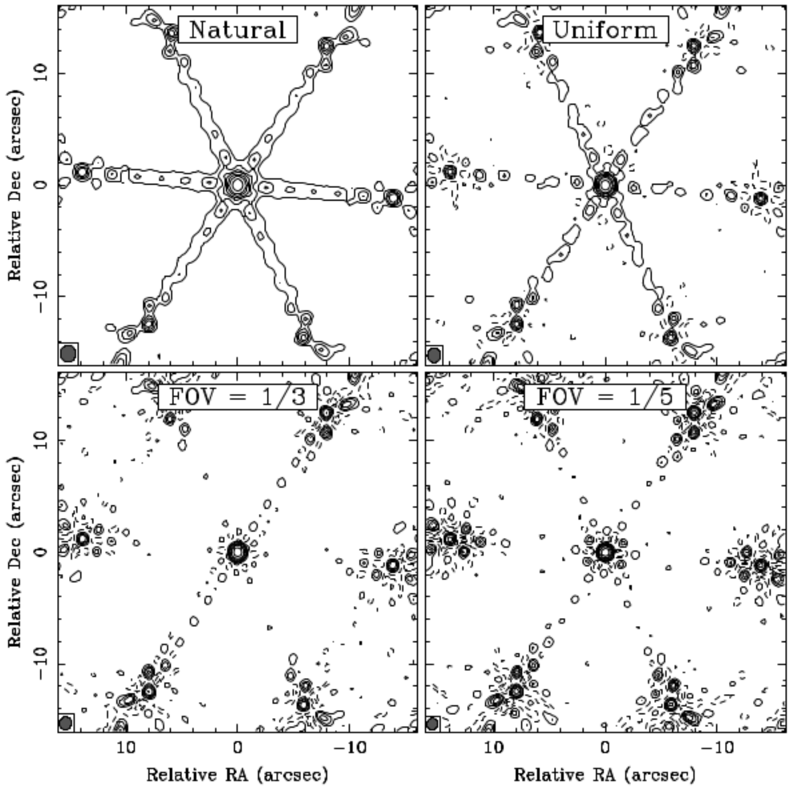


Figure 1.5: The effect of different weighting schemes on a VLA *snapshot* image of a point source (Taylor et al., 1999, Chapter 7).

To obtain a high-quality image of the sky, the effects of the dirty beam must be *deconvolved* using non-linear deconvolution techniques. The sidelobes of the dirty beam introduce artefacts in the dirty image that can easily be confused with true emission from the sky. Barring further observations, introducing *a priori* information about the true sky brightness distribution is the only way of decoupling the artefacts from real emission. The two most commonly used deconvolution techniques are CLEAN (Högbom, 1974) and the Maximum Entropy Method (MEM, Ables, 1974), which use different constraints to arrive at reasonable solutions.

CLEAN operates on the assumption that the emission from the radio source can be represented as a collection of point sources. The most basic version of the algorithm iteratively searches for the brightest point sources and records their positions to build up the *clean component* model. Once the brightest point falls below a user-specified level, the CLEAN component model is convolved with an ideal restoring beam, an elliptical Gaussian, and added back to the *residuals* from the dirty image (Högbom, 1974). Thus, CLEAN chooses a *plausible* image from a set of feasible ones (e.g. see Figure 4.3). In contrast, MEM selects the image that fits the data best within the noise level, and also has the maximum entropy (Taylor et al., 1999, Chapter 8). It must be noted that the image entropy is different from the concept of physical entropy (e.g. Jaynes, 1982; Nityananda & Narayan, 1982). CLEAN works best when the sources in the sky are compact while MEM performs better in the presence of diffuse structure. New variants and hybrid algorithms aim to combine the advantages of both techniques.

1.4 SENSITIVITY AND NOISE

The *sensitivity* of a radio telescope is a measure of the weakest signal that it can detect. Signals generated by the receiver electronics or the radiation reflected from the ground (*spill-over* radiation) are examples of unwanted *noise* that affect the system sensitivity adversely. To improve the sensitivity, we must ensure that this noise is minimised. From equation (1.13), we may define the *system temperature* T_{sys}, equivalent to the power per unit frequency due to the noise, P_N:

$$T_{sys} = \frac{P_N}{k_B}. \tag{1.48}$$

In the same vein, we may define the source temperature T_{src}, equivalent to the power per unit frequency P_ν due to the source:

$$T_{src} = \frac{P_\nu}{k_B}. \tag{1.49}$$

T_{sys} includes the noise due to the input T_A, and the receiver T_{Rx} (Wilson et al., 2009). The T_A term incorporates contributions from the sources in the sky (including the sources of interest), the Earth's atmosphere, and reflected noise from the ground, to mention a few. The term T_{Rx} comes from the thermal noise of the receiver electronics.

T_{sys} and T_{src} are statistically similar (Taylor et al., 1999) – they are broadband and mostly *stationary*[*]. According to the *central limit theorem* (e.g. Jaynes, 2003), the amplitude of such signals can be described using a Gaussian distribution and the uncertainty in their measurements can be reduced by increasing the number of independent measurements, N. For a radio telescope, this number is

---

[*] The average power of a stationary signal does not change over time though the instantaneous power produced remains unpredictable.

determined by the bandwidth $\Delta\nu$ and the integration time τ:

$$N = 2\,\Delta\nu\,\tau, \tag{1.50}$$

where the factor of 2 is introduced to meet the Nyquist sampling rate (e.g. Wilson et al., 2009). Since T_{sys} characterises the variance of the random noise, the uncertainty that is minimised is given by $\sqrt{2}\,T_{sys}$. Hence, the *root mean square (rms)* uncertainty in noise temperature is given by

$$\sigma_T = \frac{\sqrt{2}\,T_{sys}}{\sqrt{N}}. \tag{1.51}$$

This equation is called the *ideal radiometer equation*. T_{src} only needs to be a few times the rms noise given by this equation. This is why, even sources for which $T_{src} \ll T_{sys}$ can often be easily detected.

In an interferometer, this noise manifests itself as the uncertainty in the complex visibility measurements. In units of flux-density, the rms noise per visibility measurement is given by (Thompson et al., 2017)

$$\sigma_{rms} = \frac{\sqrt{2}\,k_B\,T_{sys}}{A_e\,\eta\,\sqrt{\Delta\nu\,\tau}}, \tag{1.52}$$

where A_e denotes the effective area of the telescope (equation 1.8) and η comprises any relevant efficiency terms, such as the antenna aperture efficiency (equation 1.9) and the correlator efficiency. When dealing with interferometers composed of stations of different system temperatures, the rms noise on baseline pq can be obtained by computing the geometric mean of the individual rms noise values:

$$\sigma_{pq} = \sqrt{\sigma_p\,\sigma_q}. \tag{1.53}$$

Finally, an indicator of the combined sensitivity of both the antenna and the receiver, called the *System Equivalent Flux Density* (SEFD), is defined as the equivalent flux-density of a source that would deliver twice the amount of power P_N corresponding to that given by equation (1.48) (Taylor et al., 1999):

$$\text{SEFD} \equiv \frac{2k_B T_{sys}}{\eta\,A_e}. \tag{1.54}$$

Expressed in terms of the SEFD, equation (1.52) becomes

$$\sigma_{rms} = \frac{\text{SEFD}}{\sqrt{2\,\Delta\nu\,\tau}}. \tag{1.55}$$

Probability theory is nothing but common sense reduced to calculation.

Pierre-Simon Laplace

2

The Logic of Science: Probability and Inference

The title of this chapter borrows from the comprehensive work on probability theory as an extension of logic by E. T. Jaynes (Jaynes, 2003). Unifying the works of Cox (1946), Shannon (1948), Pólya (1954), and Jeffreys (1961), Jaynes provides a quantitative description of the rules of probability as the uniquely consistent set of rules for *inference* or *plausible reasoning* of any kind. *Deductive* reasoning concerns itself with certainties and does not need the application of probability theory; the conclusions *necessarily* follow from the premises, as long as we adhere to the rules of deduction. But in most situations, we are required to reason in the face of incomplete information. For example, no amount of deduction can guarantee a *certain* answer to a proposition such as, "Will it rain tomorrow?". Viewed as extended logic, probability theory can answer such questions in terms of the *plausibility* of a conclusion; hence the name *plausible* or *inductive reasoning*.

This interpretation of probability, pioneered by Bernoulli (1713), Bayes & Price (1763), and Laplace (1774) in the 18th century, was largely forgotten in the 20th century, with the *frequentist* interpretation dominating for the most part. In the frequentist approach, probability is allowed to describe only the relative frequency of an outcome in a long series of independent repetitions of a *random* experiment. This is too idealised and narrow a definition to handle many real-world scientific problems. Lacking a foundation that stems from probability theory, we are forced to select a *statistic* such as mean or variance, from intuition and experience, and apply *ad hoc* methods to perform statistical inference (Sivia & Skilling, 2006). In contrast, formulated as an extension of logic with no reference to *chance* or *random variables*, the concept of probability acquires greater flexibility and hence, wider applicability. Armed with the ability to account for prior information and *marginalise*[*] over a subset of parameters, this interpretation is well-equipped to tackle problems of inference.

---

[*] See section 2.2.2.

Any scientific experiment undergoes an initial exploratory phase in which there is very little additional information present beyond the raw data. In such cases, the optimal conclusions are drawn by applying the principle of *Maximum Entropy* (MaxEnt) (Shannon, 1948; Jaynes, 2003). MaxEnt requires one to define only a *sample space* and strives to maximise the *information entropy*[*] which generates the optimal model out of the data. For our purposes, we concern ourselves with problems of inference that are developed beyond the exploratory phase to the point where we may deploy the powerful *Bayesian* approach, which is a particular case of the extended logic formulation. In addition to defining a sample space, this approach requires us to quantify any prior knowledge we may have about the problem, in terms of a *hypothesis space*. If the extra hypotheses are true, the results of the Bayesian inference will improve on the results of MaxEnt; otherwise, they will likely be worse (Jaynes, 2003).

2.1 Rules for Logical Consistency

Let us introduce some notation that we will use in the rest of this book. A logical proposition is denoted by A. $\overline{A}$ denotes the *negation* of A i.e. the proposition that A is false. The vertical bar '|', read as *given*, means that all propositions that appear to its right are taken as true.

2.1.1 Cox's desiderata

Richard Cox (Cox, 1946) set forward three basic desiderata for logical and consistent reasoning about propositions (see e.g. Sivia & Skilling, 2006):

(i) Comparability: To each proposition about which we reason, we must assign a degree of plausibility in such a way that when new evidence presents itself, we are able to revise our assignments consistently. This would mean that we be able to rank three or more propositions *transitively*[†]. This can be achieved by representing degrees of plausibility by *real numbers*.

(ii) Common Sense: This requirement gives a sense of direction to the way the reasoning is to be carried out. Thus, for instance, if C gets updated to C' and $(A|C') > (A|C)$, then it must produce a corresponding decrease in the plausibility of $\overline{A}$ or, if any proposition B dependent on A is unaffected by the change in C, then the plausibility that both A and B are true can only increase, never decrease. This is just a qualitative correspondence to 'common sense' (Jaynes, 2003).

(iii) Consistency: If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same conclusion. Moreover, all the evidence relevant to the current problem

---

[*] This is a property of any probability distribution. It is distinct from *experimental entropy*, which is a property of a thermodynamic state as defined by some observed quantities of a physical system (Jaynes, 2003).

[†] Transitivity is the property that between three propositions A, B, and C, if A is more plausible than B, which in turn is more plausible than C, then A is more plausible than C. Real numbers, for instance, are transitive.

must be taken into account and equivalent states of knowledge must be represented by equivalent plausibility assignments.

Jaynes (2003) shows that the quantitative rules for inference, namely, the *sum rule* and the *product rule*, can be derived from Cox's desiderata using *Boolean algebra*.

2.1.2 The Sum Rule

The sum rule states that for any proposition A, the probability that A is true plus the probability that A is false is equal to unity:

$$P(A|I) + P(\overline{A}|I) = 1, \tag{2.1}$$

where P denotes the probability function. I stands for any relevant background information. Hence $P(A|I)$ means that the probability of A is *conditional* on I. Cox's desiderata require that the function P only be any continuous monotonic increasing function. Without loss of generality, we may consider it to range from 0 (impossibility) to 1 (certainty), so that it can be written in a form familiar to us as a probability function (Jaynes, 2003).

2.1.3 The Product Rule

The product rule states that if we specify the probability of B being true, and the probability of A being true given that B is true, then we have implicitly specified the probability of A and B both being true (e.g. Sivia & Skilling, 2006):

$$P(A|I)\,P(B|A,I) = P(A,B|I) = P(B|I)\,P(A|B,I). \tag{2.2}$$

As with the sum rule, I denotes relevant background information; 'A, B' denotes the *conjunction* 'A and B'. The presence of I in both the sum and the product rules indicates that there is no such thing as an absolute probability. Any probability assignment necessarily reflects only a state of knowledge.

2.2 Bayes' Theorem and Marginalisation

Together, the sum rule and the product rule form the basis of probability theory. Let us consider two important results that follow from them, to which we will find ourselves returning throughout the course of this work.

2.2.1 Bayes' Theorem

Probability calculations can be divided into two categories: *forward* and *inverse* (Mackay, 2003). Although both involve a *generative* model of a process that gives rise to the observed data, they differ in the kind of quantity for which the probability distribution is computed. Forward probability problems compute the probability distribution of some quantity *produced* by the process, also called an *observable*. Inverse probability problems aim to compute the conditional probability of one or more *unobserved variables*, given the observables. In most real-world situations, we find ourselves in the position of having to deal with precisely such inverse problems. *Bayes' theorem*, which can be obtained from the product rule, is adequate for handling them. Rearranging the terms in equation (2.2) and replacing A and B by D (for *data*) and H (for *hypothesis*) respectively, we have that

$$P(H|D,I) = \frac{P(H|I)\,P(D|H,I)}{P(D|I)}. \tag{2.3}$$

The quantity $P(H|I)$ represents our state of knowledge about H before we analyse the data. Any probability that is conditional on I alone is called *prior probability* Π, often shortened to *the prior*. We must note that 'prior' here means 'prior to the analysis of the data', not prior in *time*. This means that new information that comes to light after the data have been collected becomes part of the background information I on which our belief in the validity of H is conditional. There is no universal rule for assigning priors. Principles such as maximum entropy and marginalisation aim to address this issue and have been successful in solving different aspects of this problem (Jaynes, 2003).

The analysis of the data modifies the prior by means of the term $P(D|H,I)$, which represents the *likelihood function* or the *sampling probability*, $\mathcal{L}$. It is important to understand that the likelihood function is not a probability distribution since it is a function of the hypothesis, not of the data (Mackay, 2003). The expression $\mathcal{L}(H|D,I)$ is often used in place of $P(D|H,I)$. The likelihood reflects how the data are obtained.

The term on the left-hand side $P(H|D,I)$ is called the *posterior probability* or *the posterior* and denotes our modified belief in hypothesis H after the data have been analysed. The denominator of equation (2.3) is a normalising constant called the *Bayesian evidence* or *marginal likelihood* denoted by $\mathcal{Z}$, which does not depend on the hypothesis being evaluated. For problems of *parameter estimation*, it is not necessary to calculate this factor since it does not have any bearing on the parameter estimates. For problems of *model selection*, where we compare models (or hypotheses) of differing complexity, the Bayesian evidence becomes a valuable tool which lets us rank models according to their respective probabilities (section 2.3).

2.2.2 Marginalisation

The marginal likelihood $\mathcal{Z}$ is obtained by a process called *marginalisation*. For two propositions A and $\overline{B}$, the product rule states that (Sivia & Skilling, 2006)

$$P(A,\overline{B}|I) = P(A|I)\,P(\overline{B}|A,I). \tag{2.4}$$

Adding equations (2.2) and (2.4), we have that

$$P(A,B|I) + P(A,\overline{B}|I) = P(A|I)\,\left[P(B|A,I) + P(\overline{B}|A,I)\right]. \tag{2.5}$$

Applying the sum rule (equation 2.1) to the quantity in the square brackets, we obtain the marginalisation equation for two propositions:

$$P(A, B|I) + P(A, \overline{B}|I) = P(A|I) = \mathcal{Z}. \tag{2.6}$$

Now, if $\{B_k\}$ is a set of M propositions such that $B_1, B_2, \ldots, B_M \equiv \{B_k\}$, then the equation above can be generalised to

$$P(A|I) = \sum_{k=1}^{M} P(A, B_k|I). \tag{2.7}$$

This generalisation holds as long as the following normalisation requirement is satisfied (for *mutually exclusive* and *exhaustive* propositions, this requirement is satisfied automatically (Sivia & Skilling, 2006)):

$$\sum_{k=1}^{M} P(B_k|A, I) = 1. \tag{2.8}$$

In the *continuum limit*, where we must consider an arbitrarily large number of propositions within a large enough range (e.g. estimating the distribution of a continuous variable), equation (2.7) generalises to an integral, with the integrand denoting the *probability density function*[*]

$$P(A|I) = \int_{-\infty}^{+\infty} P(A, B|I)\, \mathrm{d}B \tag{2.9}$$

and the normalisation constraint (equation 2.8) becomes

$$\int_{-\infty}^{+\infty} P(B|A, I)\, \mathrm{d}B = 1. \tag{2.10}$$

Marginalisation is useful in inference problems because it enables us to account for *nuisance parameters*. A nuisance parameter is any parameter whose distribution does not interest us, but which modifies the distribution of the parameters in which we *are* interested. For instance, measuring the noise in an experiment may not interest us, but it must be accounted for in the data analysis. Such parameters can be integrated out to leave us with the distribution of the parameters that do interest us. A special case is the Bayesian evidence in the denominator of equation (2.3), which is obtained by integrating the numerator or marginalising over the entire set of parameters Θ of length N belonging to the hypothesis H:

$$\mathcal{Z} = \int P(H|I)\mathcal{L}(\Theta|D, H) \quad \mathrm{d}^N\Theta. \tag{2.11}$$

---

[*] The probability density function of a continuous variable ϑ gives the probability that the value of ϑ lies between ϑ and $\delta\vartheta$.

2.3 Two Levels of Inference

Inference problems can be classified into two levels (Mackay, 1991). At the first level, we assume a model is true and compute how well it agrees with the data; typically, this is undertaken by inferring how the model parameters are distributed in the sample space. At the second level, the evidence obtained for each model during parameter estimation is used to assign the models an order of preference.

2.3.1 Parameter Estimation

Assuming that model H is true, we wish to estimate the parameters $\Theta = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_n\}$ of H given the data D, using Bayes' theorem. This involves estimating the posterior probability distribution of the model parameters:

$$P(\Theta|D,H) = \frac{\mathcal{L}(\Theta|D,H)\,P(\Theta|H)}{P(D|H)}. \tag{2.12}$$

The first step is to construct the likelihood function $\mathcal{L}$. The likelihood incorporates how the uncertainties in the measurement process are distributed. For example, for an experiment with random (i.e. Gaussian) background noise, a *Gaussian* distribution best describes the likelihood function; for an experiment that involves measuring counts, a *Poisson* distribution is most suitable*.

The next step is to choose a prior distribution that reflects our state of knowledge about the model parameters as accurately as possible. Two observers may well have different prior beliefs based on their theoretical outlook and past experiences. But as long as their priors allow for regions in the parameter space where the likelihood is large, repeated application of Bayes' theorem will lead to the posterior probabilities converging to an objective inference on the hypothesis (Trotta, 2008). This is illustrated schematically in Figure 2.1. For small sample sizes and an inadequate number of data points in cases such as image reconstruction, where the number of free parameters of the model may be more than

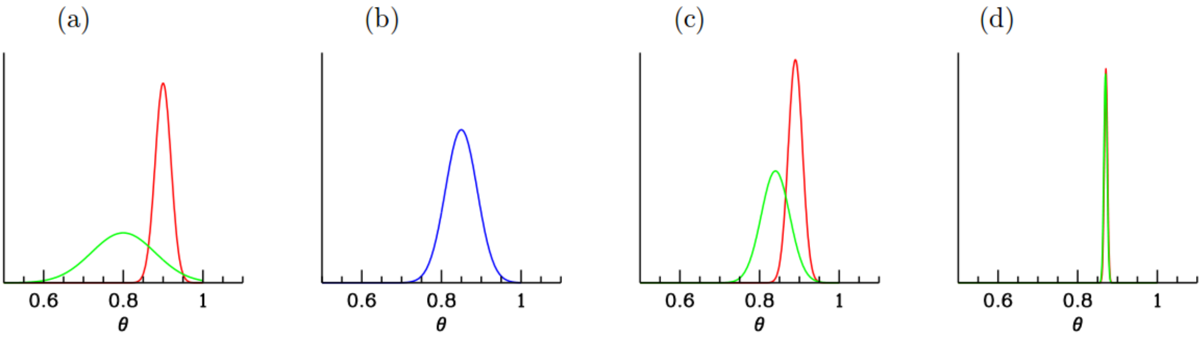


Figure 2.1: (a) The prior beliefs of two observers (green & red) about a quantity ϑ; (b) datum with likelihood $\mathcal{L}$; (c) change in beliefs (posterior distributions) after observing the datum; (d) posteriors after observing 100 data points (Trotta, 2008).

---

[*] Discussion on different probability distributions can be found in e.g. Jaynes (2003); Mackay (2003); Sivia & Skilling (2006).

the number of observations, the prior choice has more impact on the posterior. Even in such cases where different priors lead to different posteriors, we would still have learnt that the data do not have enough information to override the prior beliefs with which each observer started (Trotta, 2008).

The results of Bayesian parameter estimation may be reported in various ways. The most complete statement that one can make about posteriors is to report the probability distribution in its entirety. Under the *maximum-likelihood* (ML) method, the values of the parameters that maximise the likelihood of the observed data are taken to be their representative values (Sivia & Skilling, 2006). Another alternative is to report the maxima of the posteriors, the *Maximum A Posteriori* (MAP) values. A MAP estimate is the Bayesian equivalent of the Maximum-Likelihood estimate and corresponds to the mode of the posterior under a given parametrisation of the model[*] (Sivia & Skilling, 2006). But none of these methods is representative of Bayesian inference in general since they are all *point estimates* where the results are expressed as single values. In cases where the posterior can be approximated by a Gaussian, the results may be summarised by two parameters, the mean and the standard deviation. But this will not work for posteriors that are not easily expressed analytically or possess multiple modes and we must resort to presenting the posteriors in full.

Often, we would like to report the posteriors of a subset of Θ. This can be obtained by marginalising the joint posterior over the unwanted parameters. This marginalised posterior distribution may then be summarised by using a *statistic* such as the *mean*, *median* or *mode* along with its *confidence interval*; the resulting two-dimensional correlations between the posteriors of interest may also be presented. The evidence is not important for parameter estimation since it acts only as a normalisation constant and does not affect the shape or location of the posterior distribution in the sample space.

2.3.2 MODEL SELECTION

Unlike the more orthodox methods of sampling theory, in which no consistent method of model comparison exists[†], Bayesian inference provides a mathematical basis for assigning degrees of plausibility to models (Jaynes, 2003). Using Bayes' theorem, we can calculate the posterior probability of each model H_i, given the data D[‡]:

$$P(H_i|D,I) \propto P(D|H_i,I)\,P(H_i|I). \tag{2.13}$$

$P(D|H_i,I)$, which is the likelihood for the model H_i given the data, is the normalisation constant $\mathcal{Z}$ from equation (2.12). In model comparison, this term expresses how well the observed data predict H_i, and serves as the *evidence* for H_i. The evidence is a quantitative measure of how much the

---

[*] Unlike the ML estimate, a MAP estimate is not invariant under reparametrisation.

[†] For instance, the *chi-squared goodness-of-fit* test is an inadequate measure of performing model comparison, since a more complex model can always explain the data better (Mackay, 2003).

[‡] We omit the normalising constant since we may need to develop new models *after* the data are collected, depending on the adequacy (or lack thereof) of current models (Mackay, 2003).

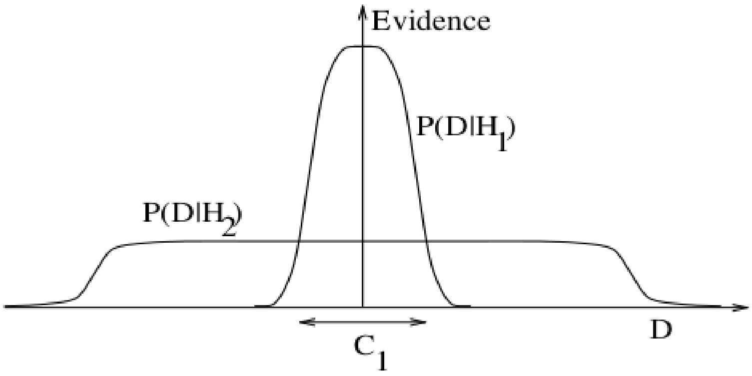


Figure 2.2: Occam's razor: The horizontal axis represents all possible data sets D. The Bayesian evidence for a model is proportional to its *predictive power* i.e. how much it predicted the data that occurred. Here, the evidence for H_1, a simpler model with fewer parameters, is higher than the evidence for H_2, a more powerful model that can predict a wider range of possible data sets. If the actual data occurred within C_1, then H_1 will have the higher evidence, as is the case here (Mackay, 2003).

data favour one model over another. Unless a complicated model with more parameters (a high-dimensional parameter space) is significantly better at explaining the data (i.e. has a higher likelihood), its evidence will be smaller than that of a simpler model with fewer parameters that can explain the data equally well (Figure 2.2). To rank two models H_1 and H_2, we may compute the ratio of their posteriors

$$\frac{P(H_1|D,I)}{P(H_2|D,I)} = \frac{P(D|H_1,I)}{P(D|H_2,I)}\,\frac{P(H_1|I)}{P(H_2|I)}. \tag{2.14}$$

The ratio of the prior distributions may often be set to unity to indicate that there is no prior preference for one model over another. The ratio of model evidences is also known as the *Bayes factor*, B_{12} (Jeffreys, 1961):

$$B_{12} \equiv \frac{\mathcal{Z}_1}{\mathcal{Z}_2} = \frac{P(D|H_1,I)}{P(D|H_2,I)}. \tag{2.15}$$

The higher this factor, the more is H_1 preferred over H_2. It is often useful to consider the natural logarithm of the Bayes factor:

$$\ln B_{12} = \ln \mathcal{Z}_1 - \ln \mathcal{Z}_2. \tag{2.16}$$

Thus the Bayes factor measures the relative success of H_1 and H_2 in predicting the data and the logarithmic Bayes factor is also called the *weight of evidence* (e.g. Good, 1985). Kass & Raftery (1995) propose an empirical scale to measure how strongly one model is preferred over another following Jeffreys (1961) (Table 2.1).

2.4 NUMERICAL SAMPLING TECHNIQUES

The probability distributions involved in Bayesian inference are seldom well-behaved i.e. adequately expressed by an analytic distribution. In many real-world problems, the likelihood function can be

Table 2.1: Criteria for model selection. B_{12} denotes the ratio of the evidences between hypotheses H_1 and H_2 (Kass & Raftery, 1995).

$2 \ln B_{12}$	B_{12}	Evidence against H_2
0 to 2	1 to 3	Not worth more than a mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very strong

multi-modal[*] or not easily expressed analytically (Mackay, 2003). Moreover, except in the cases of the simplest models, calculating the multi-dimensional integral that is the Bayesian evidence is computationally prohibitive. In such cases, we resort to numerical sampling techniques such as *Monte Carlo* methods to obtain random samples from a distribution (here, the posterior) (e.g. Neal, 1993). The distribution of these sampled values will represent the posterior more and more accurately as the number of samples increases.

The most widely used Monte Carlo methods are *Markov Chain Monte Carlo* (MCMC, Metropolis et al., 1953; Hastings, 1970) and *Nested Sampling* based techniques (Skilling, 2004). Here we concentrate on nested sampling and associated algorithms.

2.4.1 NESTED SAMPLING

MCMC algorithms focus on sampling the posterior, leaving the Bayesian evidence to be calculated by some other technique, if at all it is deemed necessary. In contrast, nested sampling focuses on calculating the evidence (Skilling, 2004). The reasoning behind this is that by computing the evidence and leaving posterior samples to be generated as a by-product, one can always present, along with the posteriors, the evidence as a quantitative measure of how well a model fits the data.

Denoting the likelihood by $\mathcal{L}(\Theta)$ and the prior by $\Pi(\Theta)$, where Θ is the vector of parameters of length Nbelonging to the model in question, the evidence may be written as

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\Pi(\Theta)\, \mathrm{d}^N\Theta. \tag{2.17}$$

We can imagine sorting the elements of the likelihood in decreasing order (Sivia & Skilling, 2006). Once sorted, we can define a *prior volume* as

$$\xi(\lambda) = \int_{\mathcal{L}(\Theta)>\lambda} \Pi(\Theta)\, \mathrm{d}^N\Theta. \tag{2.18}$$

$\xi(\lambda)$ gives the proportion of the prior with likelihood greater than λ. The evidence integral in equa-

---

[*] The distribution contains several different modes or maxima.

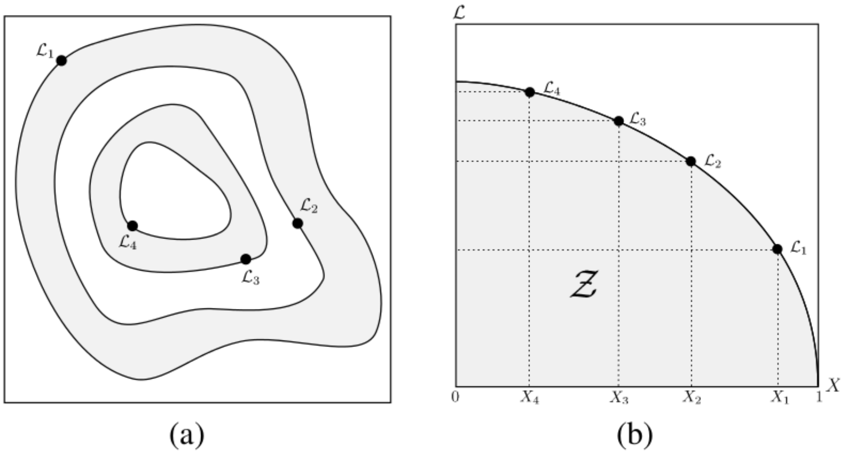


Figure 2.3: (a) The posterior of a two-dimensional problem with the iso-likelihood regions indicated by the contours; and (b) the transformed $\mathcal{L}(\xi)$ function where each ξ_i is associated with the likelihood $\mathcal{L}_i$ (Feroz & Hobson, 2008).

tion (2.17) can now be written as

$$\mathcal{Z} = \int_0^1 \mathcal{L}(\xi)\,\mathrm{d}\xi. \tag{2.19}$$

If the likelihoods $\mathcal{L}_i = \mathcal{L}(\xi_i)$ can be evaluated, where ξ_i is a sequence of decreasing values,

$$\xi_0 = 1 > \xi_1 > \xi_2 > \ldots > \xi_M > 0, \tag{2.20}$$

then the evidence can be approximated numerically using standard quadrature methods for integration (e.g. Press et al., 2007):

$$\mathcal{Z} = \sum_{i=1}^{M} \mathcal{L}_i w_i, \tag{2.21}$$

where, using the trapezium rule (Press et al., 2007), we have for the weights w_i,

$$w_i = \frac{1}{2}(\xi_{i-1} - \xi_{i+1}). \tag{2.22}$$

This is schematically shown in Figure 2.3. The summation is carried out iteratively. The iteration counter is set to $i = 0$ and N active (or *live*) points are sampled from the entire prior $\Pi(\Theta)$, so that initially $\xi_0 = 1$. The samples are then sorted according to equation (2.20) and the sample with the least likelihood $\mathcal{L}_0$ is removed from the active set. This is replaced with a new point subject to the constraint that it has a higher likelihood $\mathcal{L} > \mathcal{L}_0$. Thus, after each iteration i, the prior volume is reduced, with each new sample possessing the likelihood $\mathcal{L}_{i+1} > \mathcal{L}_i$, until the entire prior volume is traversed (Feroz & Hobson, 2008). The algorithm finally terminates when the evidence is determined to a specified precision. Once the evidence has been computed, the posterior can be generated using the full sequence of active and inactive points by simply assigning the weights

$$p_j = \frac{\mathcal{L}_j w_j}{\sum_{i=1}^{M} \mathcal{L}_i w_i} = \frac{\mathcal{L}_j w_j}{\mathcal{Z}} \tag{2.23}$$

to each point, where the sample index j runs from 1 to $M + N$, the total number of samples. The resulting joint posterior can then be marginalised and presented in various ways to study the reduced-dimensional posteriors of parameters of interest and their correlations.

2.4.2 FRIENDS OF NESTED SAMPLING

Drawing samples from the prior at each iteration i subject to the constraint $\mathcal{L} > \mathcal{L}_i$ is not trivial. Drawing blindly from the prior would mean that the *acceptance rate*[*] would decrease with decreasing prior volume as we move on to regions of higher likelihoods. *Ellipsoidal nested sampling* (Mukherjee et al., 2006) attempts to address this by approximating the iso-likelihood contour $\mathcal{L}_i$ (Figure 2.3) with a D-dimensional ellipsoid determined from the covariance matrix of the active points. The sampling is undertaken from within this ellipsoidal bound. This works for simple unimodal posteriors, but for posteriors with multiple modes, the sampling efficiency can be improved by identifying distinct *clusters* of active points and constructing ellipsoidal bounds around them (Shaw et al., 2007) (Figure 2.4). Feroz & Hobson (2008) improved on this by partitioning the set of active points into as many clusters as possible to account for highly-degenerate posteriors, at the cost of reduced efficiency for simpler problems. In Feroz et al. (2009), they proposed the MULTINEST algorithm which improves on all the above methods.

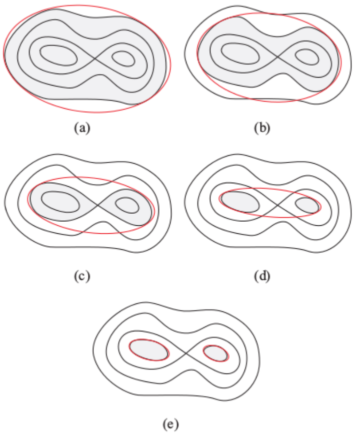


Figure 2.4: Improved ellipsoidal sampling for a simple bimodal case. (a) Approximating the active region with an ellipsoid. In (b) to (d) we see that the acceptance rate steadily worsens. (e) provides a solution by identifying clusters and constructing ellipsoids around the two modes for more efficient sampling (Shaw et al., 2007).

---

[*] The acceptance rate is the fraction of proposed samples that is accepted for use in the following iteration by the sampler.

MultiNest partitions the current set of N active points and constructs a new set of ellipsoidal bounds at each iteration i. The algorithm requires that all partitioning and ellipsoidal bound construction and sampling be performed in a *unit hypercube*. This is accomplished by letting each parameter value vary from 0 to 1 and sampling uniformly from this space. Then the N live points are partitioned by optimising the *Bayesian Information Criterion* using X-means (Pelleg & Moore, 2000); each cluster is further subdivided into subclusters by using a modified k-means* algorithm (Feroz et al., 2009).

For multimodal posteriors, it is necessary to identify which samples belong to which mode. For modes that are well-defined and located far enough apart in the parameter space, this is easy to accomplish. For modes that are close to one another, we may need to reach relatively high likelihood levels before they separate out. An illustration is provided in Figure 2.5. G_1 is the first group that remains unsplit until iteration $i = i_1$, at which stage it is split into G_2, G_3, and G_4. During $i = i_2$, G_3 is split into G_5, G_6, and G_7, and G_4 is split into G_8 and G_9. G_2 remains unsplit at $i = i_2$ but contains no active points. Any group that remains active at the end of the nested sampling process would then be promoted to a mode (Feroz et al., 2009).

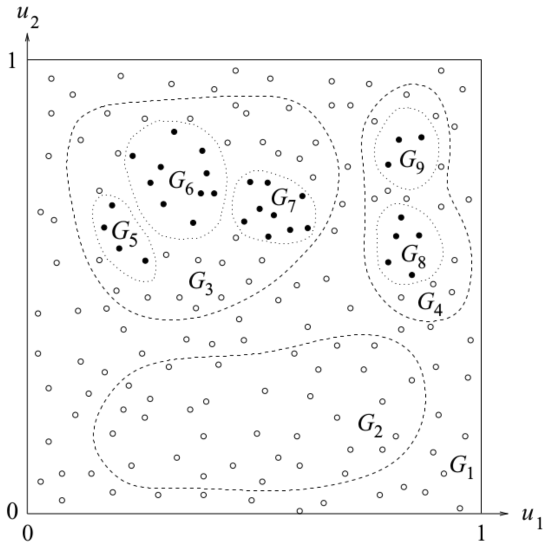


Figure 2.5: Assigning points to groups. The dashed and dotted lines denote the iso-likelihood contours $\mathcal{L}_{i_1}$ and $\mathcal{L}_{i_2}$ respectively. The solid circles are the active points at $i = 2$ and the open circles are the inactive ones (Feroz et al., 2009).

MultiNest provides reliable evidence estimates with associated uncertainties and generates accurate posteriors for high-dimensional problems at a fraction of the computational cost one would incur with other existing MCMC-based techniques. *Importance Nested Sampling* is yet another improvement on nested sampling that provides evidences up to an order-of-magnitude higher accuracy than MultiNest (Feroz et al., 2013).

---

[*] k-means is the method of partitioning n observations into k clusters such that each observation belongs to the cluster with the nearest mean value (e.g. Mackay, 2003).

In solving a problem of this sort, the grand thing is to be able to reason backwards.

Sherlock Holmes

3

Bayesian Inference for Radio Observations

ESTIMATING THE SOURCE AND INSTRUMENTAL[*] parameters from observed interferometric visibilities falls under the domain of *inverse problems*, which deal with the issue of determining the underlying phenomena (the causes) from the observed data (the results) (Parker, 1977). Inverse problems are *ill-posed* i.e., no unique solution that describes the observed data exists. More than one combination of the model parameter values (or an entirely different set of parameters described by a different model) could have given rise to the same data, and hence it becomes necessary to impose additional constraints in order to obtain the most sensible explanation for the cause. This process is known as *regularisation* and, under the Bayesian probability formalism, is achieved by introducing the concept of prior probabilities (section 2.2.1). Parametric model-fitting is one of the most widely-used techniques in handling inverse problems. We devise a model with adjustable parameters, which we believe is capable of describing the data, and choose a *figure-of-merit* or *merit function* to measure how well the data and the model agree (e.g. Press et al., 2007, Chapter 15). In the present case, the merit function is the posterior probability distribution and the data are the observed visibilities that are best modelled using the RIME (section 1.3.2).

3.1 PROBABILISTIC VISIBILITY ANALYSIS

Probabilistic analysis of visibilities complements the conventional imaging and deconvolution techniques and, if applied judiciously, can improve on them. The widely used technique of making images from visibilities (section 1.3.3.2) may not always be the best way to make sense of interferometric data. The Fourier transform involved in the imaging process distributes the localised *uv*-domain un-

---

[*] The work detailed in this chapter was published as part of *Bayesian Inference for Radio Observations* (Lochner et al., 2015).

certainties between the pixels of the image, thereby correlating the uncertainties in the image domain (Taylor et al., 1999, Chapter 16). In observations with sparse *uv*-coverage or poor calibration, the information available is inadequate to form a reasonable image of the sky, owing to the fact that the *correlations* between the sky and instrumental parameters are difficult to study using conventional imaging techniques. Moreover, the measurements are made in the *uv*-domain and so error analysis is most naturally done in that domain.

Conventional imaging methods such as CLEAN, although not Bayesian, rely on *a priori* information and models of the sky and the instrument, to choose the most plausible image of the radio sky from among other possible realisations (see section 1.3.3.2). Bayesian inference has been performed on radio observations before in more restricted contexts. (Hobson & Maisinger, 2002) presented a maximum-likelihood method for estimating the power spectrum of the cosmic microwave background directly from visibilities. MCADAM, a Bayesian inference package for analysing interferometric data was developed and used for analysing galaxy cluster data by Marshall (2003). Lancaster et al. (2005) used Bayesian inference methods to analyse galaxy cluster data observed using the Very Small Array (VSA). Feroz et al. (2009) and Zwart et al. (2011) performed Bayesian analysis of visibilities from galaxy-cluster observations with the Arcminute Microkelvin Imager (AMI) telescope for detecting clusters using the Sunyaev-Zel'dovich effect. More recently, Bayesian inference-based techniques have been developed to address various problems in imaging. Junklewitz et al. (2015, 2016) have devised an algorithm called RESOLVE (Radio Extended SOurces Lognormal deconVolution Estimator) for interferometric imaging of extended and diffuse emission in total intensity using Bayesian inference in the framework of information field theory (Enßlin et al., 2009), which shows improved performance against Multiscale-CLEAN (Cornwell, 2008) and MEM on simulations.

Here we present Bayesian Inference for Radio Observations (BIRO, Lochner et al., 2015), a fully Bayesian framework for analysing visibilities by modelling them using the RIME and deriving the posterior distributions of and correlations between model parameters. Using the RIME, in theory, enables us to model any linear propagation path effect along with the source parameters.

3.2 SETTING UP THE BIRO FRAMEWORK

Our primary assumptions are that (i) the visibility measurements are independent and (ii) the uncertainties in the measurements follow a Gaussian distribution. If the data are independent, then the likelihood $\mathcal{L}$ from equation (2.12) can be expressed as (Sivia & Skilling, 2006, Chapter 3)

$$\mathcal{P}(D|\Theta, H) = \prod_{k=1}^{N_{\text{vis}}} \mathcal{P}(D_k|\Theta, H) \quad , \tag{3.1}$$

where D_k stands for the k^{th} datum and N_{vis} denotes the number of visibilities. Under the assumption that the visibility noise is Gaussian, the probability of obtaining each individual datum is given by

$$\mathcal{P}(D_k|\Theta, H) = \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(F_k - D_k)^2}{2\sigma_k^2}\right) \quad , \tag{3.2}$$

where $\{\sigma_k\}$ denotes the expected uncertainties and $F_k(\Theta, k)$ denotes the functional form of the model for a specific set of parameter values Θ for the k^{th} datum. Combining equations (3.1) and (3.2), we obtain

$$\mathcal{P}(D|\Theta, H) \propto \exp\left(-\frac{\chi^2}{2}\right) \quad , \tag{3.3}$$

where the sum of the squares of the *normalised residuals* is represented by

$$\chi^2 = \sum_{k=1}^{N} \left(\frac{F_k - D_k}{\sigma_k}\right)^2 \quad . \tag{3.4}$$

The χ^2 measures the deviation of the data from the expected value. If we also assume a flat (uninformative) prior for the parameters to indicate that we are largely ignorant of their expected values, then, taking the natural logarithm of equation (2.12) and omitting the details of normalisation constant, we arrive at

$$\ln[\mathcal{P}(\Theta|D, H)] = \mathrm{C}(\sigma_k) - \frac{\chi^2}{2} \quad . \tag{3.5}$$

where $\mathrm{C}(\sigma_k)$ depends solely on σ_k. Thus, with some simplifying assumptions, Bayesian parameter estimation reduces to the more familiar methods of *maximum likelihood* and *least-squares estimation*. The exact form our likelihood function takes is very close to this (section 3.3.1.4). The crucial difference is that Bayesian inference is independent of the Gaussianity of the underlying process.

In BIRO, I start by setting out to apply probability theory to synthetic interferometric observations made using MeqTrees (Noordam & Smirnov, 2010). The likelihood function and the priors are supplied by the user and the software interfaces with MeqTrees and PyMultiNest (Buchner et al., 2014) using python (van Rossum, 1995).

3.2.1 MeqTrees

MeqTrees can implement an arbitrary measurement equation and solve for arbitrary subsets of its parameters, with the measurement equation implemented using *tree* data structures[*] (Knuth, 1973). A tree consists of *nodes* that are connected to each other in a parent-child hierarchy by *edges*. A node without a child is a *leaf* and a node without a parent is a *root*. No cycles are allowed and a group of trees, interlinked or otherwise, is collectively called a *forest*.

Any mathematical expression lends itself easily to being represented by trees (Figure 3.1). The

---

[*] Since MeqTrees allows for multiple parent nodes, the proper term is a *directed acyclic graph* or DAG.

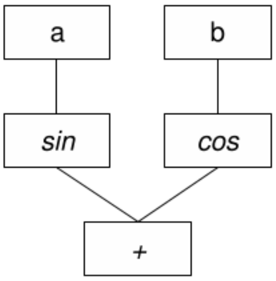


Figure 3.1: A simple mathematical expression, $\sin a + \cos b$, represented as a tree (Noordam & Smirnov, 2010).

value of the expression is computed by starting at the leaves and propagating their values through their parents until we reach the root. In MeqTrees, the results of this computation may be *functions*, not just a single value. MeqTrees implements different *classes* of nodes such as MeqConstant, MeqFreq, or MeqFITSImage, which all return different types of functions — a constant, a polynomial, or a FITS image[*] respectively. A forest of trees computes multiple functions, which together constitute a numerical model with solvable parameters (Noordam & Smirnov, 2010). A generic RIME such as equation (1.35) is implemented in MeqTrees as shown in Figure 3.2. Besides

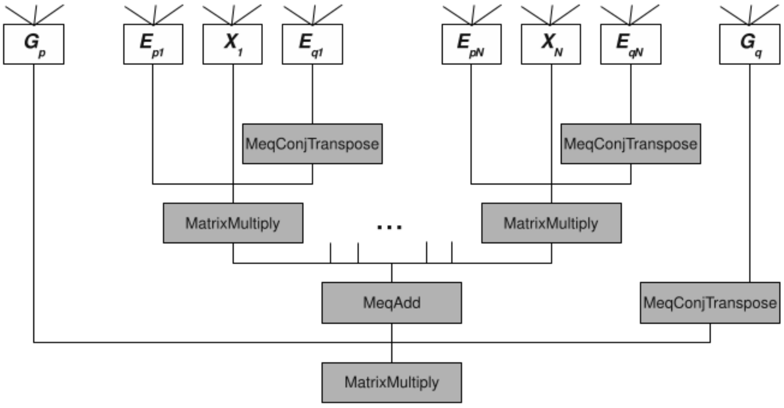


Figure 3.2: A subtree showing the generic RIME (equation 1.35) implemented in MeqTrees (Noordam & Smirnov, 2010).

generating the synthetic observations for the analysis, MeqTrees also generates the model visibilities for computing the likelihood. MeqTrees comprises two components — meqbrowser and meqserver. meqbrowser is the graphical user interface to MeqTrees. meqbrowser provides a variety of visualisation options and allows the user to explore the current RIME tree-by-tree. The *Tree Definition Language* (TDL), a python-based scripting language (Noordam & Smirnov, 2010), is used to interface with meqserver, the C++ (Stroustrup, 2013) backend to MeqTrees. TDL scripts let the user set up a simulation or a calibration run and meqserver performs the necessary

---

[*] FITS or *Flexible Image Transport System* is an image format used widely in astronomy to store information about observations as images or ASCII/binary tables. More information can be found at `https://fits.gsfc.nasa.gov/fits_home.html`

CPU-intensive computations.

3.2.2 MultiNest and PyMultiNest

The other software that I use extensively in my analyses is PyMultiNest (Buchner et al., 2014) which provides a python interface to MultiNest. PyMultiNest accepts user-written python functions for generating prior distributions for the model parameters, as well as the likelihood function. MultiNest samples the model parameters iteratively according to the prior ranges and distributions specified by the model, and converges to region(s) of high probability in the posterior (Feroz & Hobson, 2008). Finally, it returns the Bayesian evidence for each model along with the associated uncertainties and the posterior distributions (section 2.4.2). MultiNest has some tunable parameters, the most important of which are the number of live points used and the evidence tolerance level, set depending on the required accuracy and the available computational resources.

3.3 Applying BIRO to Super-resolution

One of the first problems (Lochner et al., 2015) that I applied BIRO to was that of *super-resolution* or *over-resolution*. Super-resolution addresses the issue of extracting information about the source structure at scales smaller than that of the PSF of the interferometer. The commonly used measure for the angular resolving power of an optical system is the Rayleigh criterion (equation 1.4). The theoretical limit for the resolution of an interferometer can be obtained using Fourier optics which characterise the resolving power of a system in terms of its *spatial bandwidth* (Goodman, 1968). Diffraction effects limit the maximum spatial frequency, f_{max}, transmitted by the optical system and the resolving power R may be reformulated in terms of this frequency as

$$R = \pi / f_{max}, \tag{3.6}$$

where R is the *Nyquist distance* (Bertero & de Mol, 1996). It is possible to obtain information about the spatial frequencies that lie outside f_{max} (i.e. beyond the *diffraction limit*) by incorporating *a priori* knowledge about the brightness distribution of the source. The resolving power then depends on the precision of the measuring instrument which can be quantified in terms of the signal-to-noise ratio (SNR) of the observed data (e.g., Harris, 1964; di Francia, 1969).

Techniques for super-resolution have been widely used in radio interferometry to characterise partially-resolved compact sources. Lobanov (2005) derives the resolution limits for specific brightness-distribution templates for astronomical sources and Martí-Vidal et al. (2012) extend this to the general case of super-resolution with interferometers. A discussion of model-fitting in the *uv*-domain for specific source profiles using real and synthetic data using the uvmultifit package is provided in Martí-Vidal et al. (2014). An alternative, which achieves super-resolution by *sparse* modelling[*] in

---

[*] *Sparsity* is the ratio between the number of non-zero elements in a matrix and the total number of elements.

the sky or *image* domain in the presence of Gaussian noise generated with the same standard deviation on all baselines, is explored in Honma et al. (2014). While these works explore the theoretical constraints of super-resolution, the application of these methods to real observations, made with a specific interferometer configuration in the presence of instrumental uncertainties, will not provide us with a knowledge of how the instrumental effects correlate with the source parameters and limit the resolution. I start off my investigations with the analysis of synthetic Westerbork Synthesis Radio Telescope (WSRT)* observations of compact radio sources that smaller than the width of the PSF.

3.3.1 The Model Selection Problem

To test the power of Bayesian inference applied to the super-resolution problem, I simulated observations of three different sky models with WSRT using MeqTrees. Each model had structure at scales smaller than the size of the PSF of the interferometer. Each of the synthetic observations (with simulated random† noise) was fitted against all three models and the resulting Bayesian evidences were compared to select the most probable model describing the visibilities. The posterior distributions of the parameters of the best-fitting model were then used to make informed decisions about the source brightness, position, and structure.

3.3.1.1 Details of the Simulation

WSRT is an interferometer comprising 14 equatorially-mounted parabolic dishes of 25-metre diameter each, arranged in a 2.7 km long East–West line located at Westerbork in the Netherlands. The telescope can observe at multiple frequency bands located between 120 MHz and 8.3 GHz. In the simulations, the observations were carried out for a total of 12 hours, with an *integration time*‡ of 30 seconds. A single frequency channel 125 kHz wide and centred at 1.4 GHz was used. The *uv*-coverage of the observation is shown in Figure 3.3. At this observing frequency, WSRT has a FoV of about 0.6° and a PSF with a FWHM of about 14 arcsec. A Gaussian random noise of standard deviation 100 mJy per visibility was added to each simulation.

3.3.1.2 Model Description

I simulated WSRT observations of the following three source models, with all the sources located a few arcmin away from the pointing centre (Figure 3.4):

1. PT: A single point source of flux-density 0.85 Jy,
2. GAU: A source with elliptical Gaussian morphology of size $4'' \times 3''$ and integrated flux-density 0.85 Jy, and

---

[*] `www.astron.nl/radio-observatory/public/public-0`

[†] Gaussian-distributed noise in the real and imaginary parts of the complex visibilities.

[‡] The time interval between two data dumps from the correlator.

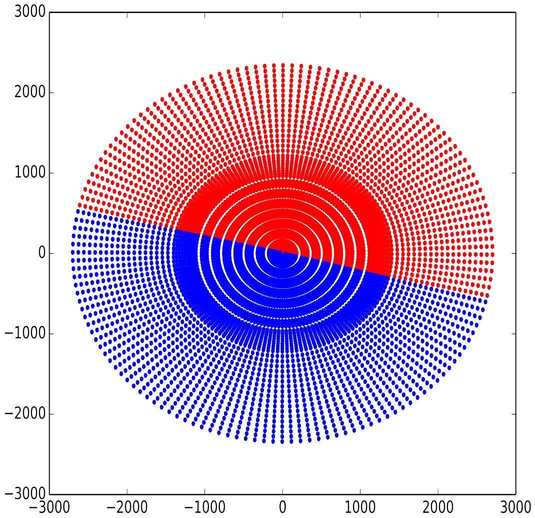


Figure 3.3: The *uv*-coverage of the WSRT simulations. Since WSRT is an East-West array, the 12-hour *uv*-coverage traces full concentric circles. The red dots are complex conjugates of the blue dots in *uv*-space.

3. 2PT: Two point sources of flux-density 0.425 Jy each and separated by a small distance (25 per cent the size of the PSF).

The three source models possess structure at scales of 0.25 the size of the PSF. I used the CASA* task CLEAN to image the simulated visibilities for comparison with BIRO. Figure 3.5 shows the CLEANed images for the models depicted in Figure 3.4. CLEAN was not able to distinguish between the three sky models. A source extractor such as PYBDSM†, which uses these images as inputs, estimates that the source is unresolved.

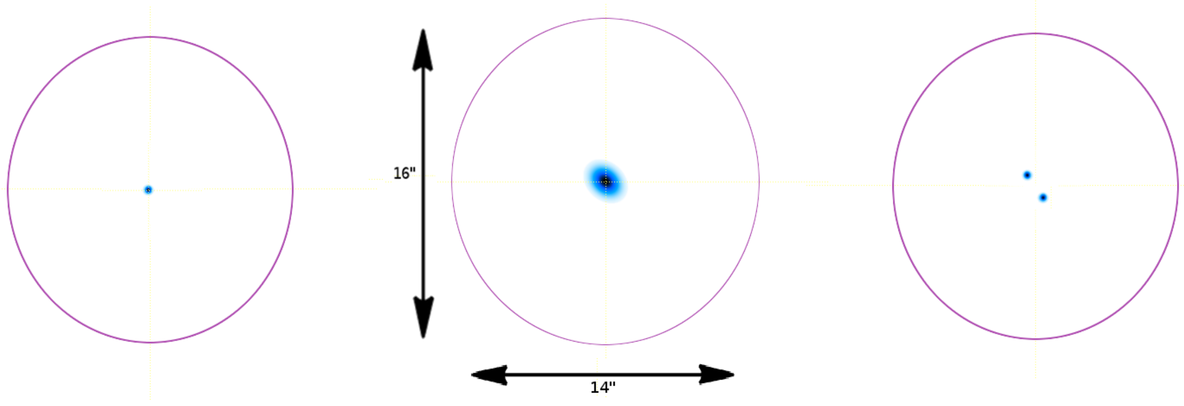


Figure 3.4: The three sky models used in the WSRT simulations. (i) a point source (PT), (ii) elliptical Gaussian (GAU), and (iii) two point sources (2PT). The red ellipses denote the FWHM of the PSF of the observations fitted by CLEAN and are $14'' \times 16''$ in size.

---

[*] https://casa.nrao.edu/

[†] www.astron.nl/citt/pybdsm

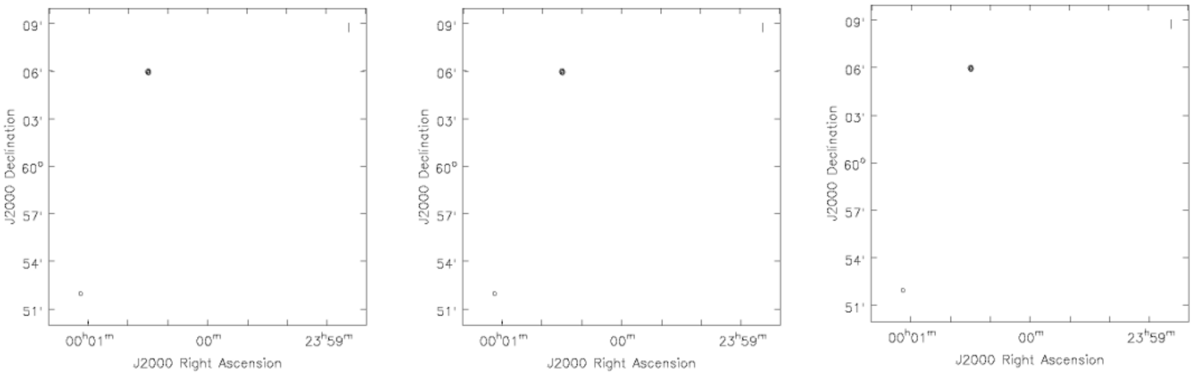


Figure 3.5: CLEANed images of the three sky models --- (i) PT, (ii) GAU, and (iii) 2PT. The contours are at 20, 40, 60, and 80 per cent of the maximum flux-density. The SNRs of the data sets are about 1000:1.

For the Bayesian analysis, we parametrised each model as follows:

PT: The single point source model consists of one flux-density parameter, S_ν, and two parameters that describe the position of the source, (l, m), the direction cosines measured with respect to the (u, v) co-ordinates.

GAU: The elliptical Gaussian model consists of S_ν and (l, m) along with three more parameters, l_p, m_p, and r, describing the shape of the source. These three parameters are related to the major axis e_{maj}, minor axis e_{min}, and position angle ϑ of the ellipse as follows:

$$\begin{aligned} l_p &= e_{\mathrm{maj}} \sin \vartheta, \\ m_p &= e_{\mathrm{maj}} \cos \vartheta, \\ r &= e_{\mathrm{min}} / e_{\mathrm{maj}}. \end{aligned} \tag{3.7}$$

l_p and m_p are the projections of e_{maj} onto the l and m axes, and r is the ratio of the minor axis to the major axis (Figure 3.6).

2PT: The two-point source model consists of the same three parameters as PT, but for two sources (a total of six parameters).

A summary of the parametrisation of each model is given in Table 3.1.

3.3.1.3 Building the RIME

We are now ready to construct the RIME for modelling the visibilities. Assuming a flat spectral index, since we use one frequency channel in the simulations, and using the flux-density parameter S_ν in the brightness matrix (equation 1.26), we have that

$$\mathrm{B} = \begin{pmatrix} S_\nu & 0 \\ 0 & S_\nu \end{pmatrix}. \tag{3.8}$$

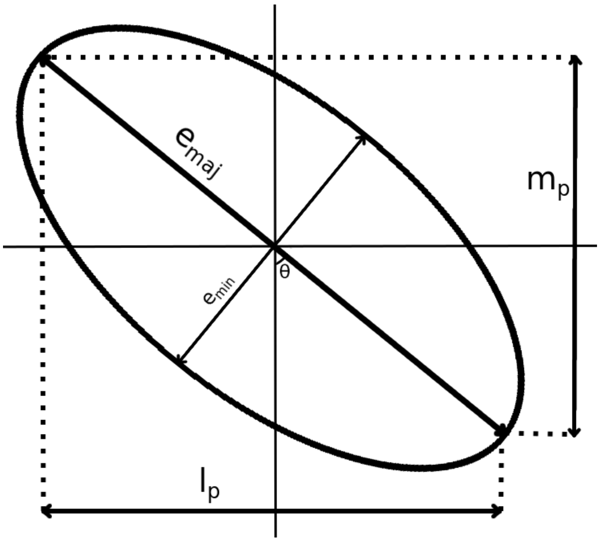


Figure 3.6: The relation between the shape parameters, l_p, m_p, and r, and the FWHM of the major axis (e_{maj}), FWHM of the minor axis (e_{min}), and position angle (ϑ) of an elliptical Gaussian profile.

Table 3.1: Models used in the analysis of the WSRT simulations.

Model	Number of parameters	Parametrisation
PT	3	Flux Density (S_ν) Position (l, m)
GAU	6	Flux Density (S_ν) Position (l, m) Shape (l_p, m_p, r)
2PT	6	Flux Density (S_{ν_1}) Position (l_1, m_1) Flux Density (S_{ν_2}) Position (l_2, m_2)

The first linear transformation this signal undergoes is represented by the phase-delay matrix K, associated with the difference in the geometric path lengths from the source to antennas p and q. Given the phase difference (κ_p) between the waves received by antenna p located at $\vec{u}_p = (u_p, v_p, w_p)$ relative to $\vec{u} = \vec{0}$, the scalar *K-Jones* term for antenna p can be written as

$$K_p = \mathrm{e}^{-i\kappa_p} \equiv \mathrm{e}^{-i\kappa_p} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.9}$$

The K-Jones term must be accounted for even under ideal conditions in which nothing else affects the signal from the source to the interferometer. Knowing this, we may write the source coherency matrix following equation (1.32) as

$$\mathrm{X}_{pq} = K_p \, \mathrm{B} \, K_q^H = \mathrm{B}\mathrm{e}^{-i\kappa_{pq}}. \tag{3.10}$$

No antenna gains or primary-beam effects such as pointing errors were added in the simulations and hence the *G-Jones* matrices for antenna gains and the *E-Jones* matrices for primary-beam effects were set to unity, simplifying the RIME considerably.

For the extended source model, the brightness distribution is integrated over the extent of the source. Expressed in terms of the direction cosines l and m, the RIME for the extended source may be written as

$$\begin{aligned} \mathrm{V}_{pq} &= \iint\limits_{lm} \mathrm{X}_{pq}(l, m) \; \mathrm{d}\Omega \; + \; \mathcal{N}(0, \sigma_{pq}^2), \\ \text{where } \mathrm{d}\Omega &= \frac{\mathrm{d}l \, \mathrm{d}m}{\sqrt{1 - l^2 - m^2}}. \end{aligned} \tag{3.11}$$

3.3.1.4 Priors and the Likelihood

We set uniform (uninformative) priors $\Pi(\Theta)$ on most parameters (Table 3.2). Since our prior ranges do not span multiple orders of magnitude, uniform priors, as opposed to logarithmic or Jeffreys priors, are sufficiently uninformative (Sivia & Skilling, 2006). The prior range for S_ν is chosen based on what we know about the flux-densities of the sources in the simulations. The prior distribution for l_p (the sine projection of e_{maj}) is restricted to non-negative values so that the position angle estimate

Table 3.2: Priors for the different parameters used. All the listed parameters were set uniform priors with the range indicated by the values in the square brackets. The parameters with delta priors are not included.

Parameter (units)	Prior distribution
S / Jy	[0, 2]
l&m / arcsec	[-25, 25]
l_p / arcsec	[0, +20]
m_p / arcsec	[-20, +20]
r	[0, 1]

is constrained to a range of 0–180°. This is done so that the sampler does not distinguish between two Gaussians inclined at position angles that are 180° apart — a Gaussian with position angle 45° should be considered equivalent to one with position angle 135°. For the model selection step, we assign equal priors to all the models considered, so that the Bayes factor B_{12} may be directly used for model comparison.

Now that we have described the models quantitatively, we can set up the likelihood function. Given the observed (V_D) and the modelled (V_M) visibilities, and the uncertainty σ denoting the standard deviation of the Gaussian-distributed noise, the likelihood function for parameter estimation for model H is given by

$$\begin{aligned} \mathcal{L}(\Theta|V_D, H) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\chi^2}{2}\right), \\ \text{where } \chi^2 &= \sum_{k=1}^{N} \left(\frac{V_{M_k} - V_{D_k}}{\sigma}\right)^2, \end{aligned} \tag{3.12}$$

and $N = 2N_{\text{vis}}$ is the total number of data points. The summation is carried out over $2N_{\text{vis}}$ since we consider the real and imaginary parts separately. Taking the natural logarithm of $\mathcal{L}$, we obtain

$$\ln(\mathcal{L}) = \ln\left[(2\pi\sigma^2)^{-N/2}\right] - \frac{\chi^2}{2}. \tag{3.13}$$

This log-likelihood is computed by the user-written likelihood function during every iteration, with a new set of sampled parameters obtained from MultiNest each time.

3.3.1.5 RESULTS

Each simulation described in section 3.3.1.2 was compared with each model. I used MultiNest for computing the evidence and the posteriors (section 2.4.2). I used 1000 live points so that the entire parameter space was sampled adequately and set the sampling efficiency to 0.3 for evidence evaluation (Feroz & Hobson, 2008). The evidence obtained for each model against each simulation was used to compute the relative evidences between models (Table 3.3). The values quoted are twice the differences in natural logarithmic evidences, computed using equation (2.16).

Table 3.3: Evidence matrix that gives twice the differences between the $\log_e$ Bayes factors of each model against others for each simulated dataset. For instance, the first row gives the odds in favour of PT against GAU and 2PT, when applied to the point source simulation. The error in relative ln-evidence in each case is less than 0.4.

Sim \ Model	PT	GAU	2PT
PT	1:1	15:1	11:1
GAU	2E4:1	1:1	4E3:1
2PT	4E4:1	34:1	1:1

Using the criteria put forth in Table 2.1, we find that the odds are *very strongly* in favour of the correct model for each simulated dataset. Once we have identified the model with the highest evidence, we may examine the corresponding posteriors returned by MultiNest. Figures 3.7, 3.8, and 3.9 show the posteriors for the parameters corresponding to the models PT, GAU, and 2PT respectively, for the corresponding simulations. We see that BIRO can not only discriminate between the models, but also estimate the source parameters with high precision and accuracy. All the posteriors closely approximate the Gaussian since these are simple simulations with only random thermal noise injected. As expected with any data affected by random noise, there are differences between the maxima of the Gaussians and the true values. Since the priors were uniform and wide enough to encompass the true values comfortably, their effects on the posterior estimates are much less significant

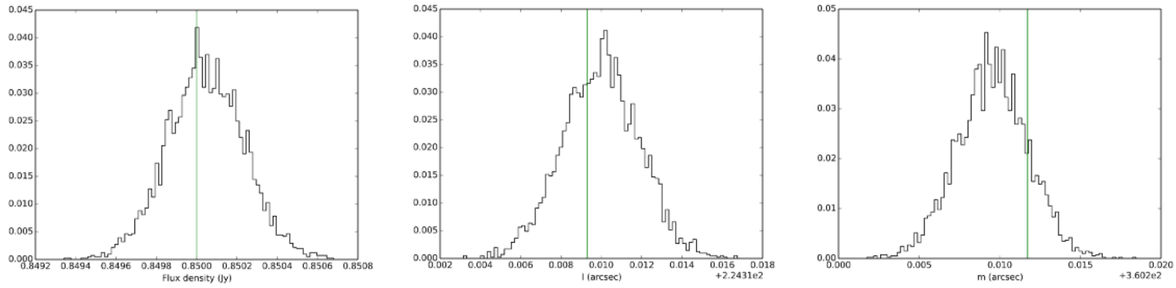


Figure 3.7: Posteriors for the parameters of the model PT. The vertical green lines indicate the `true' values that went into the simulation.

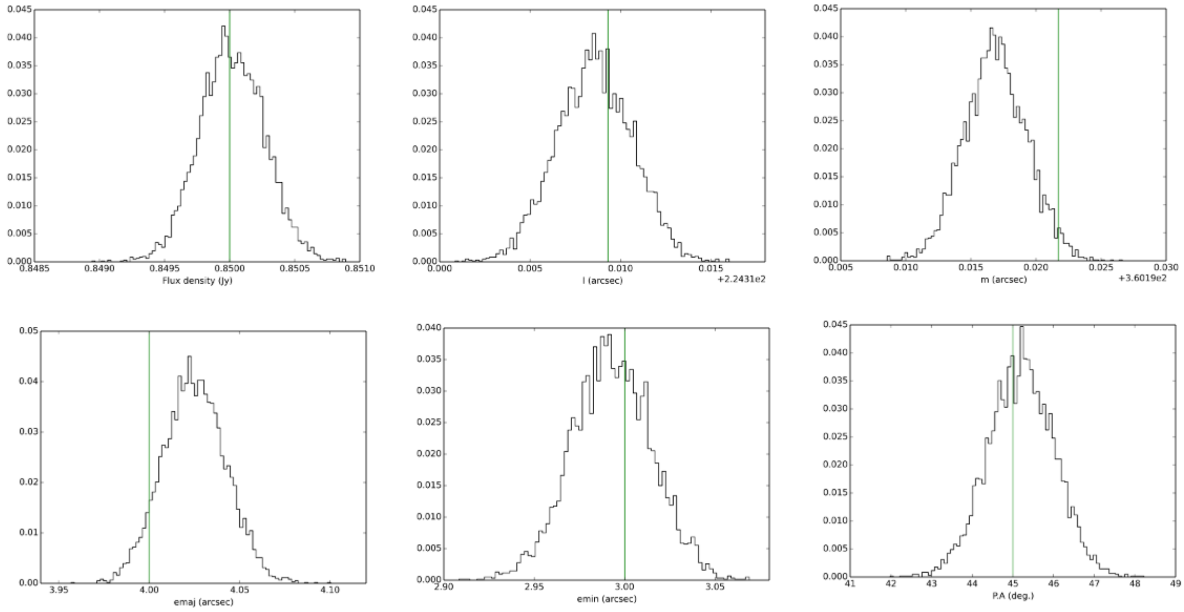


Figure 3.8: Posteriors for the parameters of the model GAU. The vertical green lines indicate the `true' values that went into the simulation.

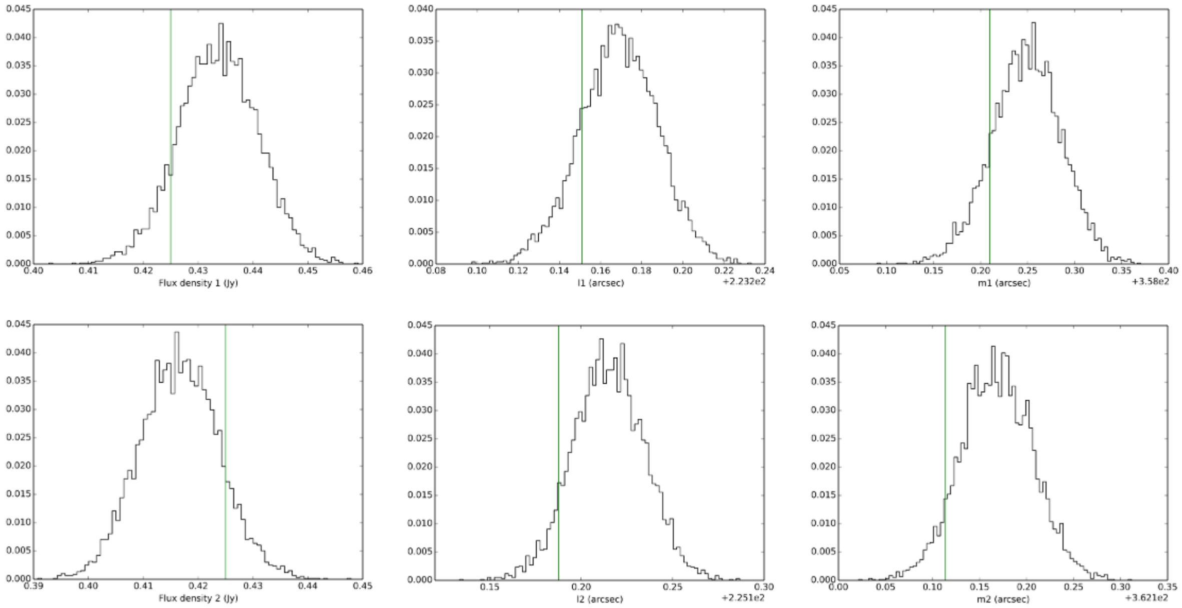


Figure 3.9: Posteriors for the parameters of the model 2PT. The vertical green lines indicate the `true' values that went into the simulation.

than those of the likelihoods. Hence, the MAP parameter estimates are not very different from the ML estimates and both lie well within the uncertainties on the means of the posteriors.

We reconstructed the skies from the MAP estimates corresponding to the parameters of the model with the highest evidence for each simulation (Figure 3.10). Comparing this with Figure 3.4, we see

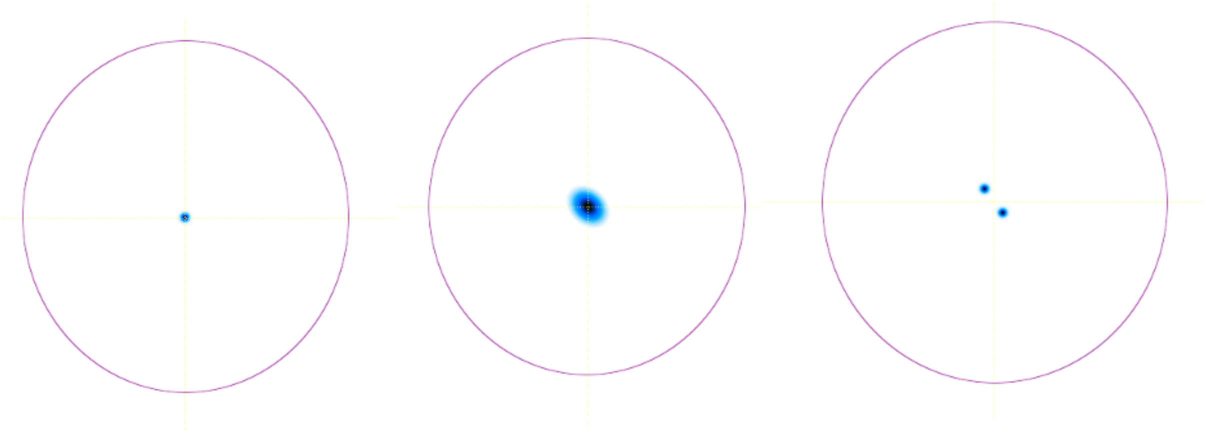


Figure 3.10: Reconstructed skies from the *maximum a posteriori* estimates from the BIRO analysis for each sky model. The red ellipses denote the FWHM of the PSF of size 14" × 16". Compare to Figure 3.4.

that the reconstruction matches the real sky closely in each case. A video showing how the reconstructed sky converges to a Gaussian over time can be found at `https://vimeo.com/117391380`.

3.3.2 Quantifying the Limits Imposed by Noise

In the previous simulations, we used a per-visibility noise of 100 mJy for sources with a peak flux-density of 0.85 Jy. The resulting datasets had an SNR of about 1000:1 and BIRO had no difficulty in estimating the source structure and discriminating between models. The theoretical limit for super-resolution with an interferometer is affected by noise and worsens with decreasing SNR (Martí-Vidal et al., 2012). To understand the effects of this noise on BIRO's capability to perform model selection, we simulated multiple datasets of varying SNRs, by increasing the per-visibility noise from 0.1 Jy to 1.0 Jy and keeping the source flux-density constant.

3.3.2.1 Details of the Simulations

The integrated flux-density of the source was kept constant at 0.85 Jy while the size of the simulated Gaussian source was increased from 0.8 arcsec to 4.0 arcsec in steps of 0.4 arcsec. A series of such simulations was carried out for five SNR levels, with the per-visibility noise set at 0.1, 0.2, 0.3, 0.5, and 1.0 Jy. A single frequency channel of width 125 kHz centred at 1.4 GHz was used. For a 12-hour simulation, the integration time used was 30 s (Figure 3.3).

3.3.2.2 Results

The RIME constructed for these analyses was identical to the one used in section 3.3.1.3. The prior ranges and the likelihood are the same as the ones given in section 3.3.1.4. We computed the relative logarithmic evidences between PT and GAU (with the same parametrisation as shown in Table 3.1). The results are shown in Figure 3.11.

The various coloured bands denote regions of varying weights of evidence and each curve in the figure corresponds to an SNR level, with multiple simulations of sources of varying sizes for each SNR. The band labelled *positive* is the point where the $2\ln(B_{ij})$ factor becomes > 2 (Table 2.1) and can reasonably be relied upon. The PSF of this interferometer at 1.4 GHz (the simulation frequency) is about 16 arcsec. For instance, when a per-visibility noise of 0.1 Jy was introduced for a simulated source of integrated flux-density 0.85 Jy at the centre, BIRO can reliably pick out a source of size 0.8 arcsec. For extended sources that are smaller than this, BIRO fails and starts preferring the single point source model PT. For a source of size 1.2 arcsec, the evidence in favour of GAU is *very strong*. This limit worsens with increasing noise as one would expect, until, when the per-visibility noise is 1.0 Jy for a source of flux-density 0.85 Jy, BIRO can pick out sources only of FWHM 3 arcsec or larger. We also see that for some cases, this trend is not as smooth as we would expect, such as the relative logarithmic evidence for a Gaussian of size 1.4 arcsec (for 0.1 Jy / vis) or a Gaussian of size 2 arcsec (for 0.2 Jy / vis). These discrepancies are due to the software not taking into account the phenomenon of peak-flux-to-noise ratio deteriorating with increasingly resolved sources. Hence the actual SNR is different from the quoted SNR in these cases. When this correction is introduced (for instance, in the experiments mentioned in section 4.6), we see that there is a monotonous increase in the sizes of the

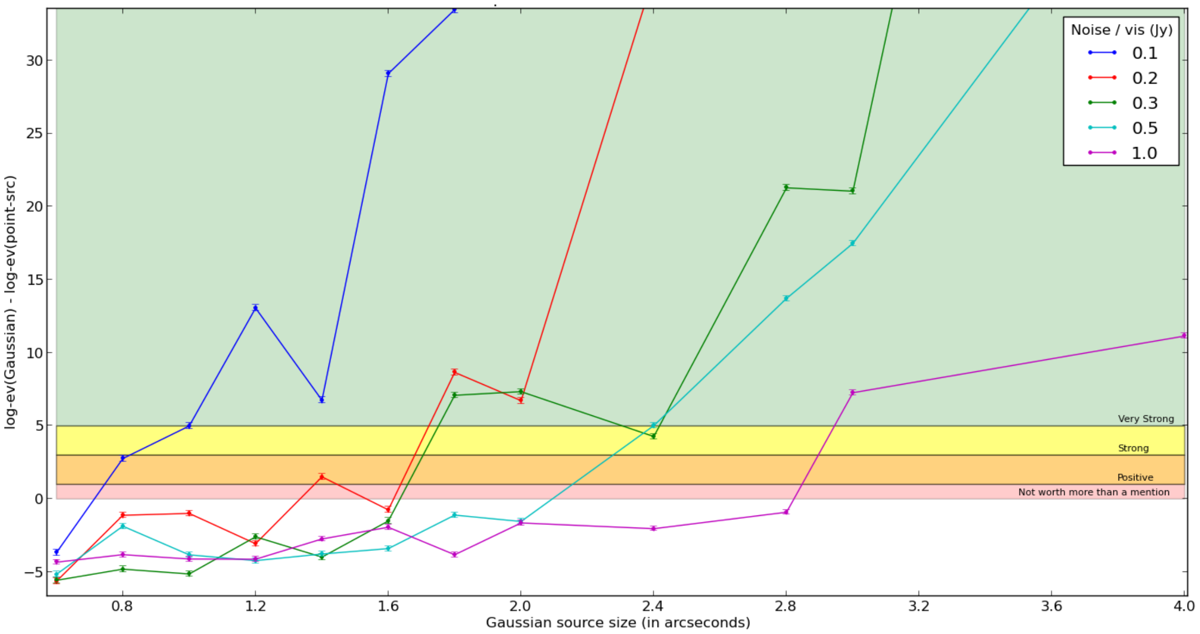


Figure 3.11: The variation in relative logarithmic evidence (y axis) with simulated source size (x axis) for multiple SNR levels. The coloured bands denote regions of varying weights of evidence according to the scale proposed by Kass & Raftery (1995).

sources that can be resolved with worsening SNR. Hence, this provides a practical way of measuring the limit to super-resolution using BIRO in the presence of random noise for a given interferometer configuration. The RIME has the potential to also study these limits in the presence of more complex instrumental effects (section 4.6).

3.4 CONCLUSIONS

We have seen that Bayesian inference can be successfully applied to the analysis of interferometric data in the visibility domain. Applying this to the problem of super-resolution in the presence of simulated random noise, we were able to discriminate between source models with high accuracy. We have also extended this to compute a practical limit for resolving compact sources by simulating sources of varying sizes with different noise levels. But the RIME is capable of modelling any propagation path effect that acts on the radio wave, from the source to the telescope. This gives us a way of incorporating more complicated instrumental effects while analysing compact source structure (see Chapter 4).

Modelling more complex effects introduces more degrees of freedom (i.e. free parameters) and hence the computational cost of the Bayesian analysis becomes a significant issue. Since a fully Bayesian estimation requires the computation of the full posteriors of all model parameters that more often than not are analytically intractable, numerical sampling techniques such as MCMC and nested sampling (section 2.4) are required to approximate the posterior. These techniques require a long time to converge on the correct solution as the dimensionality of the parameter space increases. The simple models evaluated in this chapter require 10^4 to 10^5 likelihood evaluations, with each likelihood evaluation lasting about 0.2 seconds and the entire run lasting hours on an 8-core system. At each iteration, MeqTrees computes the RIME using the sampled parameters and also the χ^2 for the likelihood. MeqTrees can use multiple CPUs to execute but it would still take hours to evaluate these simple models. To evaluate more complex models, it is necessary to speed up the computations so that the analysis can be performed in a reasonable amount of time. How this was achieved and how the improved software was instrumental in studying the structure of a compact astronomical source in the presence of instrumental effects is the subject of the next chapter.

4

Resolving the Blazar CGRaBS J0809+5341

Any multi-element interferometer is a[*] collection of two-element interferometers, the data from which are recorded independently and combined only during imaging. Typically, all the stations that take part in an observation are connected to each other and operated by the same observatory. A radio receiver *down-converts* a radio frequency (RF) signal to a lower frequency so that signal manipulation such as amplification or filtering becomes easier and the signal losses during transport are minimised (e.g. Wilson et al., 2009). This process is called *heterodyning* and is performed by mixing a monochromatic signal generated locally by a *local oscillator* (LO) with the RF signal. The LO signal frequency is tuned to be as close as necessary to the RF, so that their difference produces a low frequency signal that is more easily managed. Heterodyning is performed using a single LO whose output is connected to the individual stations. Any variation or instability in the frequency generated by the LO affects all the stations equally.

By the mid-1960s, astronomers had realised that there are many radio sources that are unresolved at angular scales of about 0.1 arcsec, which is typical of the resolution afforded by interferometers with baselines that extend up to hundreds of kilometres. To obtain very high-resolution measurements, the technique of *Very Long Baseline Interferometry* (VLBI) was introduced.

VLBI makes use of independent stations located far and wide, potentially spanning the diameter of the Earth (Thompson et al., 2017, Chapter 1). The various stations which participate in a VLBI observation belong to different observatories and are operated independently. Extremely phase-stable local oscillators avoid the need for sharing a single LO between the stations. The data are recorded on disks or tapes with precise time marks with the help of atomic clocks and are correlated later (Wilson

---

[*] This work has been published under the title *Resolving the blazar CGRaBS J0809+5341 in the presence of telescope systematics* (Natarajan et al., 2017).

et al., 2009). The more recent *e*-VLBI technique facilitates near real-time transport of the data to the correlator.

VLBI is used to observe a variety of astronomical sources such as active galactic nuclei (AGN) and their immediate structure, certain binary stars, radio stars, and young supernova remnants (Middelberg & Bach, 2008). They are also used for accurate position measurements of celestial sources (*astrometry*). VLBI observations regularly achieve sub-milli-arcsecond (sub-mas) resolutions and the super-resolution techniques discussed in Chapter 3 can only improve on this. Moreover, owing to the small FoVs processed in many VLBI observations, a typical region of interest in VLBI is devoid of other strong sources, apart from the target source. This enables one to formulate models with a manageable number of parameters for the Bayesian analysis. In this chapter we discuss the results of applying Bayesian inference to resolve the structure and measure the brightness temperature of a distant AGN observed using the European VLBI Network (EVN) under conditions of poor calibration and sampling, and derive a Bayesian criterion for the resolution limit in the presence of telescope systematics.

4.1 Active Galactic Nuclei

Active galaxies are galaxies whose energy output consists of a significant fraction of non-thermal emission, mainly synchrotron and inverse-Compton radiation[*]. If this emission is observed to emanate mainly from the core, they are called active galactic nuclei (AGN). AGNs represent only a small fraction of all galaxies, but possess high luminosities and have been studied in radio wavelengths since the inception of radio astronomy (Baade & Minkowski, 1954). The high luminosities of AGNs are believed to be the result of matter accreting on to central supermassive black holes (about 10^6 to 10^{10} the mass $M_\odot$ of the Sun) in these galaxies (Antonucci & Miller, 1985). Broad optical emission lines are observed from high-velocity gas and narrow lines from low-velocity gas near the nucleus, giving rise to the broad-line region (BLR) and narrow-line region (NLR) respectively. The central region is obscured by gas and dust, usually referred to as the *torus*, though there is no consensus on the geometrical form that the "torus" takes. Relativistic outflows at the poles called *jets* can be seen out to several tens of kiloparsecs or even megaparsecs (Urry & Padovani, 1995).

4.1.1 Classification

AGNs may be radio-loud or radio-quiet depending on how much they radiate at radio wavelengths. Radio-loud AGNs are some of the brightest radio sources in the sky while radio-quiet AGNs radiate mainly in the optical and X-ray wavelengths. Based on the nature of the emission spectra observed, AGNs are classified into Type 0, Type 1, and Type 2 sources (Table 4.1).

Seyfert galaxies are divided into Seyfert 1 and Seyfert 2, which exhibit broad and narrow line emission in the optical and X-ray wavelengths respectively. They occasionally show weak radio emis-

---

[*] The process in which fast-moving electrons impart energy to low energy photons (Rybicki & Lightman, 1985).

Table 4.1: Classification and properties of active galactic nuclei (Middelberg & Bach, 2008).

	Emission line properties		
Radio loudness	Type 2 (Narrow lines)	Type 1 (Broad lines)	Type 0 ('no' lines)
Radio quiet	Seyfert 2 LINERs	Seyfert 1 Quasars or QSOs	
Radio loud	NLRG	Quasars	Blazars

sion but are not sources of high radio luminosity. Similarly, the *low-ionization nuclear emission-line regions* (LINERs) are also radio-quiet. Among the narrow-line radio galaxies (NLRG) fall the two morphological types originally proposed by Fanaroff & Riley (1974), Fanaroff-Riley I (FR I) and Fanaroff-Riley II (FR II) galaxies: FR I galaxies have symmetrical jets that originate from the nucleus and broaden and fade away as they get farther from the nucleus; FR II galaxies show collimated jets that end in well-defined hot spots or lobes. *Quasars* are highly luminous high-redshift galaxies that can be either radio-quiet (sometimes called Quasi-Stellar Objects or QSOs) or radio-loud. The radio-loud quasars display high radio luminosities (10^{46} erg s^{-1})[*] and are some of the most energetic objects in the sky. When the jets from a quasar are aligned very close to the line-of-sight to the observer, the AGN is called a *blazar* (e.g., Antonucci, 1993; Urry & Padovani, 1995). Blazars include two types of objects: (i) *BL Lacertae* (BL Lac) objects that lack strong emission or absorption features and are highly variable and highly polarised, and (ii) *flat-spectrum radio quasars* (FSRQ), which show continuum emission properties very similar to those of the BL Lac objects[†]. Depending on the viewing angle, an AGN may be classified into a quasar, blazar, or radio-loud galaxy. The unification scheme for AGN is illustrated in Figure 4.1.

4.1.2 BRIGHTNESS TEMPERATURES AND JET SPEEDS

Equation (1.15) gives the brightness temperature T_b of a source in the limit of the Rayleigh-Jeans approximation to Planck's law. For non-thermal emission, as is the case from blazars, T_b is the equivalent temperature of a blackbody that radiates the same amount of power at the same wavelength λ. Expressed in terms of frequency rather than wavelength, equation (1.15) becomes

$$T_b = \frac{c^2}{2k_B\nu^2}\frac{S_\nu}{\Delta\Omega}, \tag{4.1}$$

where ν is the frequency of observation, S_ν is the flux-density, k_B is the Boltzmann constant, and $\Delta\Omega$ is the solid area subtended by the source.

---

[*] 1 erg = 10^{-7} Joules.

[†] Several FSRQs and some BL Lac objects have been observed to exhibit emission lines as well (Middelberg & Bach, 2008).

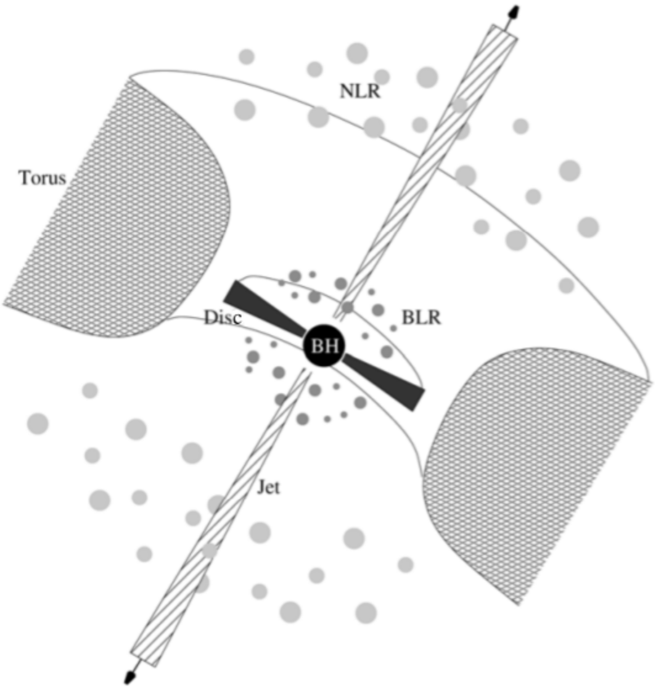


Figure 4.1: Unification scheme for AGN. The appearance of an AGN depends on the angle between the line-of-sight to the observer and the direction of the polar jets illustrated. The broad-line region (BLR) containing high-velocity gas and the narrow-line region (NLR) composed of low-velocity gas might be obscured depending on the viewing angle, resulting in the AGN being classified as one of the various types mentioned in the text (Middelberg & Bach, 2008).

Brightness temperatures as high as $10^{10} - 10^{12}$ K are obtained from VLBI observations of compact sources such as blazars, which can only be accounted for by non-thermal emission mechanisms such as synchrotron emission in which relativistic electrons spiral through magnetic fields and radiate. When $T_b > 10^{12}$ K, the energy losses due to inverse Compton scattering dominate and T_b decreases to between 10^{11} and 10^{12} K, where the inverse Compton and synchrotron losses are of the same order (Kellerman & Pauliny-Toth, 1969). Early VLBI observations showed that T_b measurements for almost all sources fell in the range of $10^{11} - 10^{12}$ but this was later found to be a consequence of Earth-based measurements[*] (Kellerman, K. I., 2003). More recent observations have shown that the brightness temperatures can be as high as 10^{13} K or even more (e.g. Zensus et al., 2002; Tingay et al., 2001; Horiuchi et al., 2004; Kovalev et al., 2016). However, Bruni, G. et al. (2017) found that such high intrinsic brightness temperatures are achieved and maintained only for short periods of time in AGN jets and have pointed out that the availability of μas resolution does not guarantee the measurement of brightness temperatures above the known limits.

Rees (1966) originally proposed that bright compact objects at cosmological distances are composed of parts moving at relativistic speeds. This bulk relativistic motion is the most probable model that explains the high brightness temperatures and apparent superluminal motion of jets observed in

---

[*] The maximum possible baseline length is limited by the diameter of the Earth and the range of flux densities observed is limited due to the presence of atmosphere, among other factors, when observed from Earth.

radio galaxies (e.g., Marscher & Scott, 1980). This enhanced T_b is a consequence of *Doppler boosting* in which the relativistic beaming of a jet moving at a small angle to the line-of-sight to the observer increases the observed flux-density without increasing the size of the jet (Kellerman & Owen, 1988). The measured brightness temperature and the internal brightness temperature $T_{b,int}$ relate as

$$T_b = \delta \, T_{b,int} \tag{4.2}$$

where δ is the *Doppler boosting factor*.

4.2 EVN Observation of CGRaBS J0809+5341

CGRaBS J0809+5341 is an FSRQ, located at J $08^{\mathrm{h}}09^{\mathrm{m}}41.733^{\mathrm{s}}$, $+53^{\mathrm{d}}41^{\mathrm{m}}25.092^{\mathrm{s}}$ at redshift $z = 2.144$ (Pâris et al., 2014), originally observed by the Candidate Gamma-Ray Blazar Survey (Healey et al., 2008). It was observed to emit a bright optical flare of magnitude $M = -30.0$ on April 12, 2014 (Balanutsa et al., 2014). High energy γ-ray outbursts have also been continuously detected as reported by Paliya (2015). These outbursts are thought to be associated with a relativistic jet and accompanied by a radio outburst as well (e.g. Jorstad et al., 2001). To study the changes in the structure of the blazar, An et al. (2016) conducted VLBI observations of the blazar on November 18, 2014 with the EVN at a frequency of 5 GHz and followed it up with 22 GHz and 43 GHz observations using the KaVA, a combination of the Korean VLBI Network (KVN) and the Japanese VLBI Exploration of Radio Astrometry (VERA). In our analysis, we use the EVN observation carried out by An et al. (2016) at 5 GHz.

The EVN observation was performed using 8 EVN stations (Table 4.2) with one of them, Sheshan (SH), located in Shanghai, China. Seven antennas of WSRT were used in *phased array* mode[*]. The blazar was observed for 130 minutes with an integration time of 2 s, using 8 × 16 MHz spectral bands of 32 channels each, located between 4926 and 5054 MHz. Both R and L polarisations were recorded. The *uv*-coverage of the synthesis is shown in Figure 4.2. The long baselines correspond to the SH station and extend over 9000 km. Without SH, the maximum baseline length is about 2200 km.

Before any analysis can be performed, invalid or corrupted data must be identified and excluded (*flagged*). Data may be corrupted due to a variety of reasons such as radio frequency interference (RFI) from various natural and artificial sources. Many RFI mitigation techniques have been proposed and implemented (e.g. Offringa et al., 2012, and references therein). After flagging, the visibilities were fringe-fitted (section 5.1) and self-calibrated[†] (section 1.3.3). For this procedure, the channels in each band were averaged together so that the data contained 8 bands of one channel each. The frequency-averaging is made possible by the fact that the source is located at the pointing centre and is therefore

---

[*] Unlike in the correlation mode, in a phased array, the antenna beams are steered electronically to point in the desired direction and the signals are added in phase.

[†] www.evlbi.org/user_guide/evn_datareduc.html. There are no primary calibrators in a VLBI observation since sources compact enough at mas scales vary a lot in the measured flux-density.

Table 4.2: EVN stations used in the observation along with the corresponding dish diameters and the nominal (SEFD) values. WB was operated as a phased array with 7 antennas.

Station	Code	Diameter (m)	Nominal SEFD (Jy)
Effelsberg	EF	100	20
Jodrell Bank	JB	25	320
Noto	NT	32	260
Onsala	ON	25	600
Torun	TR	32	220
OAN-Yebes	YS	40	160
Westerbork	WB	7×25	120
Sheshan	SH	25	720

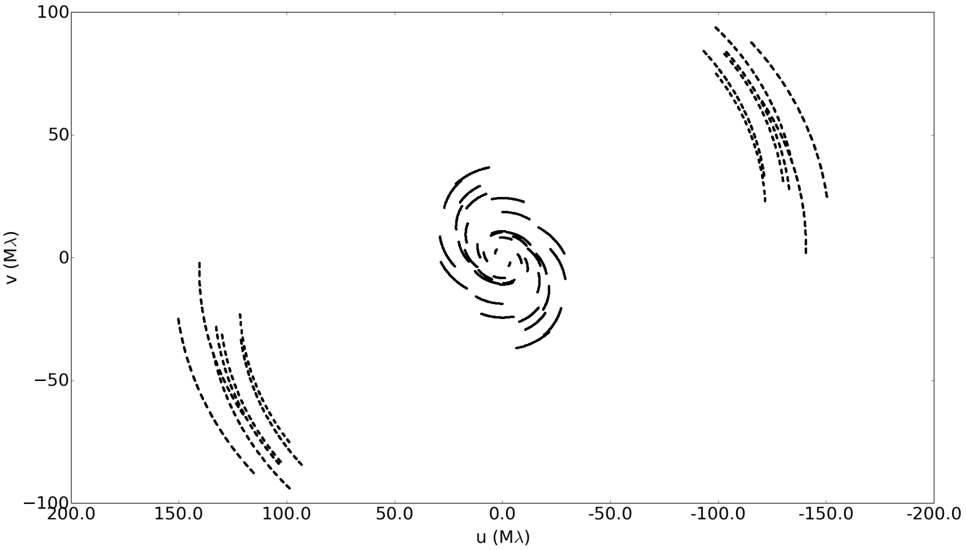


Figure 4.2: The *uv*-coverage of the EVN observation. The longest baselines correspond to the Sheshan (SH) station.

not affected by frequency smearing (Smirnov, 2011a). I extracted the spectral band centred at 4982.24 MHz for the analysis and the original UVFITS[*] file was converted into the Measurement Set (MS) format[†] and the missing[‡] baselines were introduced — with the corresponding data flagged — for compatibility with the software used (section 4.3).

4.2.1 Conventional Analysis

A *naturally*-weighted (Briggs, 1995), deconvolved image of the self-calibrated data shows a compact source (Figure 4.3). The deconvolution was performed using the CSCLEAN algorithm (Schwab, 1984) in *lwimager*[§] with 1000 iterations. The PSF of the interferometer was calculated using the Hogbom algorithm recommended for data with poor *uv*-coverage (Högbom, 1974). PYBDSM estimates that the source is unresolved with a flux-density of 161.1 ± 0.01 mJy.

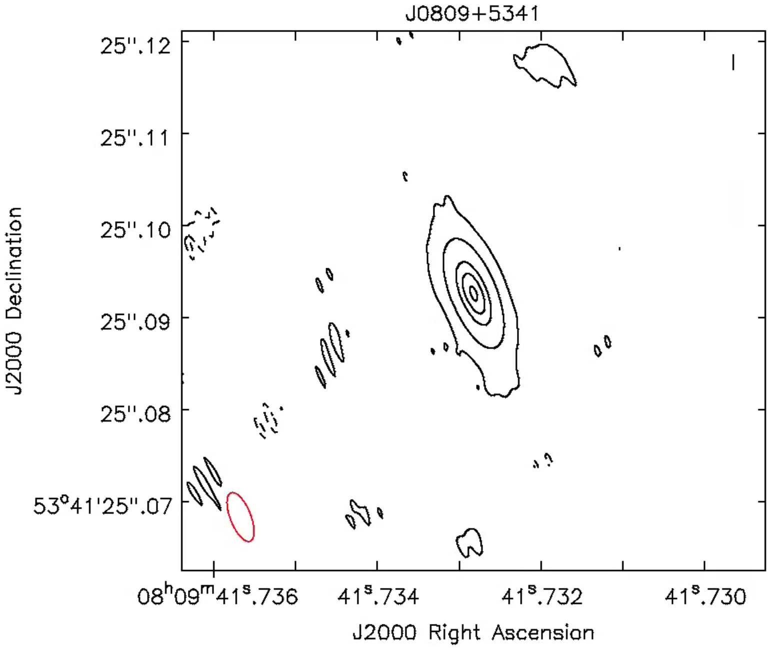


Figure 4.3: Stokes *I* image of CGRaBS J0809+5341. The contours presented are −3, 3, 100, 1000, 2000, and 3000 times the rms noise in the image ($\simeq$ 0.05 mJy), with the negative contours shown as dashed lines. The red ellipse at the bottom left corner is the PSF used for restoration by CLEAN: 5.7 × 2.2 mas, oriented at an angle of 21.7°.

An et al. (2016) fitted a circular Gaussian source model to the self-calibrated visibilities and found the best-fit size for the source to be 0.31 ± 0.06 mas, consistent with the previously-known core of the blazar. Although they measured a brightness temperature of $(0.25 \pm 0.6) \times 10^{12}$ K, expected in the presence of a relativistically beamed jet of particles from the core, they could not resolve the blazar any further than the known core, or image the jet. This measurement is further complicated by the fact

---

[*] FITS format file that records the *uv*-domain data in ASCII or binary tables.

[†] `https://casa.nrao.edu/Memos/229.html`

[‡] Due to the source not being visible to those baselines.

[§] `https://github.com/casacore/casarest` that the SH station does not have any other station close by to provide short baselines. As a result, its gain calibration is much less precise and accurate than those of the other stations. Hence, in the absence of short baselines, the low amplitudes observed at the longest baselines pertaining to SH may be the result of either the target being truly resolved or the gain calibration being wrong (Martí-Vidal et al., 2012).

Although the results of the analysis cannot resolve the jet, there is reason to believe that the source is resolved. Figure 4.4[*] shows the SH station gain amplitude and phase corrections with a 5 minute solution interval[†] for the self-calibrated data and for a single point source model. The European stations do not show any difference between the two plots. This shows that though the SH measurements provide a strong reason to suspect that the jet is resolved, we cannot be sure how much of the difference in the amplitudes is due to incorrect gain calibration.

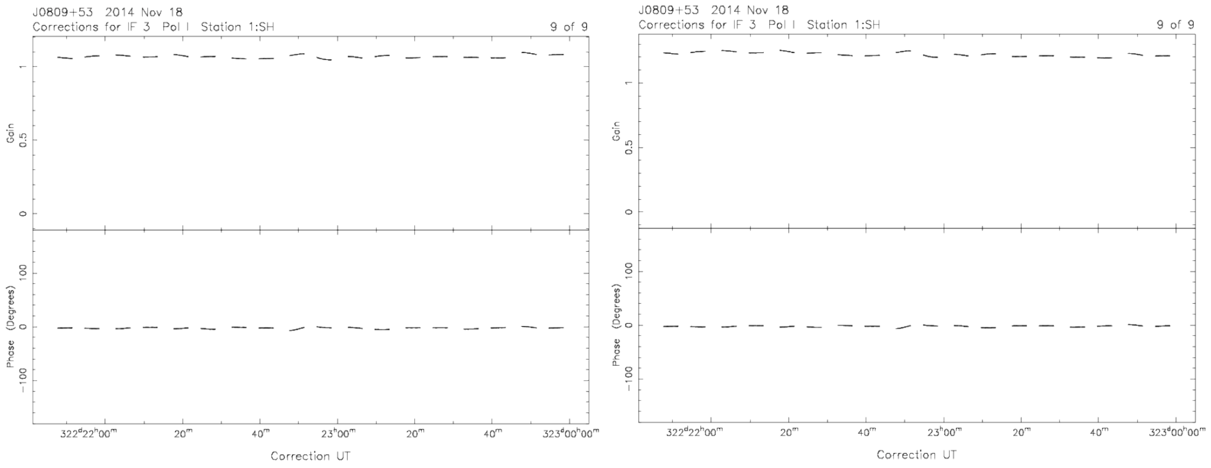


Figure 4.4: The figure on the left shows the gain amplitude (at unity) and phase corrections for the SH station after self-calibration. The figure on the right shows the same for an assumed point source model at the centre, i.e. what the gain solutions would be if the source were considered unresolved; the gain amplitudes here are noticeably higher than 1.

To estimate the uncertainties in the gains and study the correlation between the source shape and the station gains and the effect on the estimated brightness temperature, we performed Bayesian model selection on the self-calibrated visibilities. The models used here are more complex than those used for the WSRT simulations in Chapter 3, with parameters representing the individual station gains and SEFDs. The MeqTrees + MultiNest setup used there would take days to estimate the parameters for a single model. We therefore need a faster way to compute the likelihood function.

4.3 Montblanc: An Interlude

Graphics Processing Units (GPUs) consist of thousands of processing units (or *cores*) that are architected in such a way that they all execute a common program on separate data. Under Flynn's tax-

---

[*] The plots were provided by Zsolt Paragi, who was part of the team that made the observation.

[†] An independent solvable gain term per antenna per 5 minutes.

onomy of computer architectures, they fall under the category of Single Instruction, Multiple Data (SIMD) architecture (Flynn, 1972). The Compute Unified Device Architecture (CUDA), written in a variant of C, enables one to program GPUs to handle a wide variety of applications apart from graphics processing (NVIDIA Corporation, 2016). Many radio astronomy algorithms are easily parallelised and hence can benefit hugely from a GPU implementation (Barsdell et al., 2010).

MONTBLANC[*] is a GPU implementation of the RIME. Unlike OSKAR (Mort et al., 2010), another GPU-based software that implements parts of the RIME on a GPU, MONTBLANC aims to implement the entire RIME on a GPU. Additionally, it computes the chi-squared to be used in evaluating a Gaussian likelihood function such as equation (3.4). MONTBLANC is written in PYTHON, and uses the NUMPY[†] package for scientific computing and PYCUDA[‡] (Klöckner et al., 2012) for accessing the CUDA architecture. It is publicly available for download at `https://github.com/ska-sa/montblanc`.

4.3.1 BAYESIAN ANALYSIS

Bayesian analysis proceeds by repeatedly sampling from the parameter space and evaluating the likelihood function at every iteration. To speed this up, both the model visibilities using the RIME and the χ^2 for computing the likelihood are calculated on the GPU (Perkins et al., 2015).

Figure 4.5 shows the algorithm flow for one iteration. The CPU-based numerical sampler samples a set of parameter values from the parameter space of a RIME model. These parameters could pertain to either the sky or the telescope, as indicated by the green boxes in the figure. The model visibilities are computed using the RIME and are used in combination with the data (observed visibilities) to compute the χ^2 on the GPU. This value is returned to the CPU in which the likelihood function computation is completed and used by the sampler to decide on the next set of parameter values to sample. This process is repeated until a stopping criterion is reached.

4.3.2 TESTING AND PERFORMANCE

We tested MONTBLANC on a Tesla K40M GPU device. The K40M consists of 2880 CUDA processor cores and 12 GB of memory. The software was debugged and various useful features were added during the unit-testing. As of January 2017, MONTBLANC consists of versions v2 through v5, with substantial changes made to the architecture and the Application Programming Interface (API) in versions v4 and v5. One particular issue that took some time to resolve is worth mentioning here.

MONTBLANC provides a visibility-weighting mechanism that enables the user to weight the vis-

---

[*] The work described in this section was published as *Montblanc: GPU accelerated radio interferometer measurement equations in support of Bayesian inference for radio observations* (Perkins et al., 2015). I was a co-author, involved in unit-testing the software.

[†] `www.numpy.org`

[‡] `https://mathema.tician.de/software/pycuda`

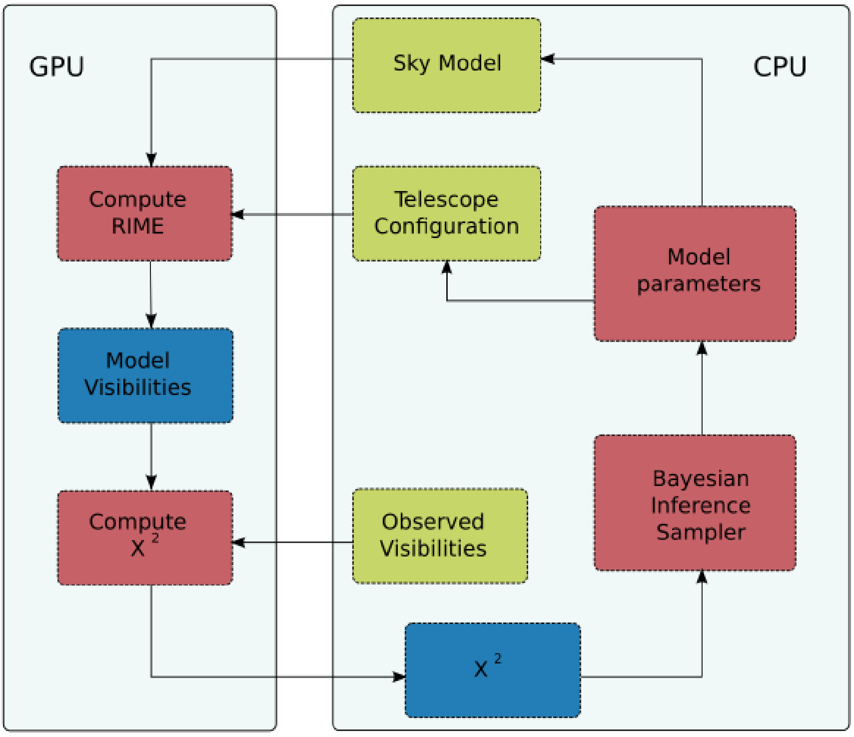


Figure 4.5: Algorithm flow for Bayesian analysis using Montblanc. Red boxes indicate computation, green boxes, input, and blue boxes, output (Perkins et al., 2015).

ibilities according to their own weighting scheme. This facility is provided by a `weight_vector`[*] which is combined with the vector of visibilities before the likelihood is computed. This vector was originally merged with a `flag_vector` that was intended to keep track of the flagged visibilities. In `flag_vector`, *De Morgan's laws* were applied to combine the flags read from the two columns FLAG and FLAG_ROW in the MS. Named after the British mathematician Augustus De Morgan (1806-1871), the two laws may be stated as, given two propositions X and Y, the following two properties hold:

$$\begin{aligned} \neg(X \wedge Y) &\iff (\neg X) \vee (\neg Y) \\ \neg(X \vee Y) &\iff (\neg X) \wedge (\neg Y) \end{aligned} \tag{4.3}$$

where $\neg$ denotes *negation* of a proposition and $\wedge$ and $\vee$ denote the operations of *conjunction* and *disjunction* respectively. Both the merging of the vectors and the application of De Morgan's laws were implemented incorrectly in the original version of Montblanc. I performed various simulations with and without the introduction of flagged baselines before I homed in on both the issues. This bug was corrected in a subsequent update[†] that separates the flag and the weight vectors, allowing the visibilities to be flagged independent of the weights, and also corrects the implementation of De

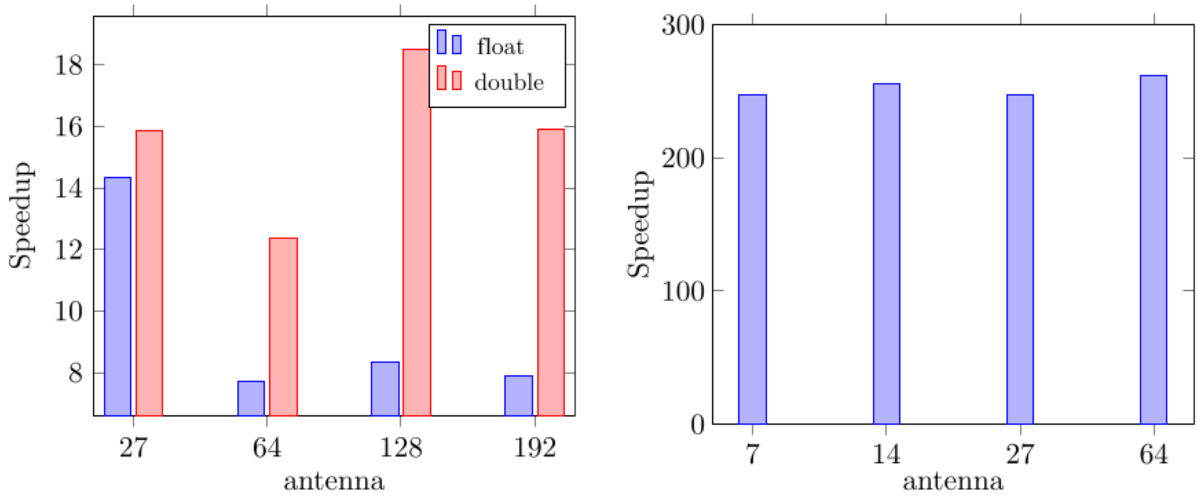
---

[*] A *vector* in computer science is closely related to the concept of an *array*, both of which may be multi-dimensional and have a certain amount of storage allocated to them. For more details about arrays and vectors, any book on programming can be referred to (e.g. Stroustrup, 2013).

[†] `https://github.com/ska-sa/montblanc/commit/d26310cdd157b47f9a51edf0164ad749f47ee382`

Morgan's laws. With this amendment, the software was ready for application to the EVN data.

MONTBLANC is written in such a way that the different Jones matrices involved (section 1.3.2) are computed in parallel and differ in computational complexity according to whether they can be separated out into antenna-based terms or not (Perkins et al., 2015). For instance, the E and the K Jones terms can be expressed as antenna-based matrices and hence their computational complexity is $O(\texttt{ntime} \times \texttt{nant} \times \texttt{nsrc} \times \texttt{nchan})$, where the four variables stand for numbers of timestamps, antennas, sources, and frequency channels respectively. In contrast, the computational complexity for B is baseline-dependent ($O(\texttt{ntime} \times \texttt{nbl} \times \texttt{nsrc} \times \texttt{nchan})$, where `nbl` is the number of baselines) and starts dominating the execution time as the number of baselines increases. A comparison of the execution speeds of MONTBLANC, OSKAR, and MEQTREES is given in Figure 4.6. MEQTREES executed on a dual hexacore E5-2620v2 system while OSKAR executed on the Tesla K40. In all cases, 64 frequency channels, 100 timesteps, 50 point and 50 Gaussian sources were used in the synthetic data on which the three software operated.



(a) OSKAR. (b) MEQTREES.

Figure 4.6: The factors of speed-up achieved by Montblanc compared to (a) oskar and (b) MeqTrees (Perkins et al., 2015).

Currently, MONTBLANC supports the evaluation of three different source morphologies — point, Gaussian, and Sérsic[*] profiles to describe brightness distributions. For our analyses, we need only the point and the Gaussian brightness distributions.

4.4 ANALYSIS

4.4.1 DESCRIPTION OF MODELS

I compare three morphological models to describe unresolved or partially-resolved sources, each differing in the assumed brightness distribution of the source:

---

[*] https://github.com/ska-sa/montblanc/commit/bafa4b6 by Marzia Rivi (2015).

PT: The point source model consists of one flux-density parameter, S_ν, and two parameters that describe the position of the source, (l, m), the direction cosines measured with respect to the (u, v) co-ordinates.

GAU: The elliptical Gaussian model consists of S_ν and (l, m) along with three more parameters, l_p, m_p, and r, describing the shape of the source. These three parameters are related to the major axis $e_{\rm maj}$, minor axis $e_{\rm min}$, and position angle or orientation ϑ of the ellipse in the same way as in equation (3.7) which is reproduced here:

$$\begin{aligned} l_p &= e_{\rm maj}\,\sin\vartheta \\ m_p &= e_{\rm maj}\,\cos\vartheta \\ r &= e_{\rm min}/e_{\rm maj}. \end{aligned} \tag{4.4}$$

Figure 3.6 shows how these three parameters describe a Gaussian brightness distribution.

CIRC: The circular Gaussian model consists of the same six parameters as GAU, with the exception that the ratio r is given a delta distribution at unity, forcing the shape to be circular.

Table 4.3: Models evaluated in this study. Besides the source parameters, there are 7 free parameters describing the station gain amplitudes and 8 parameters describing the individual station SEFDs.

Model	Number of parameters	Parametrisation
PT	18	Flux Density (S_ν) Position (l, m) Station gain amplitudes ($\lvert g_p\rvert$) SEFD$_p$
GAU	21	Flux Density (S_ν) Position (l, m) Shape (l_p, m_p, r) Station gain amplitudes ($\lvert g_p\rvert$) SEFD$_p$
CIRC	20	Flux Density (S_ν) Position (l, m) Shape (l_p, m_p) Station gain amplitudes ($\lvert g_p\rvert$) SEFD$_p$

A summary of the parametrisation of each model is given in Table 4.3. Alongside the parameters that describe the source, all three models incorporate the same instrumental parameters. In principle, any instrumental effect could be modelled using the RIME (see section 1.3.2), while in practice, we are often limited by the software used to sample the parameter space. The MultiNest algorithm performs best with low-dimensional ($\lesssim 30$) parameter spaces (Feroz & Hobson, 2008).

For the present experiment, the most important parameters that affect the source shape estimation are the amplitudes of the complex station gains and the per visibility noise term σ_{pq} that varies with baseline. The complex station gains vary with respect to time during the course of the observation and are solved for during the preliminary self-calibration (section 4.2.1). The resulting gain amplitudes and phases are sufficiently smooth in time to be modelled using one complex gain term per station. Moreover, since all the source models considered place the source at the pointing centre, the measured phases on all the baselines will be zero. Hence, I hold the gain phases constant at zero, bringing the number of gain-related parameters to one amplitude term per station.

Accounting for the per-visibility noise term σ_{pq} for each baseline pq would mean that there are as many noise parameters as there are baselines. But random (Gaussian-distributed) noise can be expressed in terms of the system temperature T_{sys} of the receiver system of each individual station, or rather, in terms of the SEFD of each station (section 1.4). VLBI observations are often conducted using heterogeneous arrays[*] composed of individual stations that are of different sizes and possibly, shapes, and are not connected to the same electronics. This results in the individual station SEFDs varying widely in any VLBI observation. The per-visibility uncertainty for one polarisation in terms of the geometric mean of the station SEFDs is given by the *radiometer equation* (Thompson et al., 2017):

$$\begin{aligned} \sigma_{pq} &= \frac{\mathrm{SEFD}_{pq}}{\sqrt{\delta\nu\, \tau_{pq}}}, \\ \text{where } \mathrm{SEFD}_{pq} &= \sqrt{\mathrm{SEFD}_p \,\mathrm{x}\, \mathrm{SEFD}_q}, \end{aligned} \tag{4.5}$$

Here SEFD_p is the SEFD of station p, $\delta\nu$ is the channel bandwidth, and τ_{pq} is the integration time for baseline pq. This is then used to weight the model visibilities using the reciprocal of the corresponding variance (natural weighting). Thus we bring the number of parameters that describe the noise from the *number of baselines* to the *number of stations*.

4.4.2 Building the RIME

Similar to section 3.3.1.3, I assume a flat spectral index and use the flux-density parameter S_ν to construct the brightness matrix:

$$\mathrm{B} = \begin{pmatrix} S_\nu & 0 \\ 0 & S_\nu \end{pmatrix}. \tag{4.6}$$

As with the previous case, the first linear transformation this signal undergoes is represented by the phase delay matrix K, associated with the difference in the geometric path lengths from the source to antennas p and q:

$$K_p = \mathrm{e}^{-i\kappa_p} \equiv \mathrm{e}^{-i\kappa_p} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.7}$$

---

[*] Some VLBI arrays such as the Very Long Baseline Array (VLBA) and the aforementioned KVN and VERA are homogeneous.

And the source coherency matrix can be written as

$$X_{pq} = K_p \, \mathrm{B} \, K_q^H = \mathrm{B}e^{-i\kappa_{pq}}. \tag{4.8}$$

Unlike the case with the models used for the WSRT simulations, here I include station gains and SEFDs as instrumental parameters. To a first order, we can assume equal gains for the two polarisation feeds, $g_p = g_{x_p} = g_{y_p}$, and reduce the diagonal *G-Jones* matrix describing station gains to a scalar matrix:

$$G_p = \begin{pmatrix} g_p & 0 \\ 0 & g_p \end{pmatrix} = g_p \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.9}$$

The PSF of the observation is 5.7 × 2.2 mas when the visibilities are weighted naturally, while the primary beams of the stations are at least 8 arcmin wide at 5 GHz. Hence, for a source located at the pointing centre, the primary beams do not affect the observation, and the *E-Jones* matrix representing the primary beam effects is set to unity, further simplifying the RIME.

Now, remembering the additive Gaussian noise term with zero mean and a variance of σ_{pq}^2 per visibility, the RIME for the point source becomes

$$\mathrm{V}_{pq} = G_p \, X_{pq} \, G_q^H + \mathcal{N}(0, \sigma_{pq}^2). \tag{4.10}$$

For the models GAU and CIRC, the summation becomes an integral performed over the extent of the source. Expressed in terms of the direction cosines l and m, the RIME becomes (Thompson et al., 2017, section 3.1)

$$\begin{aligned} \mathrm{V}_{pq} &= \iint_{lm} G_p \, X_{pq}(l, m) \, G_q^H \, \mathrm{d}\Omega \; + \; \mathcal{N}(0, \sigma_{pq}^2), \\ \text{where } \mathrm{d}\Omega &= \frac{\mathrm{d}l \, \mathrm{d}m}{\sqrt{1 - l^2 - m^2}}. \end{aligned} \tag{4.11}$$

The entire model is computed by Montblanc for all timestamps, baselines, sources, and frequency channels during every iteration of the likelihood computation.

4.4.3 PRIOR DISTRIBUTIONS

I set uniform priors $\Pi(\Theta)$ on most parameters (Table 4.4). Since the prior ranges do not span multiple orders of magnitude, uniform priors are sufficiently uninformative to not introduce unjustifiable biases. The prior range for S_ν is chosen based on what we know about the flux-density of J0809+5341 from preliminary imaging and source extraction (section 4.2). The prior distribution for l_p (the sine projection of e_{maj}) is restricted to non-negative values so that the position angle estimate is constrained to a range of 180°. This allows us to break the degeneracy between two Gaussians of the same size of which one is rotated 180° with respect to the other. I also allow the gain amplitudes to vary between ±20 per cent of unity and keep the corresponding phases fixed at zero. This is achieved by setting

Table 4.4: Prior distributions for the different parameters used. All the listed parameters were set uniform priors with the range indicated by the values in the square brackets. The parameters with delta priors are not included.

Parameter / units	Prior distribution
S_ν / Jy	[0.1, 0.2]
l&m / mas	[-4, 4]
l_p / mas	[0, +4]
m_p / mas	[-4, +4]
r	[0, 1]
$\lvert g_p \rvert$, where $p \neq$ EF	[0.8, 1.2]
SEFD / Jy	[5, 800]

delta distributions centred at the known values of the parameters. The data do not impact the posteriors of such parameters since the posteriors are also delta distributions at the same locations in the parameter space. To break the degeneracy between S_ν and $|g_p|$ estimates, the gain amplitude of the EF station is set a delta prior at unity. I choose EF because it provides some of the shortest baselines (except for the baseline with SH) in the synthesis and is not sensitive to the structure of the compact source, and the preliminary self-calibration estimates that EF has the most stable gain (Figure 4.7).

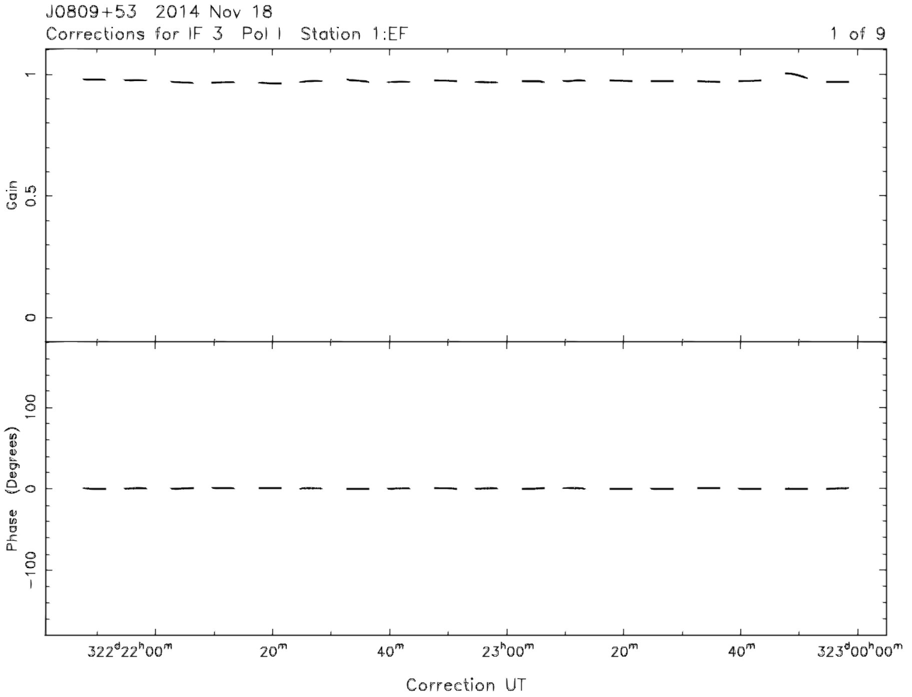


Figure 4.7: The gain amplitude and phase solutions for the EF station over the course of the observation.

For the model selection step, I assign equal priors to all the models considered, so that the logarithm of the Bayes factor B_{12} may directly be used for model comparison.

4.4.4 LIKELIHOOD FUNCTION

The likelihood for this problem is different from the one used in section 3.3.1.4 since the Gaussian noise σ is not common between baselines. Given the observed (V_D) and the modelled (V_M) visibilities, and the uncertainty σ_k that varies with baseline, the likelihood function for parameter estimation for model H may be written as

$$\begin{aligned} \mathcal{L}(\Theta|\mathrm{V}_D, H) &= \frac{1}{\prod_{k=1}^{2N_{\mathrm{vis}}} \sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{\chi^2}{2}\right), \\ \text{where } \chi^2 &= \sum_{k=1}^{2N_{\mathrm{vis}}} \left(\frac{\mathrm{V}_{M_k} - \mathrm{V}_{D_k}}{\sigma_k}\right)^2, \end{aligned} \tag{4.12}$$

and N_{vis} is the total number of complex visibilities. The summation is carried out over $2N_{\mathrm{vis}}$ since we consider the real and imaginary parts separately. Taking the natural logarithm of $\mathcal{L}$, we obtain

$$\begin{aligned} \ln(\mathcal{L}) &= \sum_{k=1}^{2N_{\mathrm{vis}}} \ln\left[(2\pi\sigma_k^2)^{-1/2}\right] - \frac{\chi^2}{2} \\ &= -\frac{1}{2}\sum_{k=1}^{2N_{\mathrm{vis}}} \ln\left[2\pi\sigma_k^2\right] - \frac{\chi^2}{2}. \end{aligned} \tag{4.13}$$

The noise is modelled such that the variance for the real and imaginary parts of a complex visibility is the same. Hence, counting each σ_k twice, we arrive at the final form of the log-likelihood function:

$$\ln(\mathcal{L}) = -\sum_{k=1}^{N_{\mathrm{vis}}} \ln\left[2\pi\sigma_k^2\right] - \frac{\chi^2}{2}. \tag{4.14}$$

The chi-squared is computed during each iteration by MONTBLANC on the GPU and the rest of equation (4.14) is evaluated in the CPU by the Python code for Bayesian analysis.

4.5 RESULTS

MULTINEST takes the prior distributions and the likelihood as its inputs, computes the natural logarithm of the Bayesian evidence for each model, and produces the joint posterior as a by-product. We evaluate each model independently and use the corresponding logarithmic evidences to compute the Bayes factor between models using equation (2.16).

The analysis reveals that the factors $2\ln(B_{ij})$, for GAU against PT and CIRC are 21.0 ± 0.8 and 26.0 ± 0.8 respectively. Applying the criteria set aside in Table 2.1, we conclude that there is a *very* *strong* preference for the elliptical Gaussian (GAU) over the other two models[*]. In linear scale, the log-odds translate to $> 10^9 : 1$ and $> 10^{11} : 1$ in favour of GAU against PT and CIRC respectively. These are very high odds against the source either being unresolved or possessing just a compact circular core. We may conclude that there is elongated structure in the source which is quite possibly, the relativistic jet. To estimate whether the brightness temperature of this source is in accordance with what is expected in the presence of a jet, we explore the posteriors of the parameters of GAU.

Figure 4.8 shows the correlations between various parameters of GAU. Using the quantities e_{maj}, e_{min}, and ϑ, obtained from equations (4.4) for visualisation, we see that the individual station SEFDs are mostly uncorrelated with the source and the gain parameters, and with each other. This is to be expected since random variations in SEFDs do not systematically affect the source shape estimates. The station gain amplitudes are correlated negatively with the estimated flux-density as one would expect: the higher the instrumental gain, the lower the true flux-density of the source. The gain amplitude of SH, $|g_{\mathrm{SH}}|$, the station that provides the longest baselines, is correlated positively with the size estimates. This indicates that there is degeneracy between these parameters (section 4.2.1) which makes it difficult to disentangle the effects of the resolution of the source from poor calibration of SH. The Noto (NT) gain amplitude, $|g_{\mathrm{NT}}|$, also shows a positive correlation with the shape parameters. Located in Italy, NT provides the longest baselines within Europe and lacks as many short baselines as the other stations. This results in a mild degeneracy between $|g_{\mathrm{NT}}|$ and e_{maj} & e_{min}[†]. There is also a strong correlation between the position estimates (although with uncertainties of the order of μas), which may be due to the position not being constrained tightly enough along the extent of the major axis of the PSF.

The 1-D and 2-D marginalised posteriors of the source and the station gain parameters, along with the derived source brightness temperature, T_b (section 4.5.3) are shown in Figure 4.9. The 2-D marginalised posterior distributions between $|g_{\mathrm{SH}}|$ and the three shape parameters show the precise nature of the correlations between them. Presenting these relationships in full is the most complete statement we can make about these parameters given the data. The SH gain amplitude is not constrained to the same precision as those of the other stations by the preliminary self-calibration process. As we have seen, SH does not form short baselines with any other station, thereby making it difficult for amplitude self-calibration to correct for the biased gains (Martí-Vidal et al., 2012). Hence, measurements involving SH have contributions from both the resolved source structure and the uncertainties in the calibration of $|g_{\mathrm{SH}}|$. We also see degeneracy between the three shape parameters e_{maj}, e_{min}, and ϑ. The shape measurements are constrained also by the fact that the entire observation lasted only for 130 minutes, with a poor *uv*-coverage (Figure 4.2). The resolution is the lowest perpendicular to the direction along which the baseline length is the longest (over 9000 km). This introduces greater uncertainty in position measurement along the major axis of the PSF, which reflects in the posteriors

---

[*] The quoted relative ln-evidences are obtained from the *importance nested sampling* results owing to their better accuracy, while the uncertainties are obtained from the *vanilla* nested sampling results (Feroz et al., 2013).

[†] See also section 4.5.1.

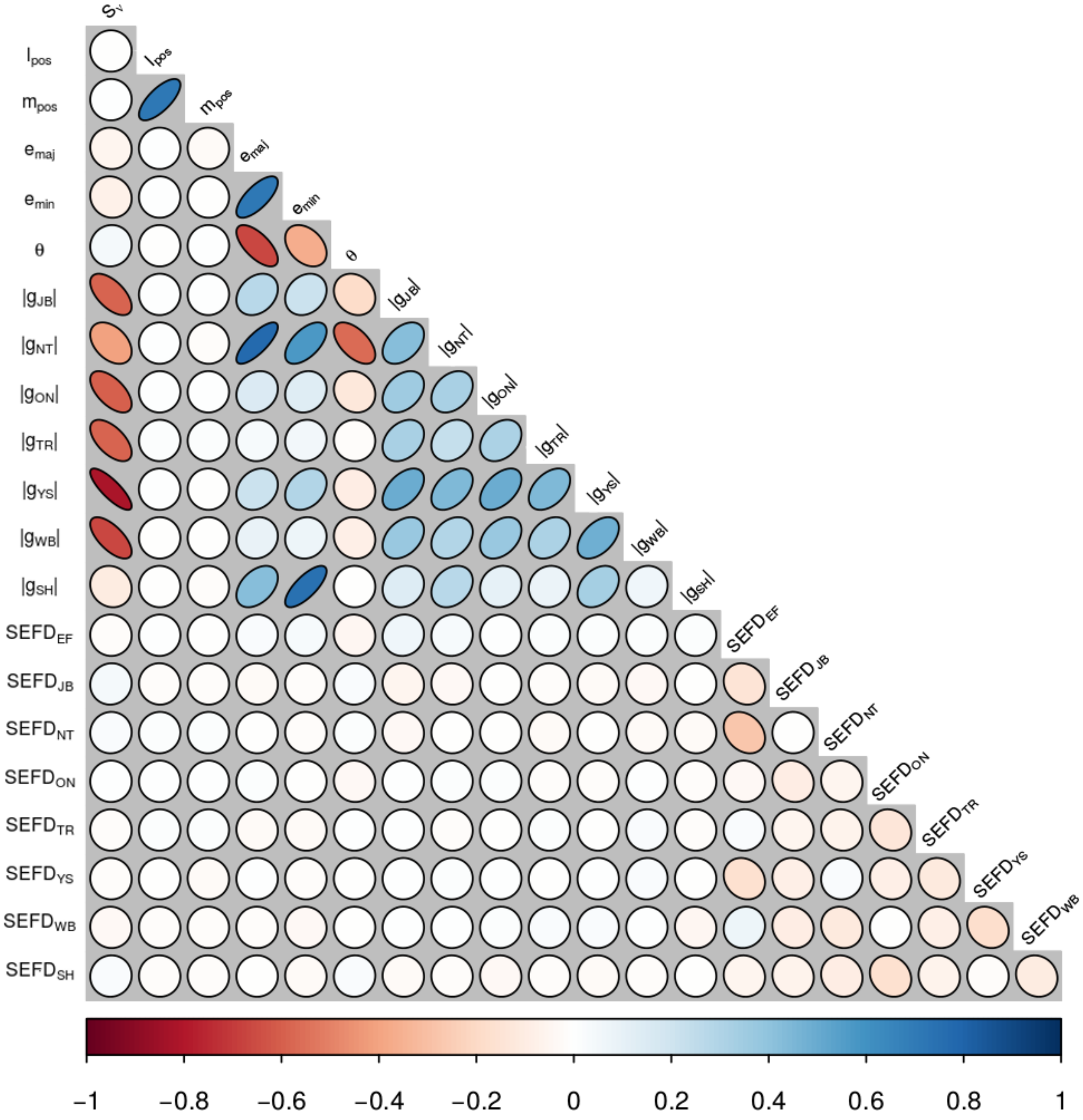


Figure 4.8: Correlations between the posteriors for model GAU. The quantities e_{maj}, e_{min}, and ϑ are derived from the shape parameters (equation (4.4)). The parameters with delta prior distributions are excluded from the figure. The coloured ellipses correspond to the correlation coefficients shown at the bottom; the higher the ellipticity, the stronger the correlation.

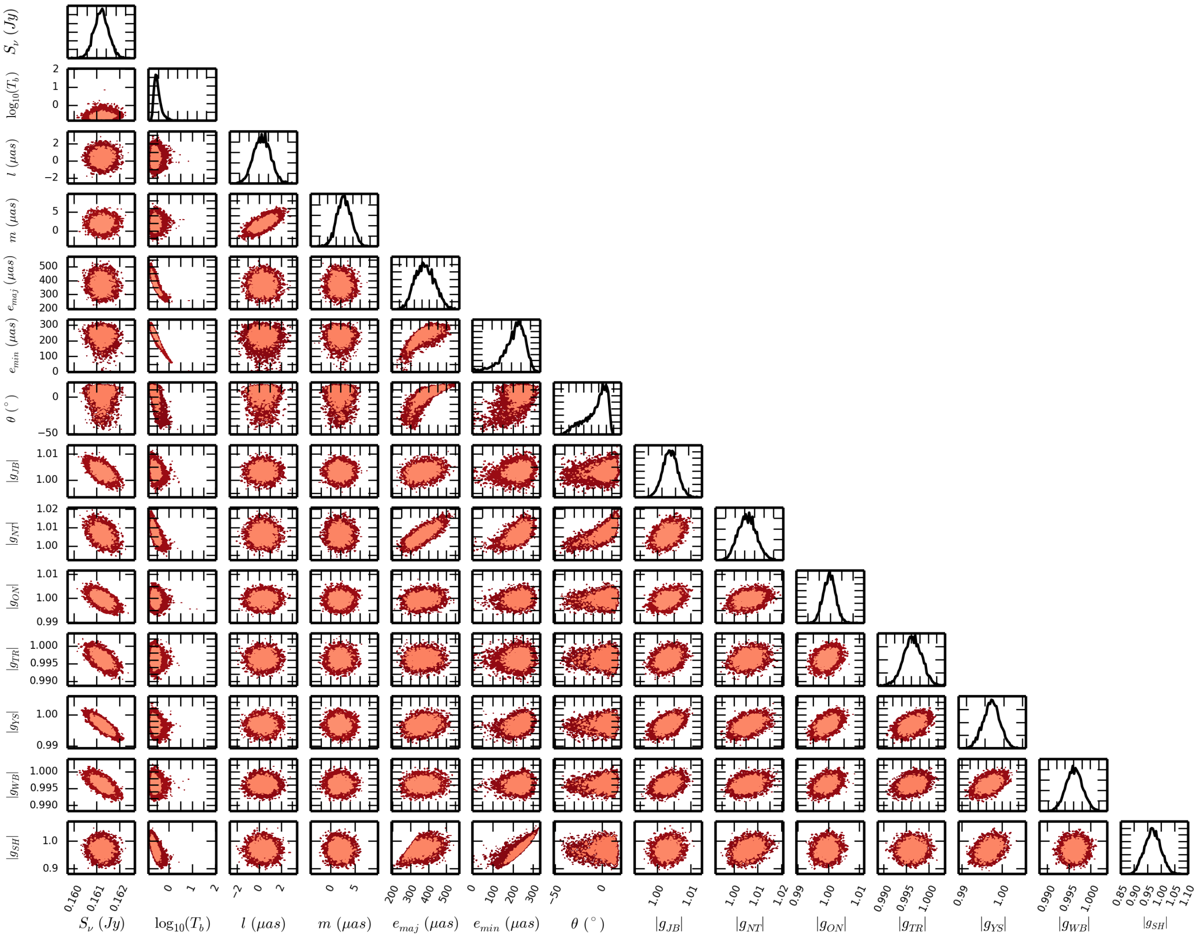


Figure 4.9: The 1-D posteriors and 2-D correlations of the source and the station gain parameters for model GAU. The principal diagonal gives the 1-D marginalised posterior distributions of the estimated flux-density (S_ν), the source brightness temperature (T_b), the derived shape quantities ($e_{\rm maj}$, $e_{\rm min}$, ϑ), and the station gain amplitudes $|g_p|$, while the lower triangular matrix gives the 2-D joint posteriors between their various combinations. The 68 and 95 per cent credible regions are indicated by the light-red and dark-red shaded regions respectively. Parameters with delta priors and the station SEFDs (which are uncorrelated with the other parameters) are excluded from the figure.

of e_{min} and ϑ which cannot be approximated by a Gaussian curve. This is precisely the reason why a Bayesian analysis which naturally produces the correlations between the parameter estimates and the uncertainties in their posteriors is necessary to resolve the source structure.

Figure 4.10 highlights the regions in the marginalised posteriors of e_{maj} and e_{min} that correspond to a 1 per cent variation in that of $|g_{\mathrm{SH}}|$ (which varies by about 12 per cent) about its mean. Providing short baselines for outlying stations such as SH by introducing more VLBI antennas close by would reduce the sparsity of the array distribution and minimise the uncertainty in the estimation of the station gains. In Shanghai, there are now two radio telescopes available for VLBI: the old Sheshan 25m (SH) and the new Tianma 65m telescope which provides the necessary short baselines for SH (Kawaguchi et al., 2015). When both are used in VLBI observations, the gain calibration of the longest EVN baselines will improve and result in reduced uncertainty in the derived shape parameter and brightness temperature distributions.

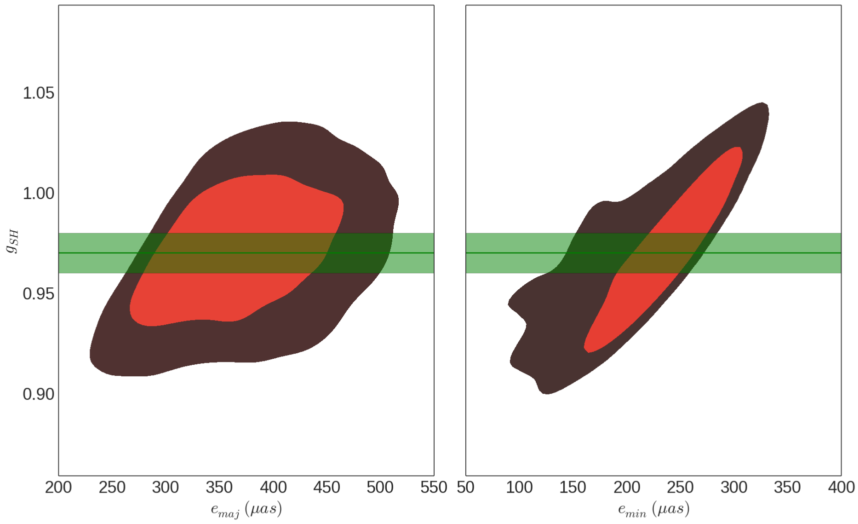


Figure 4.10: The correlations between SH gain amplitude and the shape parameters. The horizontal bars on either side of the regions shaded green correspond to ± 1 per cent about the mean of the marginalised posterior of g_{sh}. The light and dark red shaded regions indicate the 68 and 95 per cent credible regions respectively of the posteriors of the shape parameters e_{maj} and e_{min}. It is evident that a better gain calibration would constrain both e_{maj} and e_{min} better, due to the strong correlation between the two (not highlighted here; refer Figure 4.9).

Figure 4.11 shows the 1-D and 2-D marginalised posteriors of the individual station SEFDs. As with the plot of correlation coefficients, it can be seen that the SEFDs are independent of and uncorrelated with each other. Nevertheless it was necessary to estimate these values in order to quantify the baseline-dependent noise and to use the corresponding uncertainties during marginalisation to obtain the posteriors of the source parameters and the station gains.

4.5.1 Model Selection with European Baselines Only

In the absence of SH, NT provides the longest baselines in the synthesis. The NT gain amplitude is positively correlated with the shape parameters (Figures 4.8 & 4.9). This correlation provided the

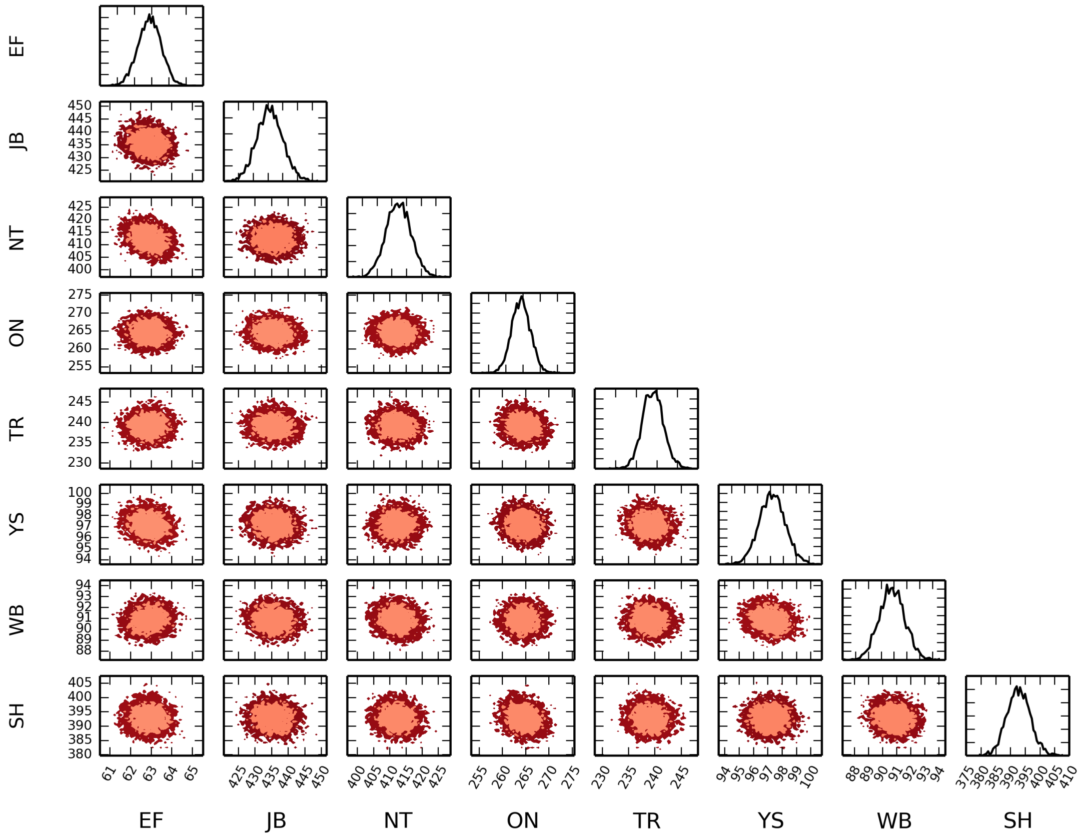


Figure 4.11: The 1-D and 2-D marginalised posteriors of the estimated individual station SEFDs pertaining to the elliptical Gaussian morphology GAU.

motivation for testing whether the source can successfully be resolved using only the baselines corresponding to the European stations. I evaluated all three models described in section 4.4.1 on these data. The relative logarithmic evidence between PT and GAU comes out to be 2.86 ± 0.75, while that between PT and CIRC is 1.61 ± 0.76. There is positive evidence for PT and a mild preference for CIRC over GAU but according to Jeffreys' criterion (Table 2.1) the model selection ratio is not strong enough to conclude that the source is resolved. So, without the measurements from the SH station, we are forced to conclude that the source is unresolved.

4.5.2 COMPARISON WITH DIFMAP RESULTS

We performed model-fitting to the data using DIFMAP[*], conventionally used in VLBI, so that we could compare our results with those reported by An et al. (2016). DIFMAP is a software package for self-calibrating visibilities and building up a model of the sky iteratively (Shepherd, 1997). It returns the best-fit parameter estimates and the reduced chi-squared, χ^2_{red}, for three different models as a measure of their goodness-of-fit[†] to the data. The χ^2_{red} values for these models, obtained with and without the SH measurements in the data, are shown in Table 4.5. These models include only the source parameters and do not account for the instrumental effects considered in the original three models (Table 4.3). As a result, they are much simpler than the models evaluated using the RIME.

Table 4.5: Reduced chi-squared values for the three models obtained with and without the SH measurements using Difmap. The asterisk indicates that these models do not take instrumental effects into account.

Model	χ^2_{red} with SH (DoF = 164173)	χ^2_{red} without SH (DoF = 121501)
PT*	2.2442	1.7266
GAU*	1.8746	1.7253
CIRC*	1.8785	1.7256

With the SH measurements included, there is a slight preference for the resolved source models. Without the SH measurements, as with the Bayesian analysis, PT is at least as good a fit as any other model if not better, because of its simplicity. Though χ^2_{red} has traditionally been used for model selection, it is a poor substitute for Bayesian evidence since it assumes that the underlying processes are Gaussian. The Bayesian approach does not depend on this assumption. Moreover, there is no measurement of the uncertainty in the value of χ^2_{red}, which becomes important when the χ^2_{red} values of two models do not differ much. In contrast, the Bayesian evidence has an uncertainty associated with it which we may use to determine the significance of the model selection ratio.

---

[*] `ftp://ftp.astro.caltech.edu/pub/difmap/difmap.html`

[†] $\chi^2_{\text{red}} = \chi^2$ / DoF, where DoF stands for *degrees of freedom*, obtained by subtracting the number of model parameters from the number of measurements.

Figure 4.12 shows the relation between the DIFMAP best-estimates and the posterior distributions of the source flux-density and shape estimates obtained using the Bayesian analysis. The DIFMAP estimates are point estimates with no associated uncertainties. From the figure, we see that they fall within the 68 per cent credible region of the Bayesian posteriors when the posteriors are approximately Gaussian (S and e_{maj}). For e_{min} and ϑ, the posteriors are non-Gaussian and must be presented and accounted for in full while drawing inferences, such as estimating the brightness temperature distribution of the blazar to detect the presence of intrinsic source structure.

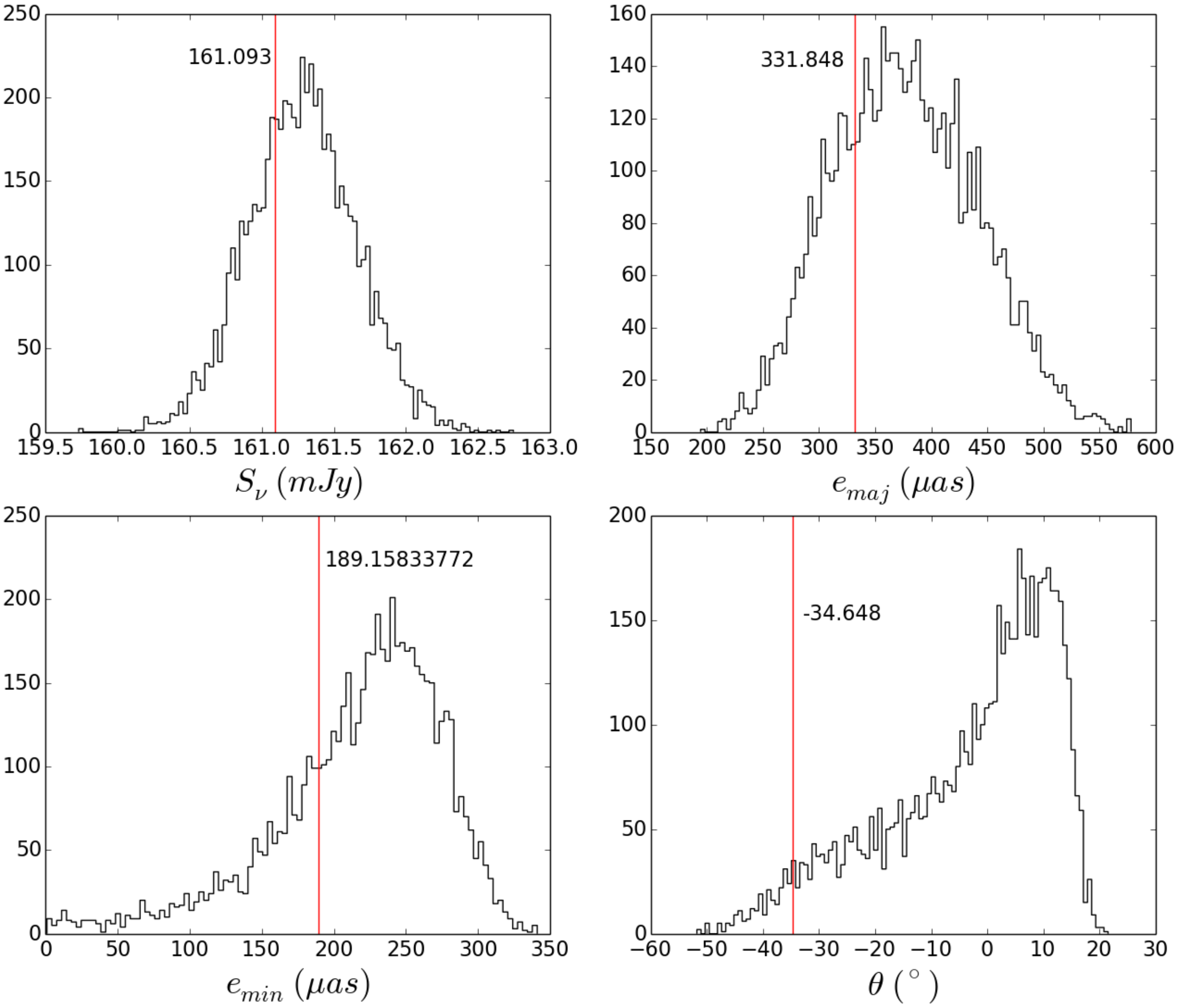


Figure 4.12: Comparison of the posteriors of the source parameters with the corresponding Difmap estimates. The vertical red lines correspond to the best-fit Difmap values printed alongside and the histograms correspond to the 1-D marginalised posteriors of the source parameters.

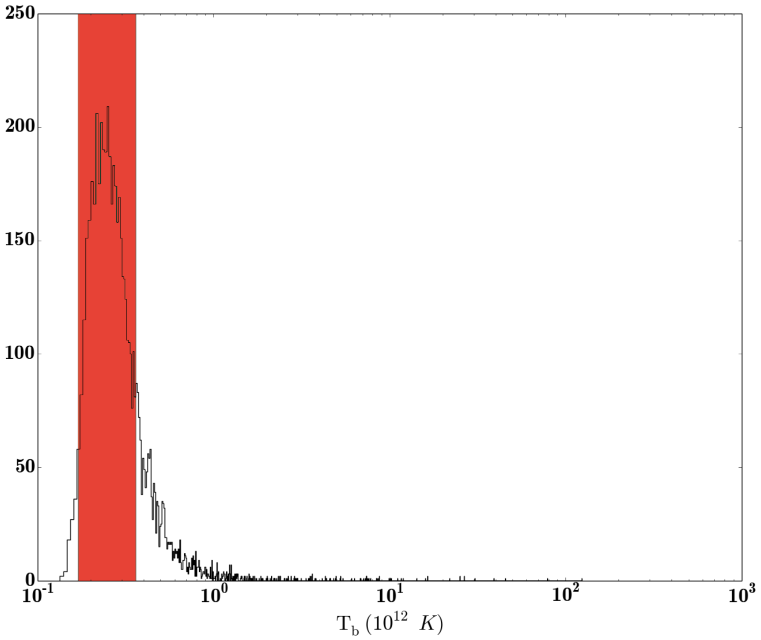


Figure 4.13: Histogram of the brightness temperature of CGRaBS J0809+5341 shown in black. The 68 per cent credible region (0.17 to 0.36) around the mode (0.25) is shaded light red.

4.5.3 WHAT DOES THIS MEAN FOR THE BLAZAR?

Equation (4.1) gives the expression for calculating the brightness temperature of a radio source in the Rayleigh-Jeans limit. For a source with Gaussian morphology at high redshifts, this relation becomes (Kellerman & Owen, 1988)

$$T_b = 1.22 \, \frac{S_\nu}{\nu^2 \, e_{\mathrm{maj}} e_{\mathrm{min}}} \, (1+z) \; 10^{12} \ \mathrm{K}, \tag{4.15}$$

With $\nu = 4.98224$ GHz and $z = 2.144$, I use the full posterior distributions of S_ν, e_{maj}, and e_{min} to infer the distribution of T_b (Figure 4.13). This ensures that our measurements account for the uncertainties in the source flux-density and shape estimates. The mode of the distribution occurs at 0.25×10^{12} K. The 68 per cent credible region places T_b between 0.17×10^{12} K and 0.36×10^{12} K. The few very high values of T_b correspond to the very low values of e_{min}.

Our T_b estimates for the blazar lie within the range corresponding to the inverse Compton limit of $\simeq 10^{11} - 10^{12}$ K for the brightness temperature (Kellerman & Pauliny-Toth, 1969). As explained in section 4.1.2, this is a consequence of *Doppler boosting* in which the relativistic beaming of a jet moving at a small angle to the line of sight to the observer increases the observed flux-density (Kellerman & Owen, 1988). Moreover the jet would be asymmetric, since the part of the jet moving towards the observer will be Doppler-boosted while the flux-density of the part that moves away from the observer will be diminished to such an extent that it might go unobserved. Since $T_b = \delta \, T_{b,int}$, where δ is the *Doppler boosting factor* and $T_{b,int}$ is the intrinsic brightness temperature, our result implies that the source is indeed relativistically Doppler-boosted i.e. $\delta > 1$ and that there is strong reason to believe that this source is composed of particles moving at relativistic speeds, beamed more or less in the direction of the Earth.

4.6 A BAYESIAN CRITERION FOR THE RESOLUTION LIMIT

Martí-Vidal et al. (2012) derived the maximum theoretical super-resolution power of an interferometer, dependent on the signal-to-noise ratio of the visibility measurements, from theoretical considerations. The minimum resolvable source size ϑ_M is given by

$$\vartheta_M = \beta \left(\frac{l_c}{2(\mathrm{SNR})^2} \right)^{1/4} \times \mathrm{FWHM}. \tag{4.16}$$

β depends on the shape and the intensity profile of the source model and takes values between 0.5 and 1. For source profiles with higher intensities at lower scales, the value of β is closer to 1. l_c is the log-likelihood value that corresponds to the critical probability of the null hypothesis, which is taken to be the point source model. It takes the values 3.84 and 8.81 for a 5 per cent and a 0.3 per cent probability cut-off of the null hypothesis respectively (Martí-Vidal et al., 2012). The FWHM is the *full-width at half maximum* of a circular PSF. For an interferometer with an elliptical PSF such as ours, the circular FWHM equivalent is given by $\sqrt{ab}$, where a and b are the major and minor axes of the PSF respectively (Lobanov, 2005).

The signal-to-noise ratio (SNR) is calculated by computing the ratio between the weighted average of the visibilities and the noise $\sigma = \sigma_{\mathrm{vis}}/\sqrt{N}$. If σ varies between measurements (equation 4.5), then by *Parseval's theorem*[*], the rms noise σ_{rms} in the naturally-weighted residual image of the sky (after the source has been subtracted out) can be equated to σ as

$$\begin{aligned} \sigma^2 &= \frac{\sigma^2_{\mathrm{vis}}}{N} = \sigma^2_{\mathrm{rms}}, \\ \sigma &= \sigma_{\mathrm{rms}}. \end{aligned} \tag{4.17}$$

Equation (4.16) gives the theoretical limit for super-resolution of an interferometer under ideal conditions. Under realistic observing conditions such as the presence of RFI or gain corruptions and antenna miscalibrations, one would expect the limit to be worse than the theoretical limit. I set out to derive a criterion for the super-resolution limit using our probabilistic approach on synthetic observations.

I set the minimum resolvable size, ϑ_B, to the size of the source at which the evidence for the resolved source model (CIRC) against the unresolved source model (PT) turns *positive* (Table 2.1). The parameters of CIRC and PT are the same as the ones used in the analysis of the CGRaBS J0809+5341 data

---

[*] Parseval's theorem, in this context, ensures that the total power of the noise remains the same regardless of whether it is computed in the image domain or in the uv-domain.

(Table 4.3). To compare ϑ_M with ϑ_B, I simulated a series of observations, each with a compact circular Gaussian source of decreasing sizes located at the pointing centre of the interferometer. For the station gains and SEFDs, I used the *maximum a posteriori* (MAP) estimates obtained from analysing the actual data. I performed model selection between CIRC and PT for three such sets of simulated data for three different SNR levels, where SNR is calculated according to equation (4.17), and computed the Bayes factor between CIRC and PT (Table 4.6).

Table 4.6: A comparison of the resolution limits denoted by ϑ_M and ϑ_B, obtained using equation (4.16) with $\beta = 1$ and $l_c = 8.81$ and by using the Bayesian approach respectively, for the sparse VLBI array described in Table 4.2 for different SNR levels at 5 GHz. The circular FWHM equivalent of the naturally-weighted elliptical PSF is $\sqrt{5.7 \times 2.2} = 3.54$ mas. For higher SNR levels, ϑ_B is limited more by the gain amplitude calibration than by the theoretical capabilities of the array. The last column gives the maximum brightness temperature one could measure for a source of 1 Jy, derived from ϑ_B.

SNR	ϑ_M (mas)	ϑ_B (mas)	ϑ_B/FWHM	T_b/S_ν (10^{12} K/Jy)
150	0.42	0.45	0.13	0.763
2000	0.11	0.29	0.09	1.837
5900	0.07	0.17	0.05	5.347

The three sets of simulated observations had SNR levels of 150:1, 2000:1, and 5900:1. From Table 4.6, we see that the source size ϑ_B at which the Bayesian method starts preferring CIRC over PT with log-Bayes factor of > 2 is larger than the theoretical limit which depends only on the SNR and assumes perfect calibration of instrumental instabilities. Martí-Vidal et al. (2012) observe how various factors such as the proportion of long baselines in an array configuration, source structure, and biased gains of the antennas providing the long baselines will further limit the resolving power of the interferometer. The Bayesian approach is sensitive to factors such as source shape, biased instrumental gains, and the associated uncertainties and provides a more realistic estimate of the minimum resolvable source size. This is made possible due to the RIME being able to model realistic instrumental effects. At low SNR, we see that ϑ_B is close to the theoretical limit ϑ_M. As the SNR improves, the instabilities in instrumental gains injected into the simulations start playing a more significant role and the ϑ_B is up to $2-3$ times larger than ϑ_M. Substituting the minimum resolvable size into equation (4.15), we have also derived the maximum T_b that could be measured given an interferometer configuration.

4.7 CONCLUSIONS

Based on the Bayes factors obtained, we have very strong evidence for a resolved source with slightly elongated structure, in agreement with what is expected for a partially synchrotron self-absorbed compact jet in a flaring blazar. By simultaneously estimating source parameters along with the antenna gains and SEFDs, we have also acquired knowledge of the precision of our estimate of the source size and its correlation with antenna gains. Without SH visibilities, the Bayesian evidence indicates that the source must be considered unresolved.

$|g_{\mathrm{SH}}|$ varies by about 12 per cent, which is about an order of magnitude worse than the precision in the gain amplitude calibration of all but one (NT) European station. The dependence of the shape parameters on $|g_{\mathrm{SH}}|$ results in the minor axis and the position angle of the Gaussian in model GAU being poorly constrained. This also illustrates the necessity of accounting for calibration errors for the stations providing the long baselines, lest we are led astray in our attempts to estimate the parameters relevant for the science goals.

Currently, we are limited only by the performance of the MultiNest sampler required to compute the evidence. The RIME can model any time or frequency variation in the source and in the instrumental effects. Within this framework, data from multiple spectral bands and epochs can be analysed together, incorporating time and frequency variation in the complex antenna gains. Future analyses will benefit from numerical samplers for evidence computation tailored for higher-dimensional parameter spaces (e.g. Handley et al., 2015a,b).

The brightness temperature distribution we have derived for CGRaBS J0809+5341 accounts for the uncertainties in the source shape and instrumental gain calibration and indicates that the source is Doppler-boosted and that the intrinsic brightness temperature must be lower than what we measure, consistent with the literature.

The analysis of synthetic observations of different SNR levels shows that, as the SNR improves (SNR$\gg$100), we are constrained less by the theoretical capabilities of the interferometer array and more by the effects of miscalibration of station gains and the discrepancy between ϑ_M and ϑ_B becomes more pronounced. The constraints derived by Martí-Vidal et al. (2012) assume perfect calibration. Our approach is more sensitive to uncertainties in source shape estimates and their correlation with instrumental gains and provides more realistic estimates that are 2 to 3 times the estimates obtained from equation (4.16). Knowledge of ϑ_B also enables us to estimate the maximum brightness temperature that one can measure for a given interferometer configuration. Such analyses will help in setting a practical limit to the super-resolution capabilities of sensitive future VLBI experiments with the Square Kilometre Array (SKA-VLBI), where the resolution of source structure down to a few tens of μas with high precision can be reasonably expected, provided the gains of stations providing long baselines are pinned down as accurately as possible.

4.8 Future Work

The gains assumed in this analysis are time-invariant. Although they prove useful in performing model selection between point and resolved source models and studying major effects of gain instabilities on visibilities measured by long baselines, this is too simplistic an assumption to enable us to perform deeper studies of intrinsic source structure and more realistic effects of time-dependent gains on visibilities. Using *Kalman filters*[*] for time-dependent gain calibration would enable us to model more realistic gain effects (Tasse, C., 2014). This long-term goal would also be relevant to other problems such as fringe-fitting (Chapter 5).

The preliminary phase self-calibration would have introduced correlations between visibility phases measured across different baselines. These, depending upon how severe the effects of self-calibration are, would leave us with data containing systematic biases. Taking these effects into consideration would give us better estimates for parameters such as the source positions. The current software setup is capable of handling these effects, at the cost of increased execution time. Nevertheless, this is a promising line of analysis for the future. Finally, the use of more complex source models would enable us to study the structure of the blazar in more detail.

---

[*] Unlike the least-squares method, which attempts to find the best-fit estimate, Kalman filters minimise the uncertainty on the estimate given information on previous states.

5

Probabilistic Fringe-fitting

An essential component of interferometry is the estimation of the spatial coherence function (the source coherency in equation (1.31)) of electromagnetic fields measured at two different locations in space. VLBI observations are similar to those made using connected-element interferometers insofar as the aim is to obtain this coherence function. But, since the stations participating in a VLBI observation are typically located thousands of kilometres apart, techniques different from those employed for connected-element interferometry are brought into play to generate and analyse VLBI data. Due to the long baselines involved ($>$ 100 km), the atmospheric (both tropospheric and ionospheric) conditions at the individual stations are different, leading to different atmospheric propagation delays that are uncorrelated (Thompson et al., 2017, Chapter 9). Moreover, the use of independent frequency standards at the participant stations (the local oscillators are not shared over the large distances involved) introduces another source of systematic uncertainty in determining the time at which a wavefront arrives at a given station. These errors, if not corrected for, may decohere the signal completely in the worst case or reduce the *coherence time*, the time interval over which the measured visibilities can be averaged without a net loss of amplitude that results in reduced sensitivity of the data (Schwab & Cotton, 1983).

A correlator must correlate the signals from different stations while ensuring that the signals correspond to the same incoming wavefront. But the stations are located at different distances from the source, so the wavefront arrives at different times (the *geometric delay* or, simply, the *delay*). Besides, the stations are moving at different speeds along the direction to the source, resulting in different Doppler shifts of the incoming wave (the *fringe rate* or, simply, the *rate*) (Taylor et al., 1999, Chapter 22). Various factors such as Earth rotation and atmospheric effects, antenna structure, station clocks, and source structure affect these time- and frequency-dependent variations in the phases[*] of the measured visibilities (Thompson et al., 2017, Chapter 13). These and many more subtle effects are accounted for in the *correlator model*, a geometrical model used by the correlator to estimate and correct for the delays and fringe rates (Taylor et al., 1999, Chapter 22). Systematic errors in the model result in residual time- and frequency- dependent phase variations, along with a *phase offset*. Correcting for these *residual* rates and delays, along with the phase offsets, in a process known as *fringe-fitting* or *fringe-search*, allows one to average data over larger intervals of time and frequency, thereby increasing the coherence time and improving the sensitivity.

Fringe-fitting is typically one of the first steps performed before self-calibrating and imaging VLBI data. The EVN observation of CGRaBS J0809+5341 described in Chapter 4 had already been fringe-fitted and self-calibrated (section 4.2) before I analysed it using our Bayesian inference framework. In this chapter, we aim to extend the application of Bayesian inference to include fringe-fitting while simultaneously estimating the source properties. Since the process involves solving for phases and their time and frequency derivatives, it can be expressed in terms of Jones matrices in the RIME formalism. The residual phase offsets, delays, and rates can be expressed in parametric form and incorporated into the phases of the individual complex terms of these Jones matrices. We then test this formalism by performing parameter estimation and model selection on synthetic observations of compact sources made using the Very Long Baseline Array (VLBA) and the Event Horizon Telescope (EHT, Doeleman et al., 2009).

5.1 FRINGE-FITTING

The atmospheric conditions over the different stations vary considerably in any VLBI observation which results in significant phase variations being introduced in the measured visibilities. This results in loss of amplitude of the visibilities called *decoherence*[†]. Fringe-fitting is the process of minimising this decoherence by solving for the phase variations that remain after the correlator model has been applied to the visibilities. This enables further averaging after the correlation to improve the SNR of the data. The old baseline-based fringe-fitting method has been replaced by 'global' fringe-fitting methods that make use of a source model to perform fringe-fitting more efficiently, especially in the low signal-to-noise ratio (SNR) region. Nevertheless, it is instructive to look at how the baseline-based method works to gain insight into how the global techniques improve on it.

5.1.1 BASELINE-BASED METHOD

The baseline-based method is applicable when the SNR is high enough on each baseline, so that solutions for the delays and rates could be obtained per baseline. Let $\widetilde{V}_{pq}(t_m, \nu_n)$ be the measured visibility corresponding to baseline pq at time t_m and frequency ν_n. This is related to the true visibility

---

[*] Differences in path lengths translate to changes in the phases of the complex-valued visibilities.

[†] In Smirnov (2011a), the same effect arising specifically due to the K-Jones term is called *smearing*.

$V_{pq}(t_m, \nu_n)$ as

$$\widetilde{V}_{pq}(t_m, \nu_n) = g_p(t_m, \nu_n)\bar{g}_q(t_m, \nu_n)V_{pq}(t_m, \nu_n) + \varepsilon_{pqmn}, \tag{5.1}$$

where ε_{pqmn} is the additive noise term arising due to thermal noise; g_p are complex-valued functions that incorporate antenna effects such as clock errors and atmospheric effects and can be expressed as

$$g_p(t, \nu) = |g_p(t, \nu)| e^{i\psi_p(t,\nu)}, \tag{5.2}$$

where ψ_p is the phase of antenna p. Assuming that the amplitudes $|g_p|$ and $|V_{pq}|$ vary slowly enough that they are constant over the time and frequency averaging intervals, we can write to first-order (Schwab & Cotton, 1983)

$$\begin{aligned} \widetilde{V}_{pq}(t_m, \nu_n) \simeq{}& |g_p||g_q|V_{pq}(t_0, \nu_0)\exp(i[(\psi_p - \psi_q)(t_0, \nu_0)]) \\ &\times \exp\left(i\left[\left.\frac{\partial(\psi_p - \psi_q + \varphi_{pq})}{\partial t}\right|_{(t_0,\nu_0)}(t_m - t_0)\right.\right. \\ &\left.\left.+\left.\frac{\partial(\psi_p - \psi_q + \varphi_{pq})}{\partial \nu}\right|_{(t_0,\nu_0)}(\nu_n - \nu_0)\right]\right), \end{aligned} \tag{5.3}$$

where $\varphi_{pq} \equiv \arg V_{pq}$. Here t_0 and ν_0 are the time and frequency relative to which the calculations are carried out and are called the *reference time* and the *reference frequency* respectively. Note that the baseline-based $\partial\varphi/\partial t$ and $\partial\varphi/\partial\nu$ are not separable from the antenna- based $\partial\psi/\partial t$ and $\partial\psi/\partial\nu$. The quantity

$$r_{pq} \equiv \left.\frac{\partial(\psi_p - \psi_q + \varphi_{pq})}{\partial t}\right|_{(t_0,\nu_0)} \tag{5.4}$$

is the expression for the fringe rate. The fringe rate is variously known as *rate*, *fringe frequency*, *rate residual*, or *delay rate*. The expression

$$\tau_{pq} \equiv \left.\frac{\partial(\psi_p - \psi_q + \varphi_{pq})}{\partial \nu}\right|_{(t_0,\nu_0)} \tag{5.5}$$

quantifies the delay, also known as *delay residual* or *time residual*. Note that the rates and the delays are relative to t_0 and ν_0. r_{pq} and τ_{pq} are estimated by Fourier-transforming the visibilities from the time-frequency domain to the delay-rate domain (Taylor et al., 1999, Chapter 22). We define a two-dimensional distribution $F_{pq}(t, \nu)$ as

$$F_{pq}(t, \nu) \equiv \sum_{m,n} \delta(t - (t_m - t_0), \nu - (\nu_n - \nu_0))\widetilde{V}_{pq}(t_m, \nu_n), \tag{5.6}$$

where δ is the Dirac delta function[*]. The delay and rate are estimated by searching for the maximum of the Fourier transform $\hat{F}_{pq}(r, \tau)$ of $F_{pq}(t, \nu)$ (Figure 5.1). The estimates of r_{pq} and τ_{pq} thus obtained enable us to obtain the phase-corrected visibilities

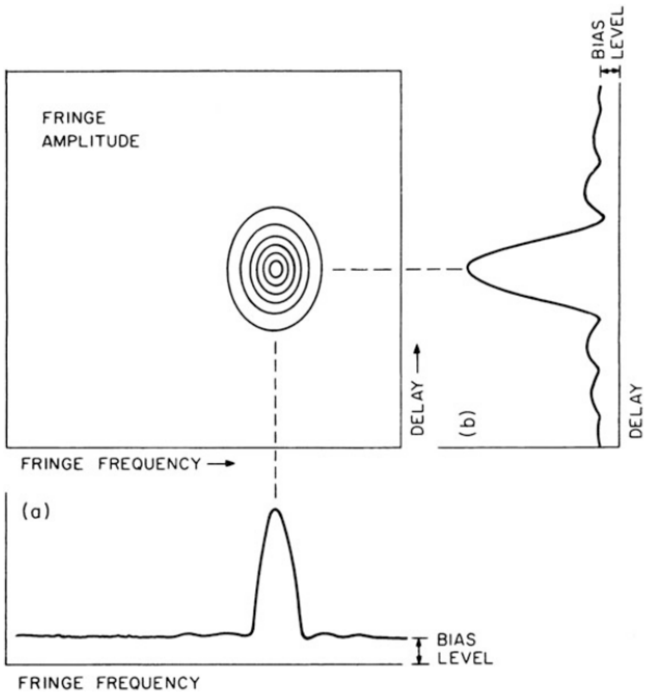


Figure 5.1: The modulus of $\hat{F}_{pq}(r, \tau)$. The 1-D plots are cross-sections of the 2-D plot along the rate and delay axes. The bias level is introduced by the presence of noise (Thompson et al., 2017, Chapter 9).

$$= \widetilde{V}_{pq}(t_m, \nu_l) e^{(-i[(t_m - t_0) r_{pq} + (\nu_n - \nu_0)\tau_{pq}])}. \tag{5.7}$$

These can be averaged coherently over time and frequency, to the extent that the first-order model in equation (5.3) is valid (Schwab & Cotton, 1983). It is important to note that the derivatives of the antenna phases ψ and those of the true visibility phases φ cannot be separated by this method.

5.1.2 GLOBAL FRINGE-FITTING

The baseline-based approach outlined above requires relatively high SNR[†] on each baseline in order to acquire reliable estimates of all the r_{pq} and τ_{pq}. To handle data with low SNR, which is often the case with VLBI, it is necessary to separate the antenna-based components $\partial\psi_p/\partial t\big|_{t_0,\nu_0}$ and $\partial\psi_p/\partial\nu\big|_{t_0,\nu_0}$ and estimate them simultaneously for all stations using the data corresponding to all baselines. To do this, we need a model of the source V^M_{pq} approximating the true visibilities V_{pq}. Defining a function

---

[*] With the property that its Fourier transform $\hat{\delta}(r, \tau) = 1$.

[†] If there are enough data per baseline, then they can be stacked together in time to improve the sensitivity.

E_{pqmn} in the delay-rate domain (Schwab & Cotton, 1983),

$$E_{pqmn} \equiv e^{(i[(\psi_{p0}-\psi_{q0})+(r_p-r_q)(t_m-t_0)+(\tau_p-\tau_q)(\nu_n-\nu_0)])}, \tag{5.8}$$

equation (5.1) can be recast as

$$\widetilde{V}_{pq}(t,\nu) = |g_p||g_q| V^M_{pq}(t,\nu) E_{pqmn} + \varepsilon_{pqmn}. \tag{5.9}$$

There are four unknowns for each antenna in the above equation: $|g_p|$, ψ_p, r_p, and τ_p. All these terms are in the form of relative phases between antennas. In the absence of further assumptions, we designate one antenna to be the *reference antenna* so that the other phases can be cast relative to this antenna. Thus, if we set $\psi_{p0} = r_p = \tau_p = 0$, then the phase, rate, and delay of any antenna q are estimated relative to those of antenna p. If the interferometer array consists of n_A (> 2) antennas, then the number of parameters reduces to $4n_A - 3$. Usually, the amplitude terms $|g_p|$ are calibrated separately and are not considered during fringe-fitting. Setting $|g_p| \equiv 1$, we therefore bring down the number of parameters to $3(n_A - 1)$ (Thompson et al., 2017, Chapter 9). Two different approaches to global fringe-fitting are in use, one of which is used to generate the starting guesses for the other, in the now standard fringe-fitting algorithms in software such as AIPS[*].

5.1.2.1 GENERALISATION OF THE BASELINE-BASED METHOD

As we did for equation (5.6), here we define the function $F_{pq}(t,\nu)$, but consider only the phases:

$$F_{pq}(t,\nu) = \sum_{m,n} w_{pqmn}\delta(t-(t_m-t_0), \nu-(\nu_n-\nu_0))e^{i[(\widetilde{\phi}_{pq}-\phi^M_{pq})(t_m,\nu_n)]}. \tag{5.10}$$

Here the weighting function w_{pqmn} is taken to be the reciprocal of the variance of $\widetilde{\varphi}_{pqmn}$ to minimise the variance of $F_{pq}(t,\nu)$. Thus larger samples (with less uncertainty) are given higher weights than smaller samples. Searching for the location of the maximum of the modulus of the Fourier transform of this function in the delay-rate domain provides an estimate for r_{pq} and τ_{pq} (since the phases are calculated relative to that of antenna p, these can simply be denoted respectively by r_q and τ_q). The argument of the Fourier transform gives an estimate of ψ_{q0}. The crucial difference here is the ability to separate the instrumental phases into antenna-based terms facilitated by the use of a source model.

A straightforward extension of the baseline-based approach outlined in section 5.1.1 to multiple baselines exploits the principle of *closure phases* (Jennison, 1958; Thompson et al., 2017). Closure relationships between complex-valued visibilities can be constructed for baselines that form a closed loop. For instance, the measured visibility phases $\widetilde{\varphi}_{pq}$ are related to the true visibility-phases φ_{pq} by

$$\widetilde{\varphi}_{pq} = \varphi_{pq} + \psi_p - \psi_q + \varepsilon_{pq}, \tag{5.11}$$

---

[*] `www.aips.nrao.edu`

where $\psi_p - \psi_q = \psi_{pq}$ is the relative instrumental phase of baseline pq and ε_{pq} is the noise on baseline pq. The closure phase of the measured visibilities $\widetilde{\varphi}_{pqr}$ is then given by

$$\begin{aligned}\widetilde{\varphi}_{pqr} &= \widetilde{\varphi}_{pq} + \widetilde{\varphi}_{qr} + \widetilde{\varphi}_{rp} \\ &= \varphi_{pq} + \varphi_{qr} + \varphi_{rp} + \varepsilon_{pqr},\end{aligned} \tag{5.12}$$

where ε_{pqr} is the additive noise term that remains after the instrumental phases ψ_p have cancelled out*.

Ignoring the noise term in equation (5.11) and remembering that we are dealing with model visibilities with phases φ^M_{pq}, we obtain

$$\psi_{pq} \equiv \psi_p - \psi_q = \widetilde{\varphi}_{pq} - \varphi^M_{pq}. \tag{5.14}$$

Now, the baseline pq can also be obtained via the path p–r–q (Figure 5.2). The instrumental phase $\psi_{\overline{prq}}$† can then be written as

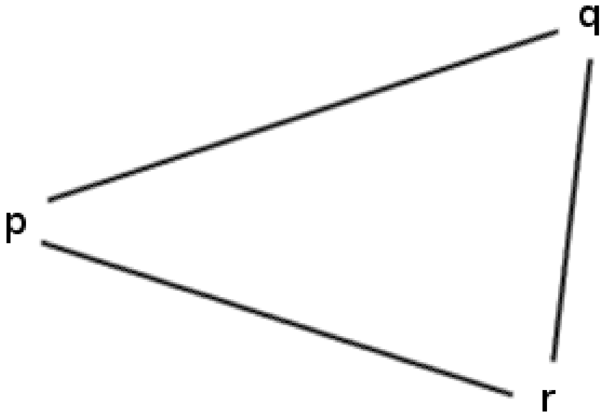


Figure 5.2: The baseline pq can also be obtained via a different path $p - r - q$.

$$\begin{aligned}\psi_{\overline{prq}} &= \psi_{pr} + \psi_{rq} = (\widetilde{\varphi}_{pr} - \varphi^M_{pr}) + (\widetilde{\varphi}_{rq} - \varphi^M_{rq}), \\ &= \psi_p - \psi_r + \psi_r - \psi_q = \psi_p - \psi_q.\end{aligned} \tag{5.15}$$

Thus $\psi_{\overline{prq}}$ provides another estimate for ψ_{pq}. Equation (5.10) can be written for all such paths pq. Since the phase differences appearing in all the 3-antenna paths are independent, the Fourier transform of

---

[*] A related concept called *closure amplitude* Γ_{pqrs} can be derived for four or more antennas as (e.g. Thompson et al., 2017, Chapter 10)

$$\Gamma_{pqrs} = \frac{|\widetilde{V}_{pq}||\widetilde{V}_{rs}|}{|\widetilde{V}_{pr}||\widetilde{V}_{qs}|}. \tag{5.13}$$

This quantity is used in amplitude self-calibration.

[†] The bar over the subscript indicates that the loop is *not* closed, in contrast with the notation used for denoting closure phase.

$F \equiv F_{pq} + \sum_{i \neq p,q} F_{\overline{piq}}$ provides better estimates of rates and delays than via any of the F's alone (Schwab & Cotton, 1983).

5.1.2.2 LEAST-SQUARES METHOD

The delay and rate estimates obtained above are used as starting guesses for a least-squares solution to the functional form shown in equation (5.9). As with the previous section, we ignore the amplitude information and consider only the phases. Choosing appropriate weights w_{pqmn} as in equation (5.10), we define the function S as

$$S(\mathrm{X}) = \sum_{m,n} \sum_{p<q} w_{pqmn} \times \left| \exp(i[\widetilde{\varphi}_{pq}(t_m, \nu_n) - \varphi^M_{pq}(t_m, \nu_n)]) - E_{pqmn} \right|^2, \tag{5.16}$$

where X is the column vector of parameters $\psi_{10}, \ldots, \psi_{n_A 0}, r_1, \ldots, r_{n_A}, \tau_1, \ldots, \tau_{n_A}$ and E_{pqmn} is given by equation (5.8). A least-squares minimisation of S with respect to X yields the solutions for the phase offsets, delays, and rates (Schwab & Cotton, 1983).

5.1.2.3 SELF-CALIBRATION AND FRINGE-FITTING

Self-calibration is similar to fringe-fitting, in that in both approaches the basic principle is to allow the station gains g_p to vary. Both rely on using *a priori* knowledge of the source to solve the overdetermined system of equations that describes the visibility data.

Self-calibration aims to create and iteratively refine a model $I^M \equiv \mathcal{F}(V^M)$ of the sky brightness distribution S_{SC} by performing a least-squares minimisation to estimate the complex gains (Taylor et al., 1999, Chapter 10):

$$S_{SC} = \sum_{m,n} \sum_{\substack{p,q \\ p \neq q}} w_{pqmn} \left| \widetilde{\mathrm{V}}_{pq}(t_m, \nu_n) - g_p(t_m, \nu_n) g^{\dagger}_q(t_m, \nu_n) \mathrm{V}^M_{pq}(t_m, \nu_n) \right|^2. \tag{5.17}$$

The estimated gains are used to correct the visibilities and form a new model of the sky and the process is carried out iteratively until an acceptable model is reached. Readhead & Wilkinson (1978) and Readhead et al. (1980) outlined methods to incorporate closure phases (equations 5.11–5.12) and closure amplitudes (equation 5.13) in iteratively developing a model of the sky. The latter case is referred to as *hybrid mapping*.

Fringe-fitting is a generalisation of self-calibration as far as the phases are concerned. Unlike self-calibration, in which the closure phases are used to eliminate the instrumental phase terms, in fringe-fitting, we estimate the first-order derivatives of the instrumental phases with respect to frequency (delay) and time (rate) along with the phase offsets to correct the visibilities. Also, because of the limitations imposed by decoherence, corrections that can be performed using a model with fewer degrees of freedom to improve phase coherence and increase the SNR, are made during fringe-fitting.

As such, it is performed before self-calibration.

5.1.3 A RIME-based Probabilistic Approach

The 3GC approach to analysing visibilities treats calibration and imaging as two aspects of the same optimisation problem (section 1.3.3). Propagation path effects are represented using 2 × 2 (full-polarisation) Jones matrices that may vary with time, frequency, and direction (i.e. source). This enables us to introduce the phase offsets, delays, and rates as parts of time and frequency varying Jones matrices. Once we have a RIME-based model, we perform a full Bayesian parameter estimation (and subsequently, model selection) on the data, akin to Chapters 3 and 4. Cornwell & Wilkinson (1981) proposed a Bayesian approach to self-calibration/hybrid mapping and Schwab & Cotton (1983) proposed a similar approach to fringe-fitting by incorporating prior knowledge about the delays and rates. Here we implement just that for fringe-fitting, while also simultaneously solving for source parameters and other instrumental effects, with the help of the RIME.

The dominant source of the residual delays and rates (i.e. imperfections in the correlator model) is the atmosphere. By convention, the phase delays due to the troposphere and the ionosphere are depicted by *T-Jones* and *Z-Jones* matrices respectively (Noordam & Smirnov, 2010). But they are not easily separated during calibration from the G-Jones matrices and hence end up being subsumed by them. Recasting the RIME for multiple discrete sources (equation 1.35) with the time- and frequency-dependence made explicit, and using the G-Jones term to depict the phase-matrix for fringe-fitting, we have that

$$\widetilde{\mathrm{V}}_{pq}(t_m,\nu_n) = G_p(t_m,\nu_n)\left(\sum_s \mathrm{X}^M_{spq}(t_m,\nu_n)\right) G^H_q(t_m,\nu_n). \tag{5.18}$$

Here X^M represents the source coherency matrix for the model used in calibration. The E-Jones terms for primary beams are taken to be unity ($E_{sp} \equiv 1$) owing to the small FoVs often used for VLBI data analysis. With the gain amplitudes set to unity ($|g_p| \equiv 1$) in the context of fringe-fitting, we construct a scalar phase-matrix that affects all polarisations equally:

$$G_p = |g_p| \exp(i[\Delta\psi_p(t,\nu)]) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{5.19}$$

where $\Delta\psi_p$ incorporates the phase offset ψ_{p0}, delay, and rate for antenna p, relative to a reference antenna, time t_{ref}, and frequency ν_{ref}:

$$\begin{aligned}\Delta\psi_p(t_m,\nu_n) &= \psi_{p0} + \frac{\partial\psi_p}{\partial\nu}(\nu_n - \nu_{\mathrm{ref}}) + \frac{\partial\psi_p}{\partial t}(t_m - t_{\mathrm{ref}}) \\ &= \psi_{p0} + \tau_p(\nu_n - \nu_{\mathrm{ref}}) + r_p(t_m - t_{\mathrm{ref}}).\end{aligned} \tag{5.20}$$

As with the global fringe-fitting methods, we use a source model that enables us to separate the antenna-based phase terms from the model visibility phases.

5.1.3.1 SOFTWARE USED

Once the model RIME for fringe-fitting is constructed, the Bayesian analysis proceeds in the same way as in Chapter 4, with MULTINEST being used for the sampling and MONTBLANC for the likelihood computation on the GPU (Figure 4.5). The simulations are performed using MEQSILHOUETTE (Blecher et al., 2017), a software package written specifically to simulate mm-VLBI observations.

MEQSILHOUETTE is written in PYTHON and is based on MEQTREES. MEQTREES can solve for any RIME as long as it is constructed suitably. MEQSILHOUETTE specialises in constructing the appropriate RIME for simulating realistic mm-VLBI observations, by introducing the ability to inject tropospheric effects, interstellar medium (ISM) scattering, and time-variable antenna pointing errors. It accepts a FITS image or a parametrised model consisting of point sources or Gaussians as input models of the sky and the observation parameters including station details, tropospheric effects, and ISM conditions in a user-defined ASCII file. The output is a Measurement Set with the simulated visibilities, along with the closure phases and uncertainties and a dirty and/or deconvolved image (Blecher et al., 2017)[*].

5.2 VERY LONG BASELINE ARRAY SIMULATIONS WITH CONSTANT DELAYS

To test the Bayesian code that incorporates the fringe-fitting parameters, we started off with a simple VLBI simulation using the VLBA [†]. The VLBA is a 10-element interferometer, with the stations located across the United States and operated remotely from Socorro in New Mexico, USA. Each VLBA station is 25 m in diameter. The longest baselines available extend over 8000 km.

We simulated a 60-minute VLBA observation of a point source of flux density 4 Jy, located at the phase centre. There were 32 channels of 31.25 MHz each, centred at a frequency of 43 GHz. The SEFD of VLBA stations at 43 GHz is 1181 Jy[‡]. The integration time used was 10 s. Delays constant in time were introduced using MEQSILHOUETTE, with the centre frequency of the 17th channel chosen as the reference frequency ν_{ref}. The rates were set to zero. Although a constant delay and zero rate over the course of an hour is relatively unphysical, particularly at 43 GHz, these settings were used purely for testing purposes.

5.2.1 BUILDING THE RIME

A point source model at the phase centre was used to generate the model visibilities. Along with the source parameters of flux-density S_ν and position (l, m), I parametrised the residual station-based phase offsets ψ_p[§] and delays τ_p, bringing the number of parameters to 23.

---

[*] Roger Deane provided the MEQSILHOUETTE simulations and performed the subsequent AIPS analyses for comparison with the Bayesian method.

[†] `https://science.nrao.edu/facilities/vlba`

[‡] `https://science.nrao.edu/facilities/vlba/docs/manuals/oss/bands-perf`

[§] We make the dependence of the phase offsets on the reference antenna implicit and drop it from the subscript used in equation (5.20).

The RIME for this analysis is constructed in much the same way as in section 4.4.2, with the exception being the definition of the G-Jones terms. As in equation (4.9), the G-Jones matrix is scalar, but here it incorporates the phase offsets and delays (no rates were added in the simulation), with $|g_p| \equiv 1$. Hence, in the equation

$$\mathrm{V}_{pq} = G_p X_{pq} G_q^H + \mathcal{N}(0, \sigma^2), \tag{5.21}$$

G_p is defined as

$$\begin{aligned} G_p &= g_p \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \text{where} \quad g_p &= \mathrm{e}^{i[\psi_p + \tau_p(\nu_n - \nu_{\text{ref}})]}. \end{aligned} \tag{5.22}$$

5.2.2 PRIORS AND THE LIKELIHOOD

Uninformative (uniform) priors $\Pi(\Theta)$ were set for most parameters of the model (Table 5.1). The position was fixed at the phase centre and given delta priors at (0, 0). The prior range for the flux-density was chosen based on the known value that was used in the simulation. The phase offsets were allowed to vary over the entire 360° range from −180° to 180° and the delays were set uniform priors between −200 and 200 picoseconds (ps). The phase offset and delay for antenna 2 were set delta priors centred at zero, indicating that antenna 2 is the reference relative to which the phases of other antennas were to be calculated.

Table 5.1: Prior distributions for the different parameters used. All the listed parameters were given uniform priors with the range indicated by the values in the square brackets. The parameters with delta priors are not included.

Parameter / units	Prior distribution
S_ν / Jy	[3, 5]
ψ_p, where $p \neq 2$ / degrees (°)	[−180, 180]
τ_p, where $p \neq 2$ / 10^{-12} s	[−200,200]

The likelihood function is set up similarly to the likelihood in section 3.3.1.4 (and not as in section 4.4.4, since the SEFDs do not vary between stations in the present case). Our model assumes that the noise is Gaussian (only random noise was injected into the simulation). Given the observed (V_D) and the modelled (V_M) visibilities, and the uncertainty σ denoting the standard deviation of the noise, derived (equation 1.54) from the theoretical SEFD of VLBA stations, the likelihood function can be written as

$$\begin{aligned} \mathcal{L}(\Theta|\mathrm{V}_D, H) &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\chi^2}{2}\right), \\ \text{where } \chi^2 &= \sum_{k=1}^{N} \left(\frac{\mathrm{V}_{M_k} - \mathrm{V}_{D_k}}{\sigma}\right)^2, \end{aligned} \tag{5.23}$$

and $N = 2N_{\rm vis}$ is the number of data points. The summation is carried out over $2N_{\rm vis}$ since we consider the real and imaginary parts separately. Taking the natural logarithm of $\mathcal{L}$, we obtain

$$\ln(\mathcal{L}) = \ln\left[(2\pi\sigma^2)^{-N/2}\right] - \frac{\chi^2}{2}. \tag{5.24}$$

5.2.3 Analysis and Results

For the Bayesian analysis, I extracted 100 seconds (10 *timestamps*, with the integration time being 10 s) of data from the 60-minute simulated data in order to reduce the processing time, with the *uv*-coverage of this subset shown in Figure 5.3. Since the delays and phase offsets are constant in time, the duration of the subset of visibilities used (100 s) served as the solution interval per phase offset/delay term. To evaluate the 23-parameter model, MultiNest required $\simeq 6 \times 10^5$ likelihood evaluations and the total execution time was $\simeq$ 150 minutes on a Tesla K40M GPU. The marginalised 1-D and 2-D posteriors of ψ_p and τ_p are shown in Figure 5.4. We see that the source flux-density is uncorrelated with the phase offsets and delays, since any phase-related errors affect only the source position. The ψ_p and the τ_p are uncorrelated with each other, indicating that there is no degeneracy between these two classes of parameters. This is what we would expect for phase offsets that are independent of delays. On the other hand, we see that the different phase offsets and delays are slightly correlated between themselves. The missing station indices correspond to the stations for which the source was too low on the sky to observe. The Bayesian estimates are consistent with the AIPS estimates and the true simulated values, with the added advantage that we obtain well-sampled posteriors of the parameters with just 100 seconds of data. Another crucial point to note is that we can estimate the source flux-density simultaneously (no AIPS estimate corresponding to the Bayesian estimate), thereby performing amplitude self-calibration while performing fringe-fitting.

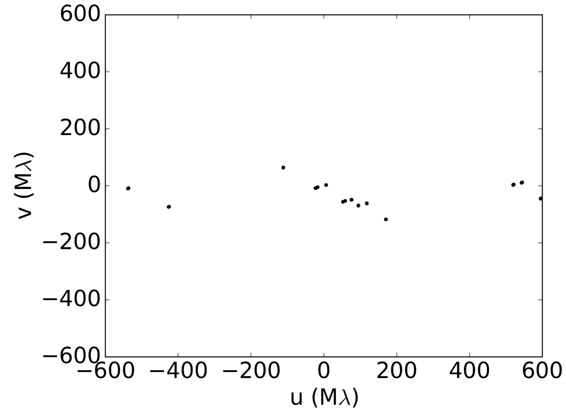


Figure 5.3: The *uv*-coverage of the VLBA 43 GHz simulated data used for the Bayesian analysis. The tracks seem short due to the short timescales involved (100 s).

We now move on to performing Bayesian fringe-fitting at much higher observational frequencies

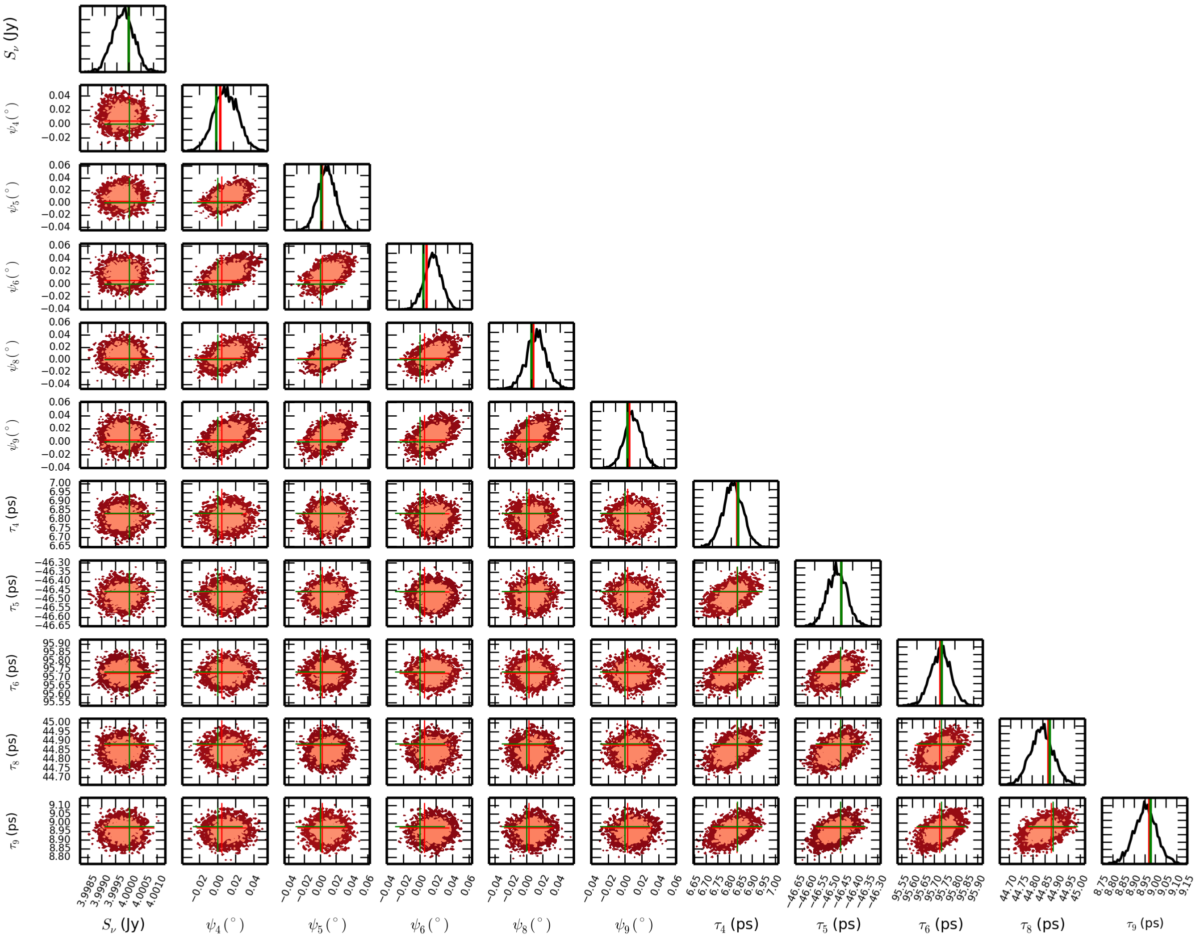


Figure 5.4: The 1-D and 2-D marginalised posteriors of the source flux-density and phase offsets and delays of all stations. Parameters with delta priors and those corresponding to stations that did not participate in the observation are not shown here. The vertical red lines indicate the phase offsets and delays estimated using AIPS. The vertical green lines indicate the true values that were used to simulate the data. There was no AIPS estimate for S_ν and the τ_5 AIPS estimate closely corresponds to the true value.

with an entirely different interferometer that has been set up to conduct high-resolution observations of black holes.

5.3 THE EVENT HORIZON TELESCOPE

The EHT is a network of mm/sub-mm facilities spread across continents to create a telescope with high angular resolution ($\simeq$ 30–10 μas), with the longest baselines spanning the Earth's diameter (Doeleman et al. (2009); see Figure 5.5). The EHT and space-based VLBI experiments provide the highest angular resolution observations in astronomy. The primary goal of the EHT is to image the gravitationally-lensed photon 'ring' or *silhouette* around the event horizon of supermassive black holes[*], for which the most promising candidates are the ones at the centres of the Milky Way (Sgr A*) and the supergiant elliptical galaxy M87. Owing to their proximity, they have the largest predicted apparent angular diameters (50 μas and 20–40 μas respectively) of the known supermassive black hole candidates, and are resolvable by the EHT (e.g. Broderick & Loeb, 2009; Falcke & Markoff, 2013). The telescope operates at a frequency of 230 GHz (a wavelength of 1.3 mm) and plans to extend operations up to 345 GHz (0.8 mm) in the future.

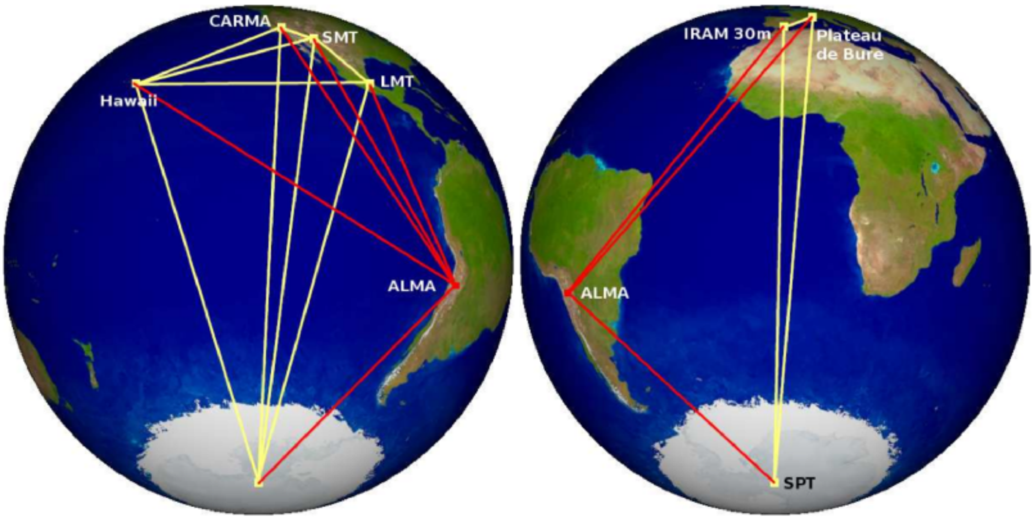


Figure 5.5: The EHT as seen from Sgr A*. The baselines span the Earth's diameter and, operating at frequencies 230 − 345 GHz, can attain angular resolutions of $\simeq$ 30 − 10 μas, comparable to the scales of Sgr A* and M87. The ALMA baselines shown in red possess an order-of-magnitude higher sensitivity compared to the other baselines. CARMA has been decommissioned, while a station in Greenland is being constructed. The stations are located at high altitudes, above as much of the Earth's atmosphere as possible (Image credit: Remo Tilanus).

At such high frequencies, the Earth's troposphere gives rise to a turbulent component to the delays, which is the major contributor to the decoherence of the visibilities. The troposphere is the lowest layer of the atmosphere, at an elevation of 7–10 km above sea level. It is composed primarily of N_2 and O_2, trace gases such as water vapour and CO_2, and particulates like liquid water and dust. As the

---

[*] For a review of supermassive black holes, see Volonteri (2010).

frequency of the electromagnetic radiation increases, so does the tropospheric absorption, mainly due to the pressure-broadened transition lines of H_2O and O_2 (Carilli & Holdaway, 1999). Unlike other chemical components, water vapour mixes poorly in the atmosphere, and varies in both space and time, introducing rapid fluctuations in the measured visibility phases.

The effects of this layer can be modelled using a mean component and a turbulent component. To obtain the phase variation due to the mean component, we introduce a quantity called *precipitable water vapour* (PWV), defined as the depth of the water vapour column if it is converted into liquid phase. Denoting this by $w_{\rm pwv}$, and assuming an atmospheric temperature $T_{\rm atm} \approx 270$ K, we get the following relation for the phase change $\delta\varphi$ experienced by an electromagnetic wave (Carilli & Holdaway, 1999):

$$\delta\varphi = \frac{12.6\pi}{\lambda} \times w_{\rm pwv}, \tag{5.25}$$

where λ is the wavelength of the electromagnetic radiation. Phase instabilities that arise due to turbulence in the troposphere are understood using *Kolmogorov theory*, which describes turbulent motions spanning a wide range of scales, from the scale at which kinetic energy is supplied to the medium, to the much smaller scale at which it is dissipated. This effectively reduces the coherence time to a few seconds ($<$ 10 s), too small for fast-switching (between calibrator and target source) calibration procedures to be effective (e.g. Blecher et al., 2017).

Searching for intrinsic sub-structure in Sgr A* at mm/sub-mm wavelengths is beset with a host of other challenges as well, beyond the contribution from the troposphere. For instance, for an observation carried out for a much longer time than the refractive timescale, the interstellar medium (ISM) induces a Gaussian blurring in the observed intrinsic structure that increases the apparent source size (e.g. Fish et al., 2014). For observations of shorter timescales, additional substructure is introduced by refraction, resulting in non-zero closure phases, which impacts the fringe-fitting process (Johnson & Narayan, 2016). In addition, antenna-pointing inaccuracies at mm and sub-mm wavelengths become significant and must be accounted for[*] in order to disentangle their effects from atmospheric effects and hence recover the source structure.

Under such difficult observing conditions, it is evident that fringe-fitting before any calibration is necessary for eliminating the nuisance phase terms. It would also mitigate effects analogous to those of self-calibration, where early decisions on the initial models used can be frozen in and become impossible to remove further down the track. This is especially true for arrays with few stations. EHT is not only that, but given the small mutual visibility windows between the stations, even fewer are observing simultaneously. All this necessitates more careful approaches to fringe-fitting, particularly in the initial stages. Our probabilistic approach also has the advantage that it can simultaneously estimate source amplitude, structure, and the fringe-fitting phase terms. To study how this method improves on traditional fringe-fitting in the case of the EHT, we analysed three synthetic observations

---

[*] In the RIME formalism, this takes the form of a modification to the E-Jones term describing the primary beam.

— a point source with constant delays (similar to section 5.2), a Gaussian source with constant delays, and a Gaussian source with time-varying delays — using both techniques, which we now proceed to discuss in detail.

5.3.1 CONSTANT-DELAY POINT-SOURCE SIMULATION

We simulated a 3-minute EHT observation of a point source of flux-density 4 Jy at the phase centre at a frequency of 230 GHz. The spectral window was divided into 32 channels of 80 MHz bandwidth each. The integration time was 2 s (Figure 5.6). Six EHT stations participated in the observation (Table 5.2). We also introduced constant-in-time delays using MEQSILHOUETTE, with the frequency corresponding to the 17th channel chosen as $\nu_{\rm ref}$. Our simulation is unrealistic to a certain extent, with constant delays, no rates, and effectively infinite coherence time due to the fact that no tropospheric turbulence was added. There are other contributions to the noise term such as non-zero opacity of the atmosphere and turbulent phases that we do not account for. To test our method in the presence of high thermal noise, we degrade the data quality by scaling up the SEFD of each station by a factor of 10[*]. With one phase offset ψ_p and one delay τ_p per station, along with the source flux-density and position parameters, there were a total of 15 parameters.

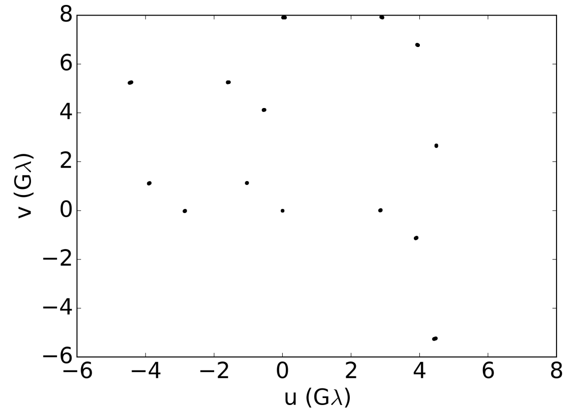


Figure 5.6: The *uv*-coverage of the 3-minute point source EHT simulation.

5.3.1.1 THE RIME, PRIORS, AND THE LIKELIHOOD

The RIME for this analysis is the same as the one that we constructed in section 5.2.1:

$$\begin{aligned} V_{pq} &= G_p X_{pq} G_q^H + \mathcal{N}(0, \sigma^2), \\ \text{where} \quad G_p &= e^{(i[\psi_p + \tau_p(\nu_n - \nu_{\rm ref})])} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{5.26}$$

---

[*] In section 5.3.3 we bring the SEFDs back to their nominal values and add the turbulent components to the phase offsets.

Table 5.2: EHT stations participating in the simulation along with the corresponding (SEFD) values. In the simulation, the SEFDs listed here were scaled up by a factor of 10.

Station	SEFD / Jy
Submillimeter Array (SMA)	4900
Submillimeter Telescope (SMT)	11900
Large Millimeter Telescope (LMT)	560
Atacama Large Millimeter Array (ALMA)	110
James Clerk Maxwell Telescope (JCMT)	5000
South Pole Telescope (SPT)	15000

The centre frequency of the 17th channel was chosen as $\nu_{\rm ref}$. All parameters were set uniform priors, except in the case of source position and parameters corresponding to the reference antenna, for which delta priors centred at zero were used.

The prior ranges used were the same as in the VLBA case, as described in section 5.2.2. The likelihood function is different since the Gaussian noise σ is not common to all baselines. Hence it must take into account the baseline dependence of σ, in the fashion of section 4.4.4. Following the argument there, we write the likelihood function as

$$\begin{aligned} \ln(\mathcal{L}) &= -\sum_{k=1}^{N_{\rm vis}} \ln\left[2\pi\sigma_k^2\right] - \frac{\chi^2}{2}, \\ \text{where } \chi^2 &= \sum_{k=1}^{2N_{\rm vis}} \left(\frac{V_{M_k} - V_{D_k}}{\sigma_k}\right)^2. \end{aligned} \tag{5.27}$$

Here, V_{M_k} and V_{D_k} are the modelled and observed (i.e. simulated) visibilities respectively, $N_{\rm vis}$ is the number of visibilities[*], and σ_k depends on baseline.

5.3.1.2 ANALYSIS AND RESULTS

The total computation time was about 2 hours on a Tesla K40M GPU, with MULTINEST requiring 2×10^5 evaluations of the likelihood to converge. A solution interval of 3 minutes i.e. the duration of the observation, was used since the delays and phase offsets are constant in time. The 1-D and 2-D marginalised posteriors of the parameters of the model are displayed in Figure 5.7. As with the case of VLBA, we see here that the flux-density is uncorrelated with the phase offsets and delays. The ψ_p and the τ_p are also uncorrelated with each other, indicating that they are not degenerate. The phase offsets and delays corresponding to different stations are correlated between themselves, but are estimated with high precision, with the phase offsets estimated to within a degree and the delays estimated to within a few picoseconds of the true values. The peaks of the posteriors are much closer

---

[*] The χ^2 is obtained over $2N_{\rm vis}$ since the real and imaginary parts of the visibilities are considered separately.

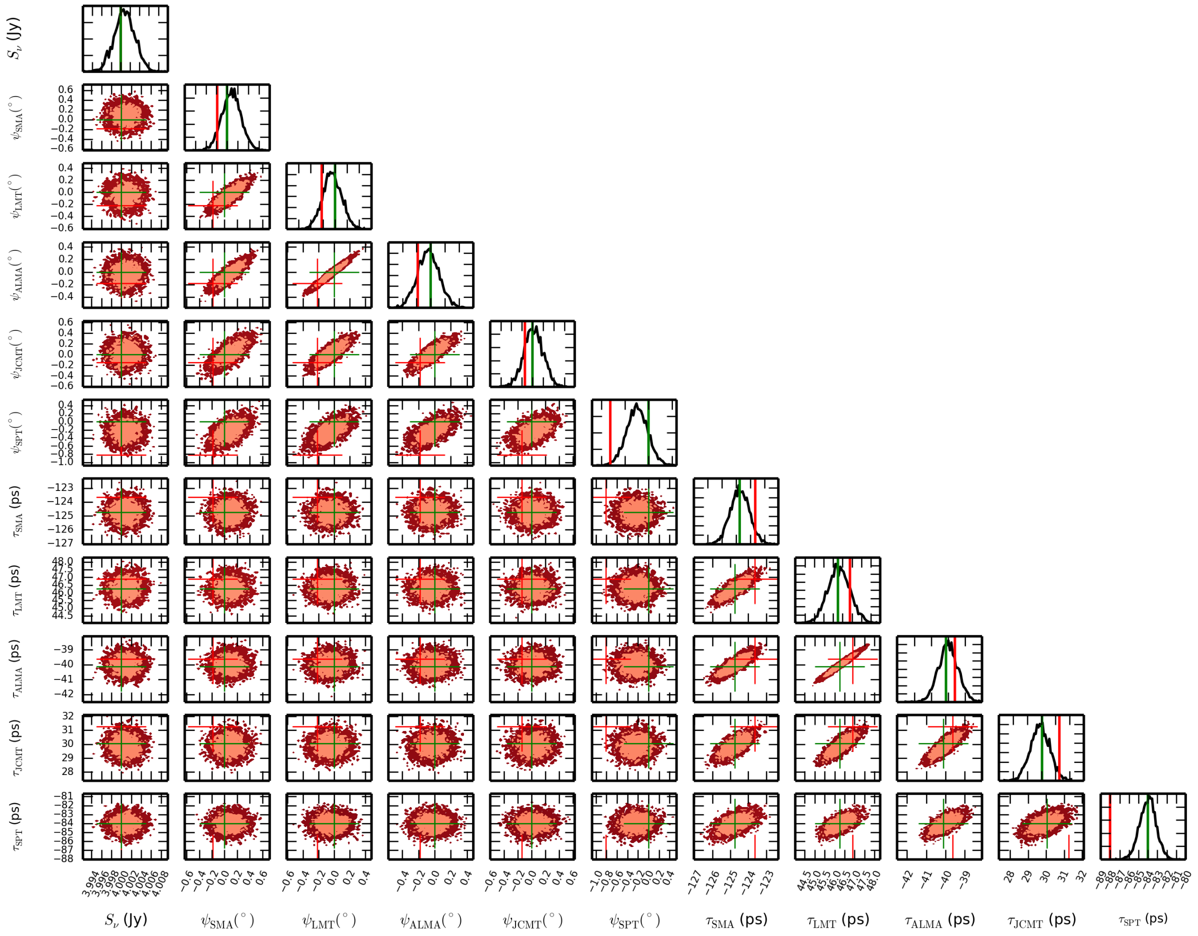


Figure 5.7: The 1-D and 2-D marginalised posteriors of the source flux-density and phase offsets and delays of all EHT stations, except the ones that were set delta priors. The vertical red lines indicate the phase offsets and delays estimated using AIPS. The vertical green lines indicate the true values that were used to simulate the data.

to the true values than the corresponding AIPS estimates, with some AIPS solutions falling almost outside the posteriors, such as the estimate for ψ_{SPT}. The delay estimates obtained by the fringe-fitter in CASA are the same as the AIPS estimates within the uncertainty. Thus, the Bayesian method performs better when compared to traditional fringe-fitting methods, while enabling us to estimate source flux-density simultaneously.

5.3.2 Constant-Delay Extended Source Simulation

The previous two analyses involved a simulated point source located at the phase centre, with phase offsets and constant delays injected into the simulation. In reality, there are hardly any unresolved sources in VLBI; especially in the case of EHT, whose goal is to study the resolved or partially-resolved black hole silhouettes, simple point source models will not suffice. Traditionally, estimation of the source structure is not performed during fringe-fitting, although the phase instabilities corrected for during fringe-fitting contribute towards improving the structure estimation later on. This was apparent in the EVN observation analysed in Chapter 4, where preliminary fringe-fitting (with point estimates for phase offsets, delays, and rates already applied) had been performed on the data. In mm-VLBI, a more general approach would be to estimate the source structure and the fringe-fitting parameters simultaneously. Since our Bayesian approach uses the RIME, it should be able to estimate partially-resolved source structure during fringe-fitting. To this end, we simulated an extended source at the phase centre, defined with a Gaussian profile, and introduced phase offsets and constant delays into the simulation.

The simulation involved observing for 3 minutes a source with a circular Gaussian profile of HPBW 5 μas. With the PSF of EHT being 20 μas, the source is one-fourth the size of the PSF. The spectral window was divided into 32 channels of 80 MHz bandwidth each and the integration time was 2 s (Figure 5.8) as before. Six EHT stations took part in the observation (Table 5.2). The SEFDs displayed in the table were scaled up by an order of magnitude (in the manner of section 5.3.1) to create a dataset

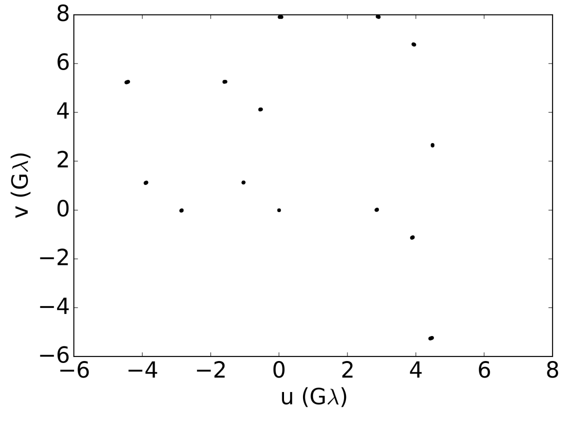


Figure 5.8: The *uv*-coverage of the constant-delay EHT simulation with a Gaussian source.

with a more realistic SNR. Delays constant in time were introduced using MeqSilhouette, with the frequency corresponding to the 17th channel chosen as $\nu_{\rm ref}$. Though constant delays are unrealistic for mm-VLBI observations, here I focus on simultaneous estimation of source structure and the fringe-fitting terms. A more realistic simulation is carried out in section 5.3.3.

Our goal is not only to estimate the shape of the Gaussian but also to perform model selection between a point source (PT) and a circular Gaussian (CIRC) model. The point source model is the same as the ones used in sections 5.2 and 5.3.1. The circular Gaussian model CIRC, has two additional parameters that describe the shape of the Gaussian:

$$\begin{aligned} l_p &= e_{\rm maj}\ \sin\vartheta, \\ m_p &= e_{\rm maj}\ \cos\vartheta. \end{aligned} \tag{5.28}$$

The third parameter, the ratio of the minor axis to the major axis (r), seen in equations (3.7) and (4.4), is set to unity since the Gaussian is circular.

5.3.2.1 THE RIME, PRIORS, AND THE LIKELIHOOD

The RIMEs for the two models are identical to the ones constructed in section 4.4.2. The RIME for PT is given by

$$\mathrm{V}_{pq} = G_p\, X_{pq}\, G_q^H + \mathcal{N}(0, \sigma_{pq}^2), \tag{5.29}$$

where $\mathcal{N}$ denotes the additive Gaussian noise term with zero mean and a variance of σ_{pq}^2 per visibility. For the model CIRC, the summation becomes an integral performed over the extent of the source. Expressed in terms of the direction cosines l and m, the RIME becomes (Thompson et al., 2017, section 3.1)

$$\begin{aligned} \mathrm{V}_{pq} &= \iint\limits_{lm} G_p\, X_{pq}(l,m)\, G_q^H\, \mathrm{d}\Omega\ +\ \mathcal{N}(0, \sigma_{pq}^2), \\ \text{where } \mathrm{d}\Omega &= \frac{\mathrm{d}l\,\mathrm{d}m}{\sqrt{1 - l^2 - m^2}}. \end{aligned} \tag{5.30}$$

The G_p terms are given by

$$G_p = \mathrm{e}^{i[\psi_p + \tau_p(\nu_n - \nu_{\rm ref})]} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{5.31}$$

The (uninformative) priors $\Pi(\Theta)$ for the various parameters of PT and CIRC are shown in Table 5.3. The position was fixed at the phase centre and the prior range for the flux-density was chosen based on the known value used in the simulation. The priors for l_p and m_p were set to vary over 20 μas, the size of the PSF of the EHT. As before (section 3.3.1.4) the prior for l_p is restricted to non-negative values so that the position angle estimate is constrained to a range of $0 - 180^\circ$. The phase offsets were allowed to vary over the entire 360° range from -180° to 180° and the delays were given uniform priors between -200 and 200 ps. The phase offset and delay for antenna 2 (SMT) were given

Table 5.3: Prior distributions for the different parameters of the models PT and CIRC. All the listed parameters were set uniform priors with the range indicated by the values in the square brackets. l_p and m_p apply only to CIRC. The parameters with delta priors are excluded.

Parameter / units	Prior distribution
S_ν / Jy	$[3, 5]$
l_p / μas	$[0, 20]$
m_p / μas	$[-20, 20]$
ψ_p, where $p \neq 2$ / degrees (°)	$[-180, 180]$
τ_p, where $p \neq 2$ / 10^{-12} s	$[-200,200]$

delta priors at zero, since SMT is the reference relative to which the phases of other antennas will be calculated.

The likelihood function assumes Gaussian error distribution and is given by equation (5.27).

5.3.2.2 Analysis and Results

Adding the two shape parameters increased the execution time to about 15 hours and MultiNest performed 1.9×10^6 likelihood evaluations before reaching convergence. The solution interval for the analysis was again set to 3 minutes, the entire duration of the observation.

The relative logarithmic evidence between CIRC and PT is 1690 ± 0.7, indicating *very strong* evidence in favour of CIRC, computed according to Jeffreys' criterion (Table 2.1). The marginalised posteriors of CIRC are shown in Figure 5.9.

We see that, here too, the Bayesian estimates correspond more closely to the true values than the AIPS estimates, with some AIPS estimates inconsistent with the true values and the Bayesian estimates. The ψ_p and τ_p posteriors are uncorrelated with each other, and the introduction of source shape parameters has not affected the estimation of the phase-related terms. The size parameter e_{maj} ($e_{\mathrm{maj}} = e_{\mathrm{min}}$ since the Gaussian is circular) is estimated with high precision (10^{-7} arcsec), and is uncorrelated with the phase offsets and delays.

5.3.3 Extended Source with Time-variable Delays

So far we have analysed only those simulations with constant delays introduced into them. In a realistic scenario, as the atmospheric conditions over the stations vary with time, the geometric path lengths from the source to the stations vary, and the corresponding delay terms evolve too. A solution interval that encompasses the whole observation will tend to smear out these variations and result in sub-optimal fringe-fitting and decoherence. Moreover, the presence of turbulence-induced delays/phases, which we have ignored so far in our simulations, also contributes to decoherence, and must be accounted for using shorter solution intervals for delay estimation. In fact, at mm wavelengths, this is the predominant source of decoherence (Carilli & Holdaway, 1999). To capture the time variation in the delays with reasonable accuracy, we must use a solution interval that is much

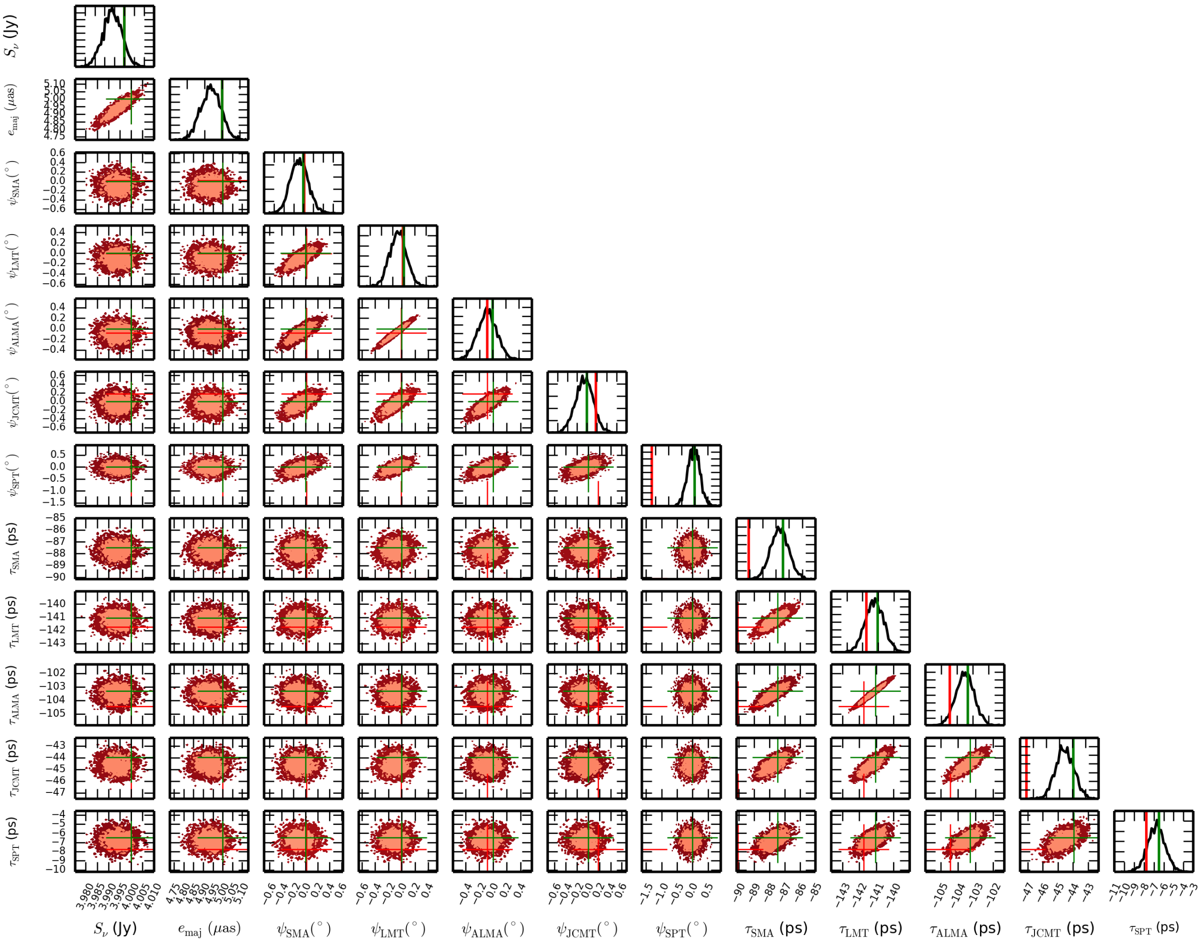


Figure 5.9: The 1-D and 2-D marginalised posteriors of the flux-density, size, and phase offsets and delays of all EHT stations, except the ones for which delta priors were set. The vertical red lines indicate the phase offsets and delays estimated using AIPS. The vertical green lines indicate the true values that were used to simulate the data. The flux-density and the shape estimates are not available for AIPS, since AIPS assumes a point source model.

shorter than this coherence time. Realistically, for very short solution intervals, the SNR of the data would be too low for us to draw reliable inferences. To bring the SNR up to a reasonable level, the data are often averaged over a larger interval, at the cost of decoherence.

As our final experiment, we would like to know whether our method can determine the time-varying delays and phase offsets, with turbulent phases of coherence time 180 s injected, while simultaneously performing model selection between PT and CIRC. We simulated an EHT observation of a Gaussian source with time-varying τ_p and ψ_p injected. A circular Gaussian source of HPBW 5 μas (25 per cent of the PSF) was observed for 10 minutes with an integration time of 20 s. The coherence time of the turbulent component of the delay was 180 s. The choice of integration time was a compromise between the need to have short enough integration times while maintaining a reasonable SNR. The spectral window was divided into 32 channels of 80 MHz bandwidth each (Figure 5.10). Six stations participated in the observation. Unlike the other EHT simulations, here the SEFDs displayed in Table 5.2 were not scaled up since we introduce time-varying and turbulent delay terms. The frequency corresponding to the 17th channel was chosen as ν_{ref}.

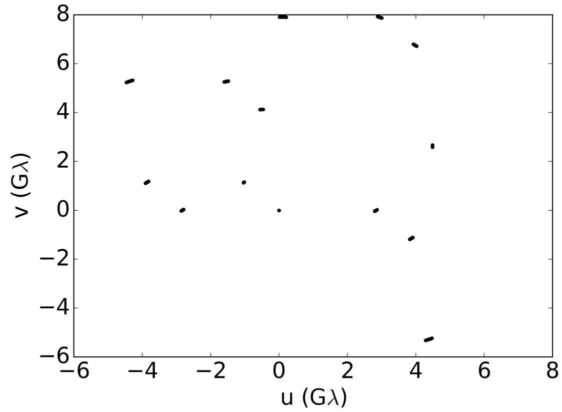


Figure 5.10: The *uv*-coverage of the EHT observation of a Gaussian source with time-varying delays. The tracks here are noticeably longer since the observation was carried out for 10 minutes.

5.3.3.1 The RIME, Priors, and the Likelihood

We may estimate the time-varying delays in two ways. Since the time-variability can be incorporated into time-varying Jones matrices in the RIME, each delay may be represented as a time-varying polynomial and its coefficients estimated. This would increase the dimensionality of the parameter space, incurring a high cost in running time. An alternative is to perform independent analyses of subsets of the simulation, with the data divided into segments of duration equal to that of the solution interval. We adopt the second method to reduce the execution time. However, we note the potential of the first method in future work.

The RIMEs corresponding to PT and CIRC for these analyses are given by equations (5.29)–(5.31). The priors used were also identical to the ones described in Table 5.3 in section 5.3.2.1. The likelihood function with baseline-dependent per-visibility uncertainty σ_{pq} is given by equation (5.27).

5.3.3.2 Analysis and Results

The 10-minute observation was divided into 28 segments of 20 s duration each[*] and each segment was analysed independently. With the solution interval (20 s) being shorter compared to that used in the previous analyses (180 s), the SNR for each subset was lower and MultiNest required about 6×10^6 likelihood evaluations to converge. The execution time for each subset was at least 30 hours, with a few subsets requiring up to 40 hours before they finished successfully.

The relative Bayesian logarithmic evidences between PT and CIRC for all data segments are *very strongly* (Table 2.1) in favour of CIRC, of the order of $10^4 : 1$ with the error not exceeding 0.9. The posteriors for the source parameters for the first data segment are shown in Figure 5.11. As with the circular Gaussian simulation with constant delays, our source parameter estimates are consistent with the true values.

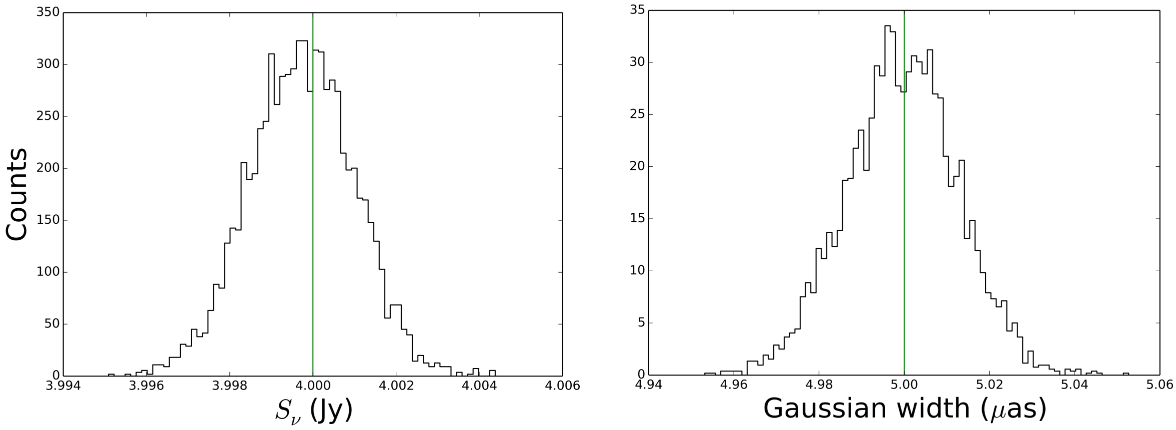


Figure 5.11: The posteriors for the source flux density and the HPBW of the circular Gaussian source for data segment 1.

Unlike the previous experiments, here I marginalise over the phase offsets and present only the source parameters and the delays, since the turbulent phase offsets added by MeqSilhouette are not accessible[†]. The 1-D and 2-D marginalised posteriors of the delays for data segment 1 are displayed in Figure 5.12. From the figure, we see that introducing turbulence into the simulation results in larger uncertainties (10–20 per cent) in the delay estimates of both the Bayesian and the traditional AIPS methods. No turbulence was introduced in the simulations discussed in sections 5.3.1 and 5.3.2, and hence the coherence time could effectively be considered infinite. Using a solution interval of 180 s did not affect the delay estimates in any way. In the present case, the solution interval (20 s) is a substantial fraction of the coherence time of the turbulence in the troposphere (180 s) and its effects

---

[*] Two data segments for which the sampler did not converge were excluded. This is a technical issue likely related to the sampler settings, which we have seen happen in the past. Using MultiNest with different settings or cross-checking with a different sampler will be necessary before we could include these data segments in the analysis.

[†] This feature is expected to be incorporated in a future version.

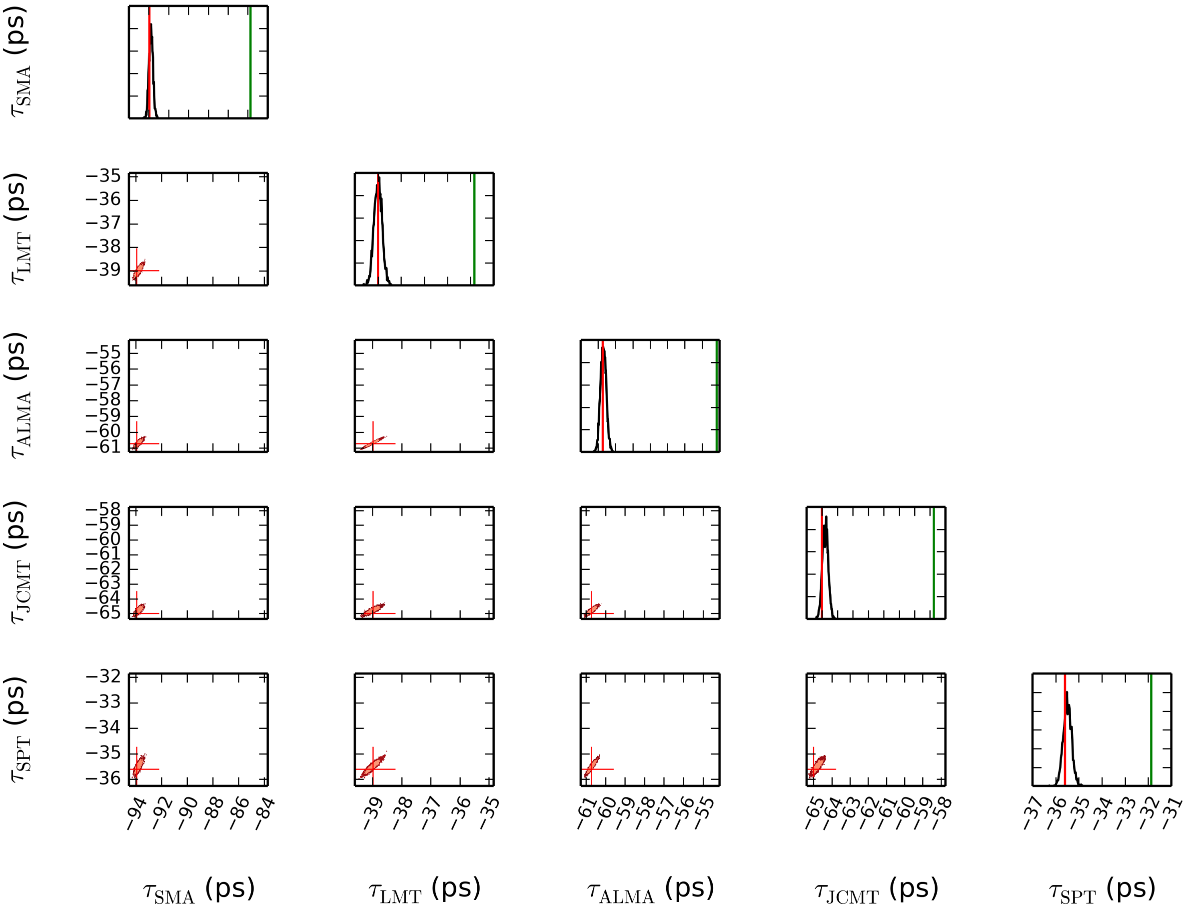


Figure 5.12: The 1-D and 2-D marginalised posteriors of delays for data segment 1. Red lines indicate AIPS estimates and green lines indicate true values used in the simulation.

do not average out to zero, as the case would be in a normal EHT observation. This results in the observed offsets in the delay estimates. Figures 5.13 and 5.14 show the delay estimates at 140 s intervals until the end of the observation, with the AIPS estimates and the true values indicated. As the delays evolve in time, our estimates of the delays do too, while retaining the systematic offsets from the true values.

To obtain a measure of the magnitude of the systematic errors in our delay estimates, we calculated the percentage by which the delays obtained by both the Bayesian method and AIPS are offset from the true values (Figure 5.15). I used the MAP values of the posteriors of the delays from all the data segments for this comparison. Presenting a point estimate such as the MAP value defeats the purpose of performing a full Bayesian analysis that returns all possible realisations of the parameters that constitute a model. Nevertheless, we use it as a metric that can easily be compared with the point estimates of AIPS (which, incidentally, do not have associated uncertainties) in computing the percentage error in the delay estimates. We see that both AIPS and the Bayesian method perform similarly, with the errors spread over 10–16 per cent of the true delays. However, with our method, the RIME enables the source structure to be estimated simultaneously with the fringe-fitting parameters and the probabilistic formulation of the problem enables us to obtain realistic uncertainties. These uncertainties are available to be propagated forward in the subsequent scientific analysis (e.g. measuring the asymmetry in the shadow of a black hole).

5.4 CONCLUSIONS AND FUTURE WORK

The work presented in this chapter is an attempt to perform complete Bayesian inference on VLBI observations, from fringe-fitting to estimating the parameters relevant to the scientific problem at hand. As such, it seeks to extend the Bayesian inference method applied to the EVN observation of the blazar CGRaBS J0809+5341 presented in Chapter 4, to incorporate fringe-fitting, an early step in the calibration of VLBI data. In Chapter 4, fringe-fitting was performed before we performed a probabilistic analysis of the visibilities; here I have formulated a RIME-based fringe-fitting algorithm and performed full Bayesian inference to estimate the phase offsets and their time- and frequency-dependent derivatives.

We see that the probabilistic method performs at least as well as or, in some cases, better than the conventional fringe-fitting algorithms implemented in AIPS, despite the large parameter space that it has to handle. The aim here is to test whether the probabilistic fringe-fitting method works rather than compare its demands on computational resources with those of AIPS. Although software such as AIPS can account for more complex source models than a point source model and return better best-fit estimates for fringe-fitting parameters in a fraction of the time required by the probabilistic analysis, this involves considerable iterative human interaction which introduces untracked subjective judgements. In contrast, the probabilistic approach eliminates the need for human intervention and more importantly, estimates the posterior distributions of both the source and the instrumental

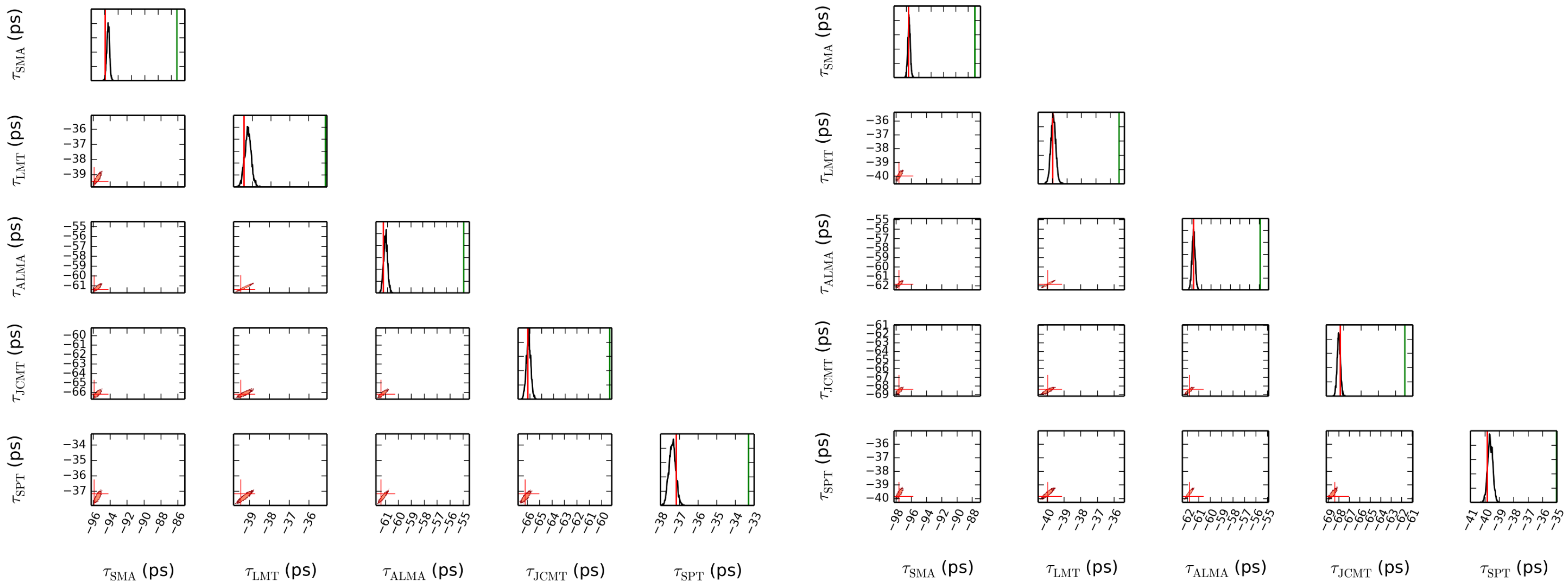


Figure 5.13: The 1-D and 2-D marginalised posteriors of delays for data segments 7 and 14, separated by an interval of 140 s.

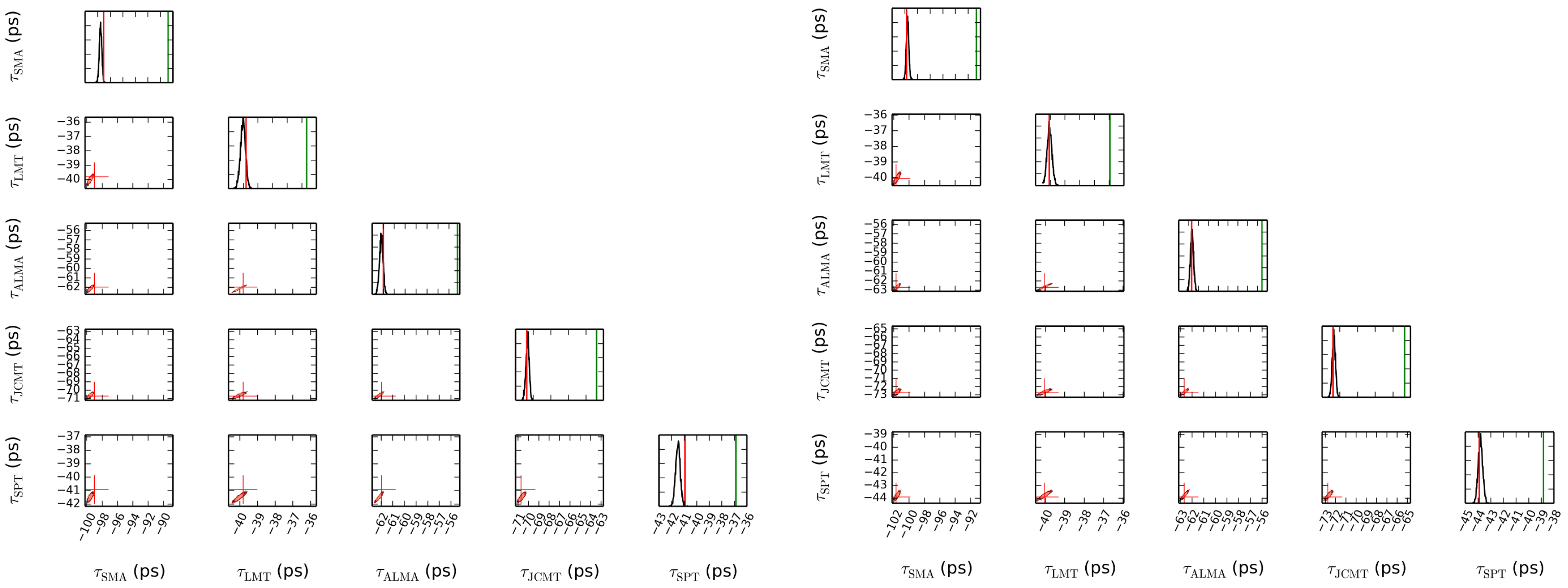


Figure 5.14: The 1-D and 2-D marginalised posteriors of delays for data segments 21 and 28, separated by an interval of 140 s.

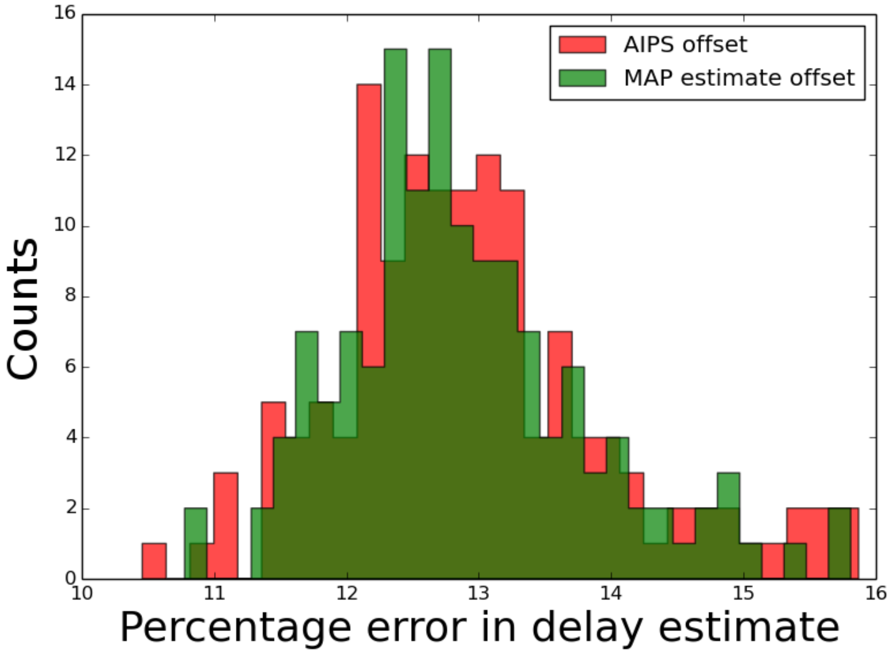


Figure 5.15: The percentage error in the delay estimates obtained by the Bayesian method and AIPS.

parameters simultaneously, which is a central objective of this work.

We started off with a simple simulation with constant delays introduced into a 43 GHz VLBA observation of a point source at the phase centre. We then moved on to a 230 GHz EHT simulation of a point source with constant delays. In both cases, we were able to estimate delays consistent with simulated values, along with the source flux-density. The primary objective of the EHT is to image the shadows or silhouettes of the supermassive black holes in the Milky Way galaxy (Sgr A*) and M87. To test whether such resolved source structures can be estimated using the Bayesian method, we analysed a synthetic observation of a partially-resolved Gaussian source with constant delays. Model selection between point source and Gaussian source models returned very strong Bayesian evidence in favour of the Gaussian source and accurate estimates of the source flux-density, size, and the delays. Moving on, we analysed a synthetic observation of a partially-resolved Gaussian source (25 per cent of the size of the PSF) with time-varying delays and turbulent tropospheric effects introduced. This results in the delay estimation being 16 per cent off from the true values of the time-varying delays introduced in the simulation, due to the short coherence times of the turbulence in the troposphere. Choosing a shorter solution interval at the cost of increasing the execution time can lead to more accurate estimates of the delays.

5.4.1 Future Work

Complete fringe-fitting involves the estimation of not only the phase offsets and the delays but also the time-derivative or the rate residual that appears in equation (5.20). We have so far refrained from estimating the rate residuals to avoid increasing the dimensionality of the parameter space, thereby keeping the execution time more manageable, since MultiNest performs well with models with 30 parameters or fewer. PolyChord, touted as a natural successor to MultiNest, can handle models with few hundreds of parameters (Handley et al., 2015a,b) and will replace MultiNest in our software in the future. Future versions of Montblanc will be able to employ multiple GPUs in modelling the RIME, which will also bring down the likelihood evaluation time.

Another simplification we have made in our simulations is that they all contain one spectral window (frequency band) each. Real VLBI observations are carried out over multiple spectral windows that may be either contiguous or widely separated in frequency. Since the different frequency bands often have at least partially independent receiver electronics associated with them, this complicates the fringe-fitting process (Cotton, 1995). To handle this effectively, we must introduce additional phase offsets that are independent of each other in frequency as well as the multi-band delay, in the Jones matrix for fringe-fitting. This would scale the number of fringe-fitting parameters by the number of spectral windows used and hence provides another strong case for implementing faster samplers and RIME computation.

For low frequency observations with wide fractional bandwidths such as the SKA at long baselines, the dispersion due to the ionosphere become significant. Modelling the effects of the ionosphere is, in principle, similar to modelling the troposphere and MeqSilhouette aims to eventually include this capability. Polarisation leakage between the two feeds is also significant at these frequencies, and hence modelling the polarisation dependence of the fringe-fitting parameters is an important step towards analysing source polarisation.

Varying solution intervals is also necessary to determine the correct trade-off point between faster execution and accurate fringe-fitting. Currently, the software does not have the capability to parametrise the solution interval and so this must be done by running multiple (preferably parallel) instances of the software on the same data with different solution intervals. Introducing the ability to parametrise the solution interval is a long-term goal for this project.

Finally, VLBI observations often resolve source sub-structures that are better modelled with more complicated models than a single point or Gaussian source. As part of the ongoing work on this project, we intend to use combinations of multiple point and Gaussian sources to describe the source structure more thoroughly, for which accurate fringe-fitting would be necessary. The steps outlined in the last few paragraphs will go a long way towards incorporating these complex models. It is also a necessary step before real EHT observations can be analysed using this technique.

Typical EHT observations last for 12 hours with a time resolution of 1 second and 1 MHz wide frequency channels. The data sizes are usually on the order of 100 GB. The software and hardware setup used in this chapter would take weeks to analyse one such EHT dataset. However, it must be noted that on larger servers, with the software taking advantage of multiple GPUs and the implementation of techniques such as non-linear Kalman filters for time-varying gain calibration, the execution time can be sped up dramatically. Also, the method would be best suited to analyse carefully designed experiments in which certain constraints can be reasonably imposed on fringe-fitting parameters. So, while a probabilistic approach to this problem is highly specialised and requires high performance hardware, it is well worth pursuing for high impact experiments such as the EHT.

6

So Far and Beyond

Mark Twain, with characteristic wit, was cautioning against the misuse of statistics when he proclaimed, “Figures often beguile me, particularly when I have the arranging of them myself”*. In the orthodox (or frequentist) interpretation, probability is conceived of as the long-run relative frequency of occurrence of an event given a large number of trials. This confines the problems to which probability theory can be applied to those that can be formulated in terms of random variables. In contrast, when formulated as an extension of logic, probability theory has widened applicability and allows us to draw plausible inferences about logical propositions. Regardless of which interpretation of probability theory (or statistical inference) one subscribes to, one must be judicious about formulating the scientific problem under consideration and the assumptions made during inference. Probability theory as extended logic forces us to make our assumptions (or biases) explicit and to quantify them, thereby making it easier to correct ourselves when we go astray in our reasoning.

The motivation behind this thesis was to explore probabilistic methods for analysing visibility data from radio interferometers. The probabilistic method provides a rigorous way of comparing hypotheses and estimating uncertainties in the model parameters, which are then available for forward propagation in the ensuing scientific analysis. The formulation of the generic radio interferometry measurement equation and its subsequent implementation in software packages such as MeqTrees and Montblanc has provided us with an efficient method of modelling interferometric observations. The code I use for performing the Bayesian analysis has evolved over time, first using the slower, but more generic, MeqTrees and later using the GPU-accelerated Montblanc as we move on to specialised VLBI applications. For the numerical sampling, we used MultiNest, a Bayesian infer-

---

[*] This is the context in which the oft-misquoted line, “There are three kinds of lies: lies, damned lines, and statistics” appears! The excerpts are from Twain’s *Chapters from My Autobiography* (1907).

ence tool for calculating the evidence and computing multi-modal posteriors. The likelihood functions and prior distributions are constructed within our software and passed on to the Bayesian sampler. These functions are problem-specific and are explained at the appropriate places in Chapters 3, 4, and 5.

6.1 Conclusions

The major conclusion that one should take away from this thesis is that it is possible to perform a complete probabilistic analysis of visibilities and estimate the source and instrumental effects simultaneously, by exploiting the power of the RIME in modelling interferometric observations. Let us now summarise the main inferences that we can draw from the results of specific applications.

6.1.1 Super-resolution

In Chapter 3 I sought to apply Bayesian inference to the analysis of visibility data. I described the theory behind using the RIME for this probabilistic approach and presented the implementation details of the software. I started by simulating simple WSRT observations of three source models — a point source, a Gaussian source, and two point sources — with structures at one-fifths the scale of the PSF at 1.4 GHz, in the presence of random noise. The SNR of the simulations was 10^3:1. I then estimated the parameters and the evidences for each model against each simulated dataset. The resulting evidences clearly indicate that our method can obtain *very strong* evidences according to Jeffreys' criterion (Table 2.1) for the correct source model in each case. The source flux-density, position, and, in the case of the Gaussian source, its shape parameters, are all estimated consistent with the inputs, with the shape and position estimates being precise to the order of tens of milli-arcseconds, for a PSF of 16 arcsec.

How Noise Limits Super-resolution

A crucial factor in estimating source structure beyond the diffraction limit is the per-visibility noise. To study the effect of noise on the super-resolution experiment, I simulated Gaussian sources of varying sizes, with the same total flux-density (0.85 Jy) and five different per-visibility noise levels: 0.1, 0.2, 0.3, 0.5, and 1.0 Jy. I then performed model selection between point and extended Gaussian source models on each one of those datasets. We see that, with increasing visibility noise, the size of the extended source that can be confidently estimated increases too. This gives us a practical way of computing the super-resolution limit of our Bayesian approach in the presence of simulated random noise for a given interferometer.

6.1.2 Estimating the Brightness Temperature of a Blazar

In Chapter 4, I applied our method to a real interferometric observation for the first time. The blazar CGRaBS J0809+5341 was observed using the EVN at a frequency of 5 GHz, after an optical flare was observed emanating from it. After conventional imaging methods failed to observe any elongated structure which might indicate the presence of a jet (An et al., 2016), I analysed the same data using the probabilistic approach, by performing model selection between a point source and an extended Gaussian jet structure.

One of the stations used in the observation, which was situated in Shanghai (SH), did not have any short baselines associated with it. As a result, preliminary self-calibration of the data could not constrain its gains to a reasonable degree of accuracy. Hence, we had to take into account the amplitude $|g|$ of the station gains in our models and let them vary by 20 per cent. The Bayes factors between the two models indicate that there is *very strong* evidence in favour of the presence of the jet, in accordance with what would be expected of a flaring blazar. We find a 12 per cent variation in the SH gains, about an order of magnitude worse than that for the other stations. The estimated size of the Gaussian is less than 500 milli-arcseconds (mas) (the PSF of the observation, when naturally-weighted, is 5.7×2.2 mas), with the minor axis being poorly constrained. This is due to the dependence of the shape parameters on $|g_{\mathrm{SH}}|$. With the introduction of a few short baselines for SH, the gains and the shape estimates would improve.

An important consequence of the Bayesian approach is that the uncertainties on the model parameters are available for propagation down the line. We use the uncertainties on the shape parameters and the flux-density estimates to compute the brightness temperature T_b of the blazar (Figure 4.13), the 68 per cent credible region of which lies within the inverse Compton limit, between 0.17×10^{12} K and 0.36×10^{12} K.

Maximum Observable Brightness Temperature

I also derived the maximum observable brightness temperature with any interferometer for a given SNR, by computing a Bayesian criterion for the resolution limit. Martí-Vidal et al. (2012), in deriving a theoretical limit for the maximum super-resolving power of an interferometer, dependent on the SNR of the data, assumed perfect calibration. In reality, uncertainties in station gain estimates and their correlation with source shape estimates will limit the minimum resolvable source size.

I simulated EVN observations of Gaussian sources of varying sizes, with visibility noise that varies with baseline, typical of VLBI observations, and perform model selection between point and extended source models with instrumental gains introduced. Setting the criterion for minimum resolvable source size as the relative evidence between the two source models at which the evidence for the Gaussian model becomes *positive* (Table 2.1), we find that, as the SNR improves ($\mathrm{SNR} \gg 100$), the effects of miscalibration of the station gains become more pronounced and the discrepancy between the theoretical limit and the practical limit obtained increases to a factor of 3. This knowledge enables us to obtain the maximum brightness temperature that can be measured by a given interferometer.

Such experiments will help in setting a practical limit to the super-resolution capabilities of existing and future VLBI arrays such as the SKA−VLBI, where resolution of source structure down to a few tens of μas with high precision can be reasonably expected. They will also help in obtaining a lower limit on how accurately the gains of stations providing the long baselines can be estimated.

6.1.3 FRINGE-FITTING

The analysis of the EVN observation in Chapter 4 was performed on data that had been previously self-calibrated. In Chapter 5, we broaden the domain of our method to include fringe-fitting, the process of solving for phase variability in time and frequency introduced by the intervening medium from the source to the telescope, such as the Earth's troposphere. We formulate a RIME-based fringe-fitting approach and include the fringe-fitting terms in our models, along with the source parameters. This ability to simultaneously estimate the source parameters and the fringe-fitting phase terms and marginalise over the unwanted parameters is the main motivation for incorporating fringe-fitting in the Bayesian method.

We simulate a 43 GHz VLBA observation and a 230 GHz EHT observation, both of point sources with constant delays and phase offsets introduced. In both cases, we are able to determine the delays and phase offsets that are consistent with the inputs. Following this, we simulate a 5 μas Gaussian source (one-fourth the size of the PSF), with the delays and phase offsets introduced in the 230 GHz EHT simulations and perform model selection between point and Gaussian source models as before. We find that partially-resolved source structure can also be estimated along with the fringe-fitting terms, with *very strong* evidence in favour of the Gaussian against a point source model. The delay estimates we obtain are better than those obtained by conventional algorithms implemented in AIPS and CASA. Moreover, we can estimate the source flux-density and shape simultaneously during fringe-fitting and marginalise over the delays and phase offsets which naturally provides the uncertainties on the source parameters.

ESTIMATING TIME-VARIABLE DELAYS

We also simulate a more realistic case of a Gaussian source observed for 10 minutes, with turbulent tropospheric delays introduced. These delays vary with time and hence the time interval over which a solution is obtained cannot be the same as the total observation time. We obtain a delay and a phase offset solution every 20 seconds, with the data split into 20-second subsets and analysed independently. In each solution interval, we obtain *very strong* evidence in favour of the resolved source model and estimate the source flux-density, shape, and the delays for the Gaussian model. The addition of a turbulent component to the delay terms results in the delay estimates being about 10–16 per cent off from the true values. A shorter solution interval at the cost of increasing the execution time will lead to more accurate estimates of the delays.

6.2 OUTLOOK

We began the thesis with a mathematical formulation of how probability theory can be employed for drawing inferences in radio interferometry. The analyses of the synthetic observations and the EVN observation of the blazar CGRaBS J0809+5341 showcase the potential of this approach. More improvements are forthcoming to the software. Let us take a brief look at what problems we can currently apply our method to before discussing the challenges we would face in expanding its horizons.

6.2.1 APPLICATIONS

With the simulation capabilities of MeqTrees and Montblanc available to us, we could perform more tests of various physical phenomena and instrumental effects that can be captured using the RIME. We have performed preliminary tests, similar to the ones described in sections 3.3.2 and 4.6 for studying the impact of SNR on the super-resolution capabilities of different interferometers, including the EHT, and future ones such as the SKA−VLBI. We have also performed some preliminary tests on estimating the polarisation properties of sources in the presence of baseline-dependent noise. The initial results look promising although we need to perform more complete tests with more realistic instrumental effects introduced.

A primary objective of the project discussed in Chapter 5 is to test the applicability of our method to the analysis of EHT observations of black holes. We have been able to incorporate the phase offsets and delays, two important components of fringe-fitting, while simultaneously estimating source amplitude and structure. To perform complete fringe-fitting, we plan to incorporate rates, the third term that captures the remaining phase-variations that afflict VLBI observations. Accounting for both the delays and the phases, capturing the phase-variations in both time and frequency for multiband datasets, is the immediate next step in the project. Factors that give rise to the above terms (such as turbulence in the Earth's troposphere and scattering by the ISM) will also be incorporated in subsequent analyses.

Astrometric studies using VLBI require precise and accurate estimation of source positions. Current methods can reach a precision of $\approx$ 10 μas, whereas we have been able to achieve just a few μas uncertainty in position measurements for the EVN observation of CGRaBS J0809+5341 and a precision of 0.1 μas in the size estimates of the source in the EHT simulations. With minimal modification to the existing software, we can easily apply our code to perform astrometry.

Finally, in all our experiments, we have used simple models to describe the sky. Seldom is the region of interest on the sky composed of single point sources or simple extended sources with Gaussian profiles. At the very least, we need to model the sky with a combination of point and extended sources. Such an experiment can be done with our current capabilities for analysing the jet from CGRaBS J0809+5341. Source models such as a core with two jets emanating from it must be tested against the simpler models to carry out in-depth studies of the blazar.

6.2.2 CHALLENGES

A major challenge in pursuing the probability-theoretic approach to data analysis has always been not one of theory, but one of implementation. Because the evidence calculation requires the computation of a multi-dimensional integral, the execution time increases exponentially with increasing dimensionality of the model. Until the mid-2000s, the computational power available publicly could not handle this calculation in a reasonable amount of time. Now, with the availability of GPUs and parallel programming techniques, we can adopt the probabilistic approach in performing scientific inference. MONTBLANC implements a generic RIME, but the current version can run only on one GPU. A version that could run on multiple GPUs making use of GPU clusters will be necessary to handle a high-dimensional parameter space or more data. This and more improvements are currently in the works.

Another challenge is that the sampler I use, MULTINEST, performs best with models containing only a few tens of parameters. Implementing the RIMEs for some of the aforementioned applications for arrays of many stations will necessitate the use of next-generation samplers that can compute the evidence for models with hundreds of parameters in a reasonable duration. One such sampler, POLYCHORD, a successor to MULTINEST, is being integrated into my software. With the completion of this, we will be able to tackle larger problem domains. And with the data rates and volumes expected to soar for future interferometers, it is imperative that our software be ready to handle them.

The radio waves from a source experience propagation path effects such as amplitude and phase effects due to the intervening medium, the Earth's atmosphere, and the primary beams of the stations of an interferometer, to mention a few. In our goal to combine current calibration and imaging procedures with drawing scientific inferences, a better understanding of these effects and the astrophysical causes of the radio emission from the source itself is necessary. Providing the ability to handle the scientific and instrumental parameters in our GPU-accelerated software is a challenge well-worth undertaking given the scientific returns expected.

The theory has been spelt out clearly for achieving the foreseeable scientific goals. With the proposed advancements, the software will get us closer to performing probabilistic analyses of radio data on a large scale. As so often happens in science, the path will be tortuous and full of surprises. Though at any point during the journey, the domain to which our theory can be applied successfully will be finite, we can hope to advance one step further from wherever we are, if we reason carefully enough. *The world is finite; our hopes spill over its rim.*

References

Ables, J. G. (1974). Maximum Entropy Spectral Analysis. *Astronomy & Astrophysics*, 15, 383.

An, T., Cui, Y.-Z., Paragi, Z., Frey, S., Gurvits, L. I., & Gabányi, K. É. (2016). VLBI observations of flared optical quasar CGRaBS J0809+5341. *Publications of the Astronomical Society of Japan*, 68, 77.

Antonucci, R. (1993). Unified models for active galactic nuclei and quasars. *Annual Reviews of Astronomy & Astrophysics*, 31, 473.

Antonucci, R. R. J. & Miller, J. S. (1985). Spectropolarimetry and the nature of NGC 1068. *The Astrophysical Journal*, 297, 621–632.

Baade, W. & Minkowski, R. (1954). On the Identification of Radio Sources. *The Astrophysical Journal*, 119, 215.

Balanis, C. A. (2005). *Antenna Theory: Analysis & Design*. Wiley-Interscience, 3rd edition.

Balanutsa, P., Denisenko, D., Lipunov, V., Gorbovskoy, E., Tiurina, N., Kornilov, V., Shatskiy, N., Chazov, V., Kuznetsov, A., Vladimirov, V., Yecheistov, V., Ivanov, K., Yazev, S., Budnev, N., Konstantinov, E., Chuvalaev, O., Poleshchuk, V., Gress, O., Yurkov, V., Sergienko, Y., Varda, D., Sinyakov, E., Gabovich, A., Parkhomenko, A., Tlatov, A., Dormidontov, D., Senik, V., Krushinsky, V., Zalozhnih, I., Popov, A., Bourdanov, A., Podvorotny, P., Shumkov, V., Shurpakov, S., Levato, H., Saffe, C., Mallamaci, C., Lopez, C., & Podest, F. (2014). Possibly the most powerful persistent object in the Universe detected by MASTER. *"The Astronomer's Telegram"*, 6096.

Barsdell, B. R., Barnes, D. G., & Fluke, C. J. (2010). Analysing astronomy algorithms for graphics processing units and beyond. *Monthly Notices of the Royal Astronomical Society*, 408(3), 1936.

Bayes, T. & Price, R. (1763). An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.

Bernoulli, J. (1713). *Ars conjectandi:*. Impensis Thurnisiorum, fratrum.

Bertero, M. & de Mol, C. (1996). Super-resolution by data inversion. *Progress in Optics*, XXXVI, 129.

Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. (2008). Correcting direction-dependent gains in the deconvolution of radio interferometric images. *Astronomy & Astrophysics*, 487, 419–429.

Blecher, T., Deane, R., Bernardi, G., & Smirnov, O. (2017). MEQSILHOUETTE: a mm-VLBI observation and signal corruption simulator. *Monthly Notices of the Royal Astronomical Society*, 464, 143–151.

Born, M. & Wolf, E. (1999). *Principles of Optics*. Cambridge University Press, 7th edition.

Bracewell, R. N. (2000). *The Fourier Transform and its Applications*. McGraw-Hill, 3rd edition.

Briggs, D. (1995). *High Fidelity Deconvolution of Moderately Resolved Sources*. PhD thesis, New Mexico Institute of Mining and Technology.

Broderick, A. E. & Loeb, A. (2009). Imaging the Black Hole Silhouette of M87: Implications for Jet Formation and Black Hole Spin. *The Astrophysical Journal*, 697, 1164–1179.

Bruni, G., Gómez, J. L., Casadio, C., Lobanov, A., Kovalev, Y. Y., Sokolovsky, K. V., Lisakov, M. M., Bach, U., Marscher, A., Jorstad, S., Anderson, J. M., Krichbaum, T. P., Savolainen, T., Vega-García, L., Fuentes, A., Zensus, J. A., Alberdi, A., Lee, S.-S., Lu, R.-S., Pérez-Torres, M., & Ros, E. (2017). Probing the innermost regions of AGN jets and their magnetic fields with RadioAstron - II. Observations of 3C273 at minimum activity. *Astronomy & Astrophysics*, 604, A111.

Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., Merloni, A., Salvato, M., Donley, J., & Kocevski, D. (2014). X-ray spectral modelling of the agn obscuring region in the cdfs: Bayesian model selection and catalogue. *Astronomy & Astrophysics*, 564, A125.

Carilli, C. L. & Holdaway, M. A. (1999). Tropospheric phase calibration in millimeter interferometry. *Radio Science*, 34(4), 817–840.

Christiansen, W. N. & Högbom, J. A. (1985). *Radiotelescopes*. Cambridge University Press, 2nd edition.

Condon, J. J. & Ransom, S. M. (2016). *Essential Radio Astronomy*. Course Notes.

Cornwell, T. J. (2008). Multiscale clean deconvolution of radio synthesis images. *IEEE Journal of Selected Topics in Signal Processing*, 2(5), 793–801.

Cornwell, T. J. & Perley, R. A. (1992). Radio interferometric imaging of very large fields. *Astronomy & Astrophysics*, 261, 353–364.

Cornwell, T. J. & Wilkinson, P. N. (1981). A new method for making maps with unstable radio interferometers. *Monthly Notices of the Royal Astronomical Society*, 196, 1067–1086.

Cotton, W. D. (1995). Fringe Fitting. In J. A. Zensus, P. J. Diamond, & P. J. Napier (Eds.), *Very Long Baseline Interferometry and the VLBA*, volume 82 of *Astronomical Society of the Pacific Conference Series* (pp. 189).

Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, 14.

di Francia, G. T. (1969). Degrees of freedom of an image. *Journal of the Optical Society of America*, 59(7), 799–804.

Doeleman, S., Agol, E., Backer, D., Baganoff, F., Bower, G. C., Broderick, A., Fabian, A., Fish, V., Gammie, C., Ho, P., Honman, M., Krichbaum, T., Loeb, A., Marrone, D., Reid, M., Rogers, A., Shapiro, I., Strittmatter, P., Tilanus, R., Weintroub, J., Whitney, A., Wright, M., & Ziurys, L. (2009). Imaging an Event Horizon: submm-VLBI of a Super Massive Black Hole. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, volume 2010 of *ArXiv Astrophysics e-prints*.

Enßlin, T. A., Frommert, M., & Kitaura, F. S. (2009). Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis. *Physical Review D*, 80, 105005.

Falcke, H. & Markoff, S. B. (2013). Toward the event horizon - the supermassive black hole in the Galactic Center. *Classical and Quantum Gravity*, 30(24), 244003.

Fanaroff, B. L. & Riley, J. M. (1974). The morphology of extragalactic radio sources of high and low luminosity. *Monthly Notices of the Royal Astronomical Society*, 167, 31P–36P.

Feroz, F. & Hobson, M. P. (2008). Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis. *Monthly Notices of the Royal Astronomical Society*, 384, 449–463.

Feroz, F., Hobson, M. P., & Bridges, M. (2009). MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398, 1601–1614.

Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. (2013). Importance Nested Sampling and the MULTINEST algorithm. *astro-ph/1306.2144v2*.

Feroz, F., Hobson, M. P., Zwart, J. T. L., Saunders, R. D. E., & Grainge, K. J. B. (2009). Bayesian modelling of clusters of galaxies from multifrequency-pointed Sunyaev-Zel'dovich observations. *Monthly Notices of the Royal Astronomical Society*, 398, 2049–2060.

Feynman, R. P., Leighton, R. B., & Sands, M. (1977). *The Feynman Lectures on Physics*. Addison-Wesley.

Fish, V. L., Johnson, M. D., Lu, R.-S., Doeleman, S. S., Bouman, K. L., Zoran, D., Freeman, W. T., Psaltis, D., Narayan, R., Pankratius, V., Broderick, A. E., Gwinn, C. R., & Vertatschitsch, L. E. (2014). Imaging an event horizon: Mitigation of scattering toward sagittarius a*. *The Astrophysical Journal*, 795(2), 134.

Flynn, M. J. (1972). Some computer organizations and their effectiveness. *IEEE Transactions on Computers*, C-21(9), 948–960.

Good, I. J. (1985). Weight of evidence: A brief survey. In Bernardo J. M., deGroot M. H., Lindley D. V., & Smith A. F. M. (Eds.), *Bayesian Statistics 2* (pp. 249–270).

Goodman, J. (1968). *Introduction to Fourier Optics*. McGraw-Hill Book Company, 1st edition.

Hamaker, J. P. & Bregman, J. D. (1996). Understanding Radio Polarimetry III. Interpreting the IAU/IEEE definitions of the Stokes parameters. *Astronomy & Astrophysics*, 117, 161.

Hamaker, J. P., Bregman, J. D., & Sault, R. J. (1996). Understanding Radio Polarimetry I. Mathematical Foundations. *Astronomy & Astrophysics*, 117, 137.

Handley, W. J., Hobson, M. P., & Lasenby, A. N. (2015a). POLYCHORD: nested sampling for cosmology. *Monthly Notices of the Royal Astronomical Society*, 450, L61–L65.

Handley, W. J., Hobson, M. P., & Lasenby, A. N. (2015b). POLYCHORD: next-generation nested sampling. *Monthly Notices of the Royal Astronomical Society*, 453, 4384–4398.

Harris, J. L. (1964). Diffraction and resolving power. *Journal of the Optical Society of America*, 54(7), 931–936.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.

Healey, S. E., Romani, R. W., Cotter, G., Michelson, P. F., Schlafly, E. F., Readhead, A. C. S., Giommi, P., Chaty, S., Grenier, I. A., & Weintraub, L. C. (2008). CGRaBS: An All-Sky Survey of Gamma-Ray Blazar Candidates. *The Astrophysical Journal Supplement Series*, 175, 97–104.

Hobson, M. P. & Maisinger, K. (2002). Maximum-likelihood estimation of the cosmic microwave background power spectrum from interferometer observations. *Monthly Notices of the Royal Astronomical Society*, 334, 569–588.

Högbom, J. A. (1974). Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy & Astrophysics Supplement Series*, 15, 417.

Honma, M., Akiyama, K., Uemura, M., & Ikeda, S. (2014). Super-resolution imaging with radio interferometer using sparse modeling. *arXiv e-prints, astro-ph/1407.2422*.

Horiuchi, S., Fomalont, E. B., Taylor, W. K., Scott, A. R., Lovell, J. E. J., Moellenbrock, G. A., Dodson, R., Murata, Y., Hirabayashi, H., Edwards, P. G., Gurvits, L. I., & Shen, Z.-Q. (2004). The VSOP 5 GHz Active Galactic Nucleus Survey. IV. The Angular Size/Brightness Temperature Distribution. *The Astrophysical Journal*, 616, 110–122.

Jansky, K. G. (1932). Directional Studies of Atmospherics at High Frequencies. *Proceedings of the Institute of Radio Engineers*, 20.

Jansky, K. G. (1933). Electrical Disturbances Apparently of Extraterrestrial Origin. *Proceedings of the Institute of Radio Engineers*, 21.

Jaynes, E. T. (1982). On the Rationale of Maximum Entropy Methods. *IEEE Proceedings*, 70, 939–952.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, 1st edition.

Jeffreys, H. (1961). *Theory of Probability*. Clarendon Press, 3rd edition.

Jennison, R. C. (1958). A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent. *Monthly Notices of the Royal Astronomical Society*, 118, 276.

Johnson, M. D. & Narayan, R. (2016). The Optics of Refractive Substructure. *The Astrophysical Journal*, 826, 170.

Jones, R. C. (1941). New calculus for the treatment of optical systems I. Description and discussion of the calculus. *Journal of the Optical Society of America*.

Jorstad, S. G., Marscher, A. P., Mattox, J. R., Aller, M. F., Aller, H. D., Wehrle, A. E., & Bloom, S. D. (2001). Multiepoch Very Long Baseline Array Observations of EGRET-detected Quasars and BL Lacertae Objects: Connection between Superluminal Ejections and Gamma-Ray Flares in Blazars. *The Astrophysical Journal*, 556, 738–748.

Junklewitz, H., Bell, M. R., & Enßlin, T. (2015). A new approach to multifrequency synthesis in radio interferometry. *Astronomy & Astrophysics*, 581, A59.

Junklewitz, H., Bell, M. R., Selig, M., & Enßlin, T. A. (2016). RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy. *Astronomy & Astrophysics*, 586, A76.

Kass, R. E. & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90, 773–795.

Kawaguchi, N., Jiang, W., & Shen, Z.-Q. (2015). VLBI observation of the M 81 core with the CVN in X-band and the Pseudo-Closure Analysis. *Publications of the Astronomical Society of Japan*, 67, 112.

Kellerman, K. & Owen, F. (1988). Radio galaxies and quasars. *Galactic and Extragalactic Radio Astronomy*, (pp. 653).

Kellerman, K. I. & Pauliny-Toth, I. I. K. (1969). The spectra of opaque radio sources. *The Astrophysical Journal*, 155, L71.

Kellerman, K. I. (2003). Variability, Brightness Temperature, Superluminal Motion, Doppler Boosting, and Related Issues. *Radio Astronomy at the Fringe, ASP Conference Series*, 300.

Klöckner, A., Pinto, N., Lee, Y., Catanzaro, B., Ivanov, P., & Fasih, A. (2012). PyCUDA and PyOpenCL: A Scripting-Based Approach to GPU Run-Time Code Generation. *Parallel Computing*, 38 (3), 157–174.

Knuth, D. E. (1973). *The Art of Computer Programming, Vol 1: (2Nd Ed.) Fundamental Algorithms*. Addison Wesley Longman Publishing Co., Inc.

Kovalev, Y. Y., Kardashev, N. S., Kellermann, K. I., Lobanov, A. P., Johnson, M. D., Gurvits, L. I., Voitsik, P. A., Zensus, J. A., Anderson, J. M., Bach, U., Jauncey, D. L., Ghigo, F., Ghosh, T., Kraus, A., Kovalev, Y. A., Lisakov, M. M., Petrov, L. Y., Romney, J. D., Salter, C. J., & Sokolovsky, K. V. (2016). RadioAstron Observations of the Quasar 3C273: A Challenge to the Brightness Temperature Limit. *The Astrophysical Journal Letters*, 820(1), L9.

Kraus, J. D. (1986). *Radio Astronomy*. Cygnus-Quasar Books, 2nd edition.

Lancaster, K., Genova-Santos, R., Falcòn, N., Grainge, K., Gutièrrez, C., Kneissl, R., Marshall, P., Pooley, G., Rebolo, R., Rubiño-Martin, J.-A., Saunders, R. D. E., Waldram, E., & Watson, R. A. (2005). Very Small Array observations of the Sunyaev-Zel'dovich effect in nearby galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 359, 16–30.

Lang, S. (1986). *Introduction to Linear Algebra*. Springer, 2nd edition.

Laplace, P.-S. (1774). Memoir on Inverse Probability.

Lobanov, A. P. (2005). Resolution limits in astronomical images. *arXiv e-prints, astro-ph/0503225*.

Lochner, M., Natarajan, I., Zwart, J., Smirnov, O., Bassett, B., Oozeer, N., & Kunz, M. (2015). Bayesian Inference for Radio Observations. *Monthly Notices of the Royal Astronomical Society*, 450, 1308–1319.

Mackay, D. J. C. (1991). *Bayesian Methods for Adaptive Models*. PhD thesis, California Institute of Technology.

Mackay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 1st edition.

Marscher, A. P. & Scott, J. S. (1980). Superluminal motion in compact radio sources. *Publications of the Astronomical Society of the Pacific*, 92, 127–133.

Marshall, P. (2003). *Bayesian Analysis of Clusters of Galaxies*. PhD thesis, University of Cambridge.

Martí-Vidal, I., Pérez-Torres, M. A., & Lobanov, A. P. (2012). Over-resolution of compact sources in interferometric observations. *Astronomy & Astrophysics*, 541, A135.

Martí-Vidal, I., Vlemmings, W. H. T., Muller, S., & Casey, S. (2014). UVMULTIFIT: A versatile tool for fitting astronomical radio interferometric data. *Astronomy & Astrophysics*, 563, A136.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21, 1087.

Middelberg, E. & Bach, U. (2008). High resolution radio astronomy using very long baesline interferometry. *Reports on Progress in Physics*, 71, 6901.

Mort, B. J., Dulwich, F., Salvini, S., Adami, K. Z., & Jones, M. E. (2010). Oskar: Simulating digital beamforming for the ska aperture array. In *2010 IEEE International Symposium on Phased Array Systems and Technology* (pp. 690–694).

Mukherjee, P., Parkinson, D., & Liddle, A. R. (2006). A Nested Sampling Algorithm for Cosmological Model Selection. *The Astrophysical Journal*, 638, L51–L54.

Natarajan, I., Paragi, Z., Zwart, J., Perkins, S., Smirnov, O., & van der Heyden, K. (2017). Resolving the blazar CGRaBS J0809+5341 in the presence of telescope systematics. *Monthly Notices of the Royal Astronomical Society*, 464, 4306–4317.

Neal, R. M. (1993). *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Technical report, University of Toronto.

Nityananda, R. & Narayan, R. (1982). Maximum entropy image reconstruction - A practical non-information-theoretic approach. *Journal of Astrophysics and Astronomy*, 3, 419–450.

Noordam, J. E. (1996). *The Measurement Equation of a Generic Radio Telescope*. Technical report, AIPS++ Implementation Note.

Noordam, J. E. (2004). LOFAR Calibration Challenges. *Ground-based Telescopes, SPIE Conference Series*, 5489.

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

NVIDIA Corporation (2016). NVIDIA CUDA C programming guide. Version 8.0.

Offringa, A. R., de Bruyn, A. G., & Zaroubi, S. (2012). Post-correlation filtering techniques for off-axis source and RFI removal. *Monthly Notices of the Royal Astronomical Society*, 422, 563–580.

Osgood, B. (2016). *The Fourier Transform and its Applications*. EE261 Lecture Notes.

Paliya, V. S. (2015). Fermi-Large Area Telescope Observations of the Exceptional Gamma-Ray Flare from 3C 279 in 2015 June. *The Astrophysical Journal, Letters*, 808, L48.

Pâris et al. (2014). The Sloan Digital Sky Survey quasar catalog: tenth data release. *Astronomy & Astrophysics*, 563, A54.

Parker, R. L. (1977). Understanding Inverse Theory. *Annual Review of Earth and Planetary Sciences*, 5, 35–64.

Pelleg, D. & Moore, A. W. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML ’00 (pp. 727–734). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Perkins, S., Marais, P., Zwart, J., Natarajan, I., Tasse, C., & Smirnov, O. (2015). Montblanc: GPU accelerated radio interferometer measurement equations in support of Bayesian inference for radio observations. *Astronomy and Computing*, 12, 73–85.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3rd edition.

Pólya, G. (1954). *Mathematics and Plausible Reasoning*. Princeton University Press.

Rau, U., Bhatnagar, S., Voronkov, M. A., & Cornwell, T. J. (2009). Advances in Calibration and Imaging Techniques in Radio Interferometry. *Proceedings of the IEEE*, 97.

Rayleigh (1879). Investigations in optics, with special reference to the spectroscope. *Philosophical Magazine and Journal of Science*, 8, 49.

Readhead, A. C. S., Walker, R. C., Pearson, T. J., & Cohen, M. H. (1980). Mapping radio sources with uncalibrated visibility data. *Nature*, 285, 137–140.

Readhead, A. C. S. & Wilkinson, P. N. (1978). The mapping of compact radio sources from VLBI data. *The Astrophysical Journal*, 223, 25–36.

Reber, G. (1940). Cosmic Static. *Astrophysical Journal*, 91, 621.

Rees, M. J. (1966). Appearance of relativistically expanding radio sources. *Nature*, 211, 468.

Rybicki, G. & Lightman, A. (1985). *Radiative Processes in Astrophysics*. Wiley-VCH.

Ryle, M. (1955). The Application of Interferometric Methods in Radio Astronomy. *Vistas in Astronomy*, 1, 532–541.

Schwab, F. R. (1984). Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry. *The Astronomical Journal*, 89, 1076–1081.

Schwab, F. R. & Cotton, W. D. (1983). Global fringe search techniques for VLBI. *The Astronomical Journal*, 88, 688–694.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379–423.

Shaw, J. R., Bridges, M., & Hobson, M. P. (2007). Efficient Bayesian inference for multimodal problems in cosmology. *Monthly Notices of the Royal Astronomical Society*, 378, 1365–1370.

Shepherd, M. C. (1997). Difmap: An interactive program for synthesis imaging. *Astronomical Data Analysis Software and Systems VI, ASP Conference Series*, 125.

Sivia, D. S. & Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial*. Oxford University Press, 2nd edition.

Skilling, J. (2004). Nested Sampling. In F. R., P. R., & T. U. V. (Eds.), *24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (pp. 395–405).

Smirnov, O. M. (2011a). Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism. *Astronomy & Astrophysics*, 527, A106.

Smirnov, O. M. (2011b). Revisiting the radio interferometer measurement equation. II. Calibration and direction-dependent effects. *Astronomy & Astrophysics*.

Smirnov, O. M. (2011c). Revisiting the radio interferometer measurement equation. III. Addressing direction-dependent effects in 21 cm WSRT observations of 3C 147. *Astronomy & Astrophysics*.

Smirnov, O. M. (2011d). Revisiting the radio interferometer measurement equation. IV. A generalized tensor formalism. *Astronomy & Astrophysics*.

Stroustrup, B. (2013). *The C++ Programming Language*. Addison-Wesley Professional, 4th edition.

Tasse, C. (2014). Nonlinear kalman filters for calibration in radio interferometry. *Astronomy & Astrophysics*, 566, A127.

Taylor, G. B., Carilli, C. L., & Perley, R. A., Eds. (1999). *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*. Astronomical Society of the Pacific Conference Series, 1st edition.

Thompson, A. R., Moran, J. M., & George W. Swenson, J. (2017). *Interferometry and Synthesis in Radio Astronomy*. Springer Verlag, 3rd edition.

Tingay, S. J., Preston, R. A., Lister, M. L., Piner, B. G., Murphy, D. W., Jones, D. L., Meier, D. L., Pearson, T. J., Readhead, A. C. S., Hirabayashi, H., Murata, Y., Kobayashi, H., & Inoue, M. (2001). Measuring the Brightness Temperature Distribution of Extragalactic Radio Sources with Space VLBI. *The Astrophysical Journal Letters*, 549, L55–L58.

Trotta, R. (2008). Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49, 71–104.

Urry, C. M. & Padovani, P. (1995). Unified schemes for radio-loud active galactic nuclei. *Publications of the Astronomical Society of the Pacific*, 107, 803.

van Rossum, G. (1995). *Python reference manual*. Report CS-R9525.

Volonteri, M. (2010). Formation of supermassive black holes. *The Astronomy and Astrophysics Review*, 18, 279–315.

Wilson, T., Rohlfs, K., & Hüttemeister, S. (2009). *Tools of Radio Astronomy*. Springer, 5th edition.

Zensus, J. A., Ros, E., Kellermann, K. I., Cohen, M. H., Vermeulen, R. C., & Kadler, M. (2002). Sub-milliarcsecond Imaging of Quasars and Active Galactic Nuclei. II. Additional Sources. *The Astronomical Journal*, 124, 662–674.

Zwart, J. T. L., Feroz, F., Davies, M. L., Franzen, T. M. O., Grainge, K. J. B., Hobson, M. P., Hurley-Walker, N., Kneissl, R., Lasenby, A. N., Olamaie, M., Pooley, G. G., Rodríguez-Gonzálvez, C., Saunders, R. D. E., Scaife, A. M. M., Scott, P. F., Shimwell, T. W., Titterington, D. J., & Waldram, E. M. (2011). Sunyaev-Zel'dovich observations of galaxy clusters out to the virial radius with the Arcminute Microkelvin Imager. *Monthly Notices of the Royal Astronomical Society*, 418, 2754–2772.

Index