Identification of Symmetric Achromatic Variability in Blazar Light Curves using Bayesian Inference

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF SCIENCE AND ENGINEERING

2021

Thomas C. McAloone
Department of Physics and Astronomy

Contents

Word Count: 24158

List of Figures

List of Tables

Abstract

The work presented in this thesis is centred around Symmetric Achromatic Variability, a rare phenomenon observed in the light curves of blazar sources. SAV is hypothesised to arise through gravitational milli-lensing when relativistically moving components traverse the lensing caustics created by some intervening massive object(s). It was first identified due to the presence of symmetric U-shaped features within 15GHz blazar light curves produced using the Owens Valley Radio Observatory 40m telescope. In this thesis I present a model to describe the light curve of a lensed blazar source. I then introduce a fitting procedure which uses nested sampling to fit this model to blazar light curves, with the goal of using this procedure to identify SAV within the the OVRO data set. I also demonstrate how this procedure can then be used to generate artificial blazar light curves based on real OVRO data. I then show the results of running the fitting procedure with both real and simulated blazar light curves.

Additionally, I propose a framework for a transdimensional alternative to standard nested sampling algorithms, where the number of model parameters, N, is itself included as a free parameter. Such an approach could have wide-ranging applications, including for the problem of SAV identification. I describe the process by which we explore the parameter space, including the introduction of a novel concept - the 'flattened' particle space. Using both data and an analytical approach, I investigate the performance of this method.

Finally, I detail an ongoing campaign to monitor the SAV candidate J1415+1320 using e-MERLIN at both L- and C-band, in order to gather evidence for or against the achromaticity of the SAV events. I then present and analyse light curves produced from the most recently available data.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Copyright

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, the University Library's regulations (see http://www.library.manchester.ac.uk/aboutregulations/) and in the University's policy on Presentation of Theses.

Acknowledgements

The Author

The author graduated from the University of Warwick in July 2017 with an Upper Second-Class Honours degree (MEng) in Computer Science. In September 2017, the author began a PhD in the Jodrell Bank Centre for Astrophysics at the University of Manchester supervised by Keith Grainge. This PhD was part of the 4IR STFC Centre for Doctoral Training in Data Intensive Science.

Chapter 1

Introduction

Symmetric Achromatic Variability (SAV) is a rare phenomenon, in which the emission from blazar sources is believed to be gravitationally lensed by an intervening massive object. In this chapter I introduce the material necessary to understand this process, giving an overview of Active Galactic Nuclei (AGN), their classifications and the mechanisms through which they emit electromagnetic radiation. I then introduce the Owens Valley Radio Observatory (OVRO) 40m blazar monitoring program, which provided the first evidence of SAV activity. I then recount the origins of the SAV hypothesis, giving a description of a typical proposed lensing scenario. Finally, I give an overview of the structure and content of the rest of this thesis.

1.1   Active Galactic Nuclei

Active Galactic Nuclei (AGN) are the most luminous sustained sources of electromagnetic radiation in the universe, observed across every frequency band from

radio to gamma rays. The current consensus understanding of AGN is that their emission results from a rapidly spinning super massive black hole (SMBH) at the centre of a host galaxy. This SMBH, typically with a mass between $10^4 - 10^{10} M_\odot$ (Beckmann & Shrader 2012), is orbited by a disk of diffuse material called an accretion disk. Through an accretion process, the SMBH channels the released potential energy of the matter in this disk into an outflow of radiation and particles in the form of relativistic jets aligned with the spin axis of this 'central engine'. These jets are highly collimated, often to distances over 100 kpc, and are so energetic that their kinetic power can, in some cases, contribute a significant fraction of the total bolometric luminosity of the system. Although we can observe AGN back to redshifts of $z > 7$, it seems that the maximum AGN activity is found around $z = 1 - 3$.

1.1.1 AGN Classification

A multitude of AGN types have been observed and, over the years, a number of classifications have arisen to distinguish between their observational characteristics. These classifications are intrinsically linked to the history of AGN research, and so, to understand the modern view of AGN, it is important to consider that history. In the early 20th century, the physical properties of a number of sources had been explored, however these objects had not yet been identified as AGN. Broad lines were found in the emission spectra of nebulae, indicating material moving with high velocity dispersions (Slipher 1917, Seyfert 1943). These *Seyfert galaxies* were later split into two categories: Type I and Type II. Seyfert I galaxies show both narrow and broad lines in their emission spectra, whereas Seyfert II galaxies

only exhibit narrow emission lines. Here, broad emission lines are those with a velocity width of $\gtrsim 1000 \, \mathrm{km \, s^{-1}}$, whereas narrow emission lines have a velocity width of $\lesssim 1000 \, \mathrm{km \, s^{-1}}$.

In 1959, members the Radio Astronomy Group of the University of Cambridge published the Third Cambridge Catalogue of Radio Sources (3C), which was later revised in 1962 (Edge et al. 1959, Bennett 1962). For many years this catalogue was considered to be the definitive listing of bright radio sources in the Northern Hemisphere. It proved difficult to identify optical counterparts for these sources until Schmidt (1965) identified a star-like counterpart for 3C 273. This counterpart, along with others identified later, was dubbed a *quasi-stellar object*, or *quasar*. Although the optical source was point like, it had a very different spectrum and was identified as having a high redshift of $z = 0.158$.

Osterbrock (1978) suggested that the lack of broad emission lines in Seyfert II galaxies could be due to absorption of most of the ionizing photons in the dense gas near the central source. This idea paved the way for a "unified scheme", which suggests that observationally different classes of AGN are fundamentally the same type of object, viewed from different angles (Orr & Browne 1982, Readhead et al. 1978, Antonucci 1993). This leads to the modern picture of the structure of AGN which is depicted in fig. 1.1. Here we see a broad line region (BLR), which is comprised of gas clouds close to the accretion disk in a rapid orbit, as well as narrow line region (NLR) clouds which are further out from the central engine. The accretion disk is surrounded by a torus of optically thick dust, which is typically larger than the accretion disk and so, from some angles, would obscure the BLR. The NLR, being further out, is visible from a much wider range of viewing angles, thus explaining the observational differences in Seyfert I and II galaxies.
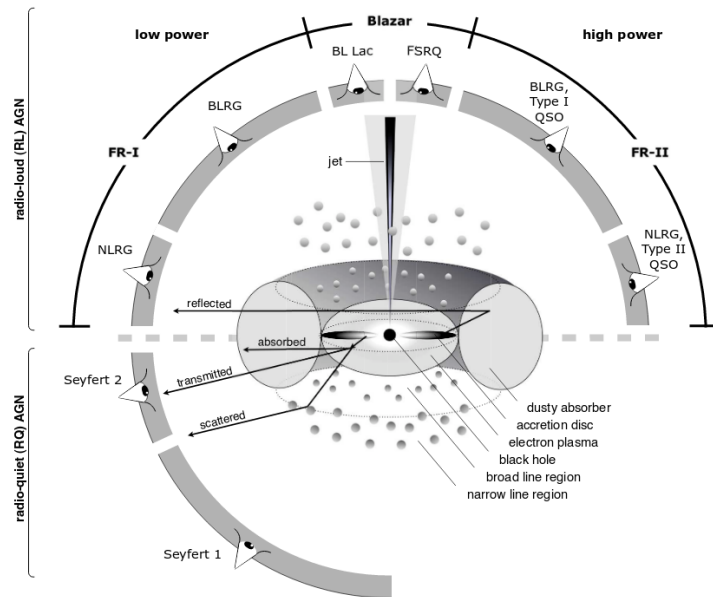
Figure 1.1: Schematic representation of our understanding of the AGN phenomenon in the unified scheme. The type of object we see depends on the viewing angle, whether or not the AGN produces a significant jet emission, and how powerful the central engine is. Note that radio loud objects are generally thought to display symmetric jet emission (Figure from Beckmann & Shrader 2012).

We also see that AGN can be classified as radio-loud objects, indicating the presence of large scale radio jets and lobes, as well as radio-quiet objects, in which no well-defined radio jets can be identified. Interestingly, it seems that radio-loud AGN tend to be found within elliptical galaxies, whereas radio-quiet AGN are more typically associated with spiral galaxies. The exact reason for this large range in radio powers remains elusive, but Wilson & Colbert (1995) suggest that the difference between the classes is associated with the spin of the black hole. Assuming that the process of accretion is not enough to produce rapidly spinning holes, the merger of two SMBHs following the merger of their host galaxies could be. This idea conforms with the generally-accepted opinion that elliptical galaxies are formed by the merger of spiral galaxies, and would therefore explain the

association of radio-loud AGN with elliptical galaxies.

Radio-loud AGN are then further divided into 3 categories. Fanaroff & Riley (1974) identified two of these categories based on the brightness and structure of the object's radio emission: Fanaroff-Riley type I (FR I) galaxies, which are low-luminosity and the FR II galaxies which are high-luminosity. The third group of radio-loud AGN are *blazars*. These are AGN which have a relativistic radio jet aligned closely with our line of sight from Earth. It is generally accepted that because of this alignment, anisotropic boosting of the radiation along the direction of motion gives rise to an apparent enhanced luminosity at all wavelengths (Kellermann et al. 2003). This also leads to the measured flux density of these sources being highly variable (in comparison with the more stable measurements of other AGN). This phenomenon involves the observation of apparent faster-than-light motion, however this is merely an 'optical illusion' which can be simply explained without any incompatibilities with the theory of special relativity. Because AGN jets are emitting light at every point of their path and are travelling at relativistic speeds, if they are aligned at a very small angle relative to the observer then the light they emit does not approach the observer much more quickly than the jet itself. Hence, light emitted over large time scales (hundreds of years) over the course of the jet's travel does not have hundreds of light-years between the earliest and latest light emitted in the jet (front and back ends). This means the complete "train" of light arrives at the observer over a much smaller time period, giving the illusion of faster-than-light travel.

1.1.2   Jet Emission Process

Blazar broadband emission is characterised by an observed spectral energy distribution (SED) with two components. The first component is widely accepted to be the result of synchrotron emission and ranges from radio through ultraviolet/X-ray radiation, with a peak around the infrared (IR) band. Synchrotron radiation involves highly relativistic charged particles (electrons and possibly positrons) being subjected to a radial acceleration due to a magnetic field. Larmor's equation tells us that any charged particle radiates when accelerated, so the accelerated particles in the AGN jets will radiate as they travel. While synchrotron is understood to be the dominant emission process from the radio to the optical/UV/X-ray domain, the second component is less well-understood, but typically it is thought to originate from inverse Compton scattering. This involves low-energy photons being scattered to higher energies by ultrarelativistic electrons, meaning the photons gain energy and the electrons lose energy. At present, however, it is unclear where the seed photons for this second SED component emerge from (Beckmann & Shrader 2012). One possibility is that blazars are predominantly one-zone synchrotron self-compton emitters (SSC), that is to say that the electromagnetic radiation from the synchrotron emission provides the very same seed photons for inverse Compton scattering. This would explain the SEDs observed, since we would likely see contributions from both emission mechanisms. Alternatively, it could be that the seed photons emerge from e.g. the accretion disk or broad line region, indicating an external Compton process (Dermer et al. 1992).

1.2 Symmetric Achromatic Variability

Symmetric Achromatic Variability refers to a very rare form of AGN variability, where the brightness of an active galaxy is observed to fluctuate in a specific manner, resulting in time-symmetric U-shaped features in its light curve. It is currently hypothesised that this phenomenon is the result of gravitational millilensing due to some intervening, massive lens.

1.2.1 The OVRO 40m Blazar Monitoring Program

The Owens Valley Radio Observatory (OVRO), located in California, USA, is one of the largest university-operated radio observatories in the world. Since it was established in 1958, it has been involved in decades of research, spanning many fields of radio astronomy. In late 2007, OVRO embarked upon a new research campaign involving the 40m telescope. Originally this involved supporting the then-recently launched Fermi Gamma-ray Space Telescope by monitoring 1158 blazars every two days. By performing measurements at 15 GHz (with a bandwidth of 3GHz), the resulting light curves could be compared with those produced using Fermi gamma-ray measurements of the same sources. This allowed for analysis of correlations in the variability, in the hopes of bettering our understanding of the AGN emission mechanisms.

One key element of such a campaign is the selection of sources to monitor. It is desirable to draw conclusions that are statistically robust, which may be meaningfully extrapolated to the parent population. Therefore the sample drawn was designed to be complete with respect to to the physical characteristics that could affect those conclusions. If it is not possible for the sample to be complete in this

regard, the impacts must be understood thoroughly before making claims about any conclusions drawn. In the case of OVRO, the sample selection was driven by three considerations (Richards et al. 2011). Firstly, since the goal was the detailed study of the radio variability properties of the blazar population and the dependence of those properties on observables such as redshift, the sample had to be large enough to divide into sub-samples (e.g. in redshift or luminosity bins), with each sub-sample containing enough members to permit statistical characterisation.

Secondly, to allow for the evaluation of the confidence level of any correlations or variable dependencies identified in the data through Monte Carlo simulations, and the generalisation of any findings to the parent blazar population, the sample would have to be statistically well-defined, using uniform and easily repeatable criteria. It would not be sufficient for robust statistical study to simply choose bright, easily observable sources.

Finally one of the major goals of the program was the cross-correlation of 15GHz light curves with Fermi gamma-ray data. Therefore, it would be ideal for the sample to include a large number of gamma-ray-loud blazars. At the same time, however, it was also preferable that the sample could help to answer the question of why some blazars are gamma-ray-loud, while others, with apparently similar properties, are not. For this reason the sample would ideally be preselected, before Fermi data biased any understanding of what constitutes a likely gamma-ray-loud blazar. Therefore a comparable number of blazars which are not gamma-ray loud would be preferable. Blazars in the Candidate Gamma-Ray Blazar Survey (CGRaBS) satisfy all of the requirements above (Healey et al. 2008).
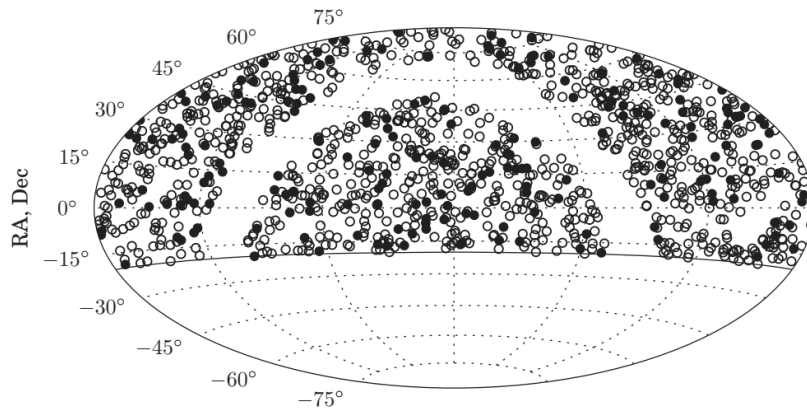
Figure 1.2: Positions of the CGRaBS sources in the OVRO 40m program in equatorial coordinates. Filled circles indicate CGRaBS that are also in the 1LAC sample. The solid line marks the $-20°$ declination limit of the program. The semi-circular gap is excluded to avoid the galactic plane. From Richards (2012).

The CGRaBS sample is a total of 1625 AGN over the whole sky outside a $\pm 10°$ band around the galactic plane. This sample was compiled before the launch of Fermi and was expected to contain a large fraction of the extragalactic sources that would be detected by the Fermi Large Area Telescope (LAT). The sample then used in the OVRO 40m monitoring program included all CGRaBS objects with a declination $> -20°$, a total of 1158 sources. Figure 1.2 shows the sky positions of these sources. Each of these sources has been continuously monitored (twice a week, where possible) since the program started, with publication-quality data available since 1st January 2008. An example light curve of the source J1415+1320 is shown in Figure 1.3.
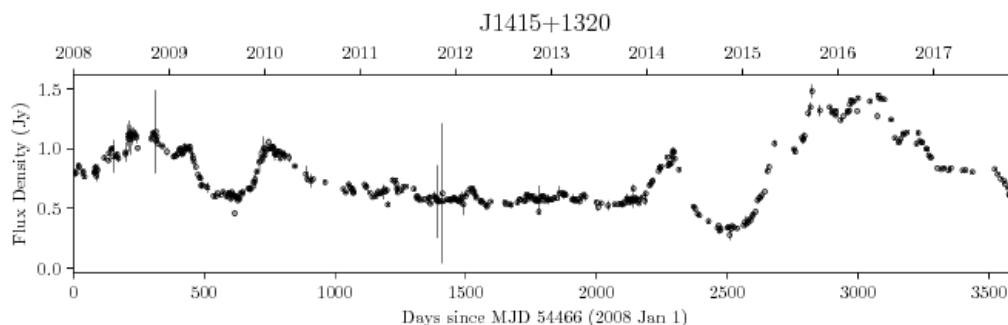
Figure 1.3: Light curve of the blazar J1415+1320 showing measurements between January 2008 and December 2017, the same span of data that was available when SAV activity was first identified.

Once Fermi-LAT had begun its operation, the Fermi-LAT collaboration routinely published source catalogues, each with an associated study of the blazars and other AGN. After 3 months of data had been collected, the LAT Bright Source List was released (0GFL, Abdo et al. 2009*b*) alongside the LAT bright AGN sample (LBAS, Abdo et al. 2009*a*). The first Fermi-LAT catalogue (1FGL, Abdo et al. 2010*a*) and first LAT AGN catalogue (1LAC, Abdo et al. 2010*c*) were released using 11 months of data, and the second Fermi-LAT catalogue (1FGL, Abdo et al. 2010*a*) and second LAT AGN catalogue (2LAC, Ackermann et al. 2011) were released using 24 months of data. With more data, a better understanding of the instrument and improved methods to analyse the data, each new catalogue iteratively improved on its predecessor. As such, the original OVRO 40m sample was augmented to include all of the original CGRaBS sources plus the blazars in the 1LAC and 2LAC "clean samples", which contain sources detected with high statistical significance and that are not affected by analysis issues like multiple associations (Richards 2012). Today, the OVRO 40m program has grown to regularly monitor over 1800 blazar sources.

1.2.2   Extreme Scattering Events

In early 2017, working with the OVRO 40m data set, Vedantham et al. (2017*b*) (V2017B henceforth) identified particular transient events of unknown origin which resulted in time-symmetric U-shaped features in the light curve of the blazar J1415+1320 (fig. 1.3). It was thought that these unusual variations were likely due to some form of lensing. Lensing can be caused by various mechanisms, such as rays being deflected by spatial fluctuations in the refractive index of the medium of propagation. In terms of interstellar ray propagation, electron density fluctuations in interstellar plasma provide the refractive index variation. One typical signature of lensing is the presence of multiple magnified images of the lensed source. Even if individual images cannot be resolved (as is the case with the OVRO 40m telescope, due to its low, arcmin-scale angular resolution) flux-density variations due to the relative motion in the source-lens-observer system may be observed in light curves. An example of how this can occur is shown in fig. 1.4, which demonstrates how relative motion in such a lensing system may produce a U-shaped feature like the ones identified in the OVRO 40m data. Moreover, it must be the case that, on average, lensing yields time-symmetric features. A manual search was undertaken to find such events in the OVRO data, which revealed several possible candidates. One source in particular, J1415+1320 (see fig. 1.3), stood out for a number of reasons. Firstly, it shows recurring, highly symmetric, U-shaped features in its light curve, which seemed like they may be chromatic (frequency dependent). V2017B illustrates that a collation of multi-frequency data from the Metsähovi radio observatory, the Sub-Millimeter Array (SMA), and OVRO shows that the U-shaped events persist even in the mm-wave band.
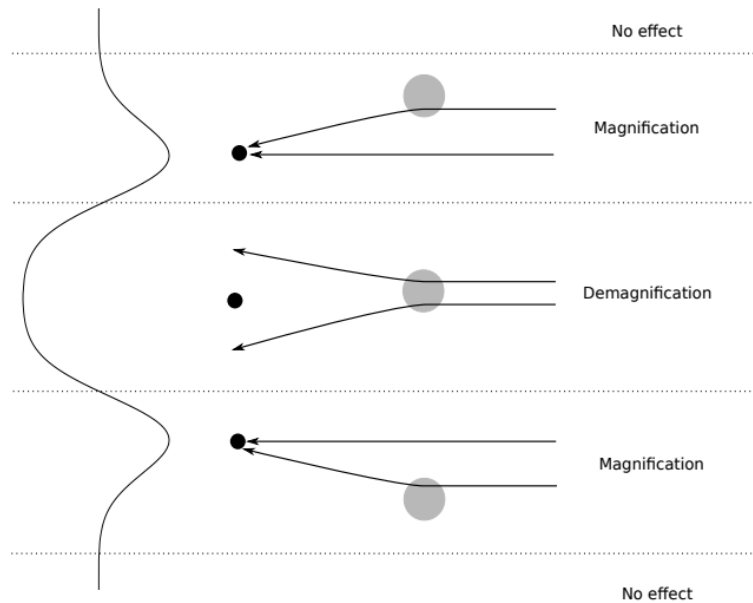
Figure 1.4: A cartoon depiction (not to scale) of the effect of a diverging plasma lens on a background source. Black circle is the Earth, and the gray circle is the lens. Relative motion in the source-lens-observer system leads to symmetric magnification events that bracket a de-magnification trough as shown in the U-shaped curve on the left (From V2017B).

It was originally thought that the features seen in J1415+1320 and several other blazar sources could be explained as Extreme Scattering Events (ESEs) (Fiedler et al. 1987). Upon initially being reported, these highly symmetric, chromatic events were explained to be caused by lensing due to intervening plasma. The plasma lenses were inferred to possess high electron column densities ($N_e \sim 10^{17} cm^{-2}$) over relatively small transverse extents ($\lesssim 10$AU). Such lenses would therefore be highly over-pressurized (density of $n_e \sim 10^3 cm^{-3}$), and therefore cannot exist in pressure balance with the ambient interstellar medium ($n_e \sim 0.03 cm^{-3}$). To explain this, two classes of models were put forth. In the first, the lens is the photo-ionized 'skin' of an underlying cool, self-gravitating cloud in the galactic halo (Walker & Wardle 1998). It was found, however, that such objects would

form a considerable fraction of Galactic dark matter. It therefore seems somewhat unlikely that dark matter could exist in this form and not have previously been detected. The second class of models invokes highly elongated plasma sheets seen edge-on, such that the required column density can be obtained with comparably low over-densities (Goldreich & Sridhar 2006). Whether plasma sheets with such large axial ratios indeed exist remains unknown, and ESEs have remained an unsolved mystery for over 3 decades (V2017B).

The implied ESE rate based on previous surveys is about $7 \times 10^{-3} \text{source}^{-1} \text{year}^{-1}$ (Fiedler et al. 1987). From this we should therefore see an ESE event from J1415+1320 every 140 years on average. Instead, four ESE-like events were observed in about 25 years. J1415+1320 showed this ESE-like variability down to a wavelength of 0.1cm, which implies a very strong and compact plasma lens. Additionally, it is debatable as to whether the events seen in its light curve are truly chromatic. In fact, the observed magnitude of the (de-)magnification was almost achromatic over a factor of 20 in wavelength. The question of the chromaticity of these events is explored in chapter 5. Overall, however, these inconsistencies indicate that the plasma lensing hypothesis is not a good fit to the events observed here.

The plasma lensing interpretation was critically examined in V2017B, in which a range of possible lensing geometries were considered, with the authors presenting three key conclusions. First, any extragalactic plasma lens would require unrealistically high axial-ratios (width/depth) of $\gtrsim 10^3$ to yield enough free-free opacity. Second, the lens would have to be placed well beyond the galactic disk, since there is an absense of asymmetry one would expect to see from the Earth's orbit around the Sun. The placement of this putative lens would be in a region devoid

of dense interstellar material. Finally, the U-shaped events seen are not well described as de-magnification events. The model light curves provide poor fits to the data, since the flux density of the source at the bottom of the 'trough' (the centre of the U-shaped dip) is comparable to that away from the U-shaped events. V2017B therefore rejects the ESE hypothesis for J1415+1320. The evidence, however, does still seem to indicate that these events are caused by some type of lensing. The only other lensing mechanism known is gravitational lensing, which, being achromatic, may fit the events we see.

1.2.3 The SAV Hypothesis

With the rejection of the ESE hypothesis, the next logical step is to investigate an alternative lensing mechanism: gravitational lensing. Vedantham et al. (2017*a*) (V2017A henceforth) investigated this proposal. They proposed that SAV is caused by the modulation of the flux density of a luminal or superluminal compact emission region as it traverses the magnification pattern cast by an intervening gravitational lens. The proposed lensing mechanism is similar to microlensing of stars by stellar-mass lenses (Alcock et al. 1993), but the lenses would need to be in the milli-lensing mass range $\sim 10^3 - 10^6 M_\odot$ with a projected surface mass density of $\gtrsim 10^4 M_\odot \mathrm{pc}^{-2}$. Potential lens candidates with these properties include the dense cores of globular clusters and molecular clouds, and massive black holes. The lenses likely consist of multiple components that are projected close to the line of sight to the source, but are not necessarily gravitationally bound to each other. Figure 1.5 demonstrates how a gravitational lens can result in U-shaped flux density variation in a similar manner to the plasma lensing case shown in fig. 1.4.
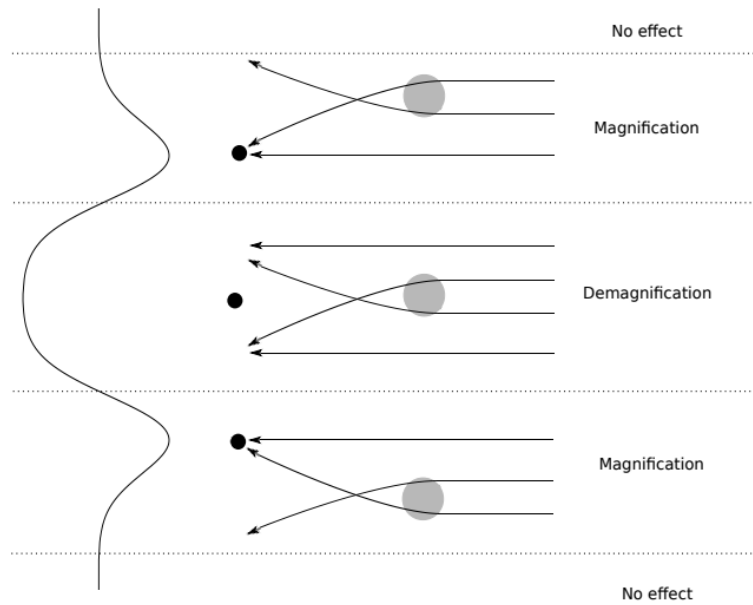
Figure 1.5: A recreation of fig. 1.4 for the case of a gravitational lens, again showing how relative motion in the source-lens-observer system can lead to symmetric magnification events that bracket a de-magnification trough as shown in the U-shaped curve on the left.

V2017A established that we can distinguish two types of U-shaped events within the context of gravitational lensing: a "volcano" type in which the first spike displays a slow rise and a fast decline and the second spike shows a fast rise and slow decline (SRFD-FRSD); and a "crater" type in which a fast rise, slow decline is followed by a slow rise, fast decline (FRSD-SRFD). This is demonstrated in fig. 1.6. Initially, in order to identify these features within the data from the OVRO light curve catalogue, two of the co-authors independently examined a sample of 981 high quality light curves which were determined to have a high enough signal-to-noise ratio for symmetrical U-shaped features to be identified. Both experts selected all those U-shaped features deemed to have sufficient symmetry to be of interest. One co-author picked out 23 U-shaped features, and the other 25 U-shaped features. Out of these selected features, only ten were selected by both authors.

These were accepted as candidates. The large disparity of possible candidates between the two experts invites the possibility of a less-subjective approach for identifying SAV in blazar light curves. This idea is explored in chapter 3. The ten selected features were seen across the light curves of seven objects, of which one had two features, and another had three. From this we can gather that the U-shaped features in the OVRO light curves are rare. We can determine the random probability of a U-shaped feature in one of the 981 light curves looked at is 1.02%. J1415+1320 stood out amongst these seven objects, having two clear, isolated U-shaped features relative to the rest of the light curves (over the course of 8 years). It is highly unlikely that the U-shaped features observed in this source are random events. V2017A identified four events in the light curves of J1415+1320 over the course of 27 years. Assuming these events are randomly distributed throughout the light curves of all sources, we can calculate a probability of $\sim 4 \times 10^{-6}$ that three or more events occur in one object in a 27 year time window. It is very likely, therefore, that these are either unusual intrinsic features of the objects themselves, due to propagation effects along the line of sight, or a combination of both.
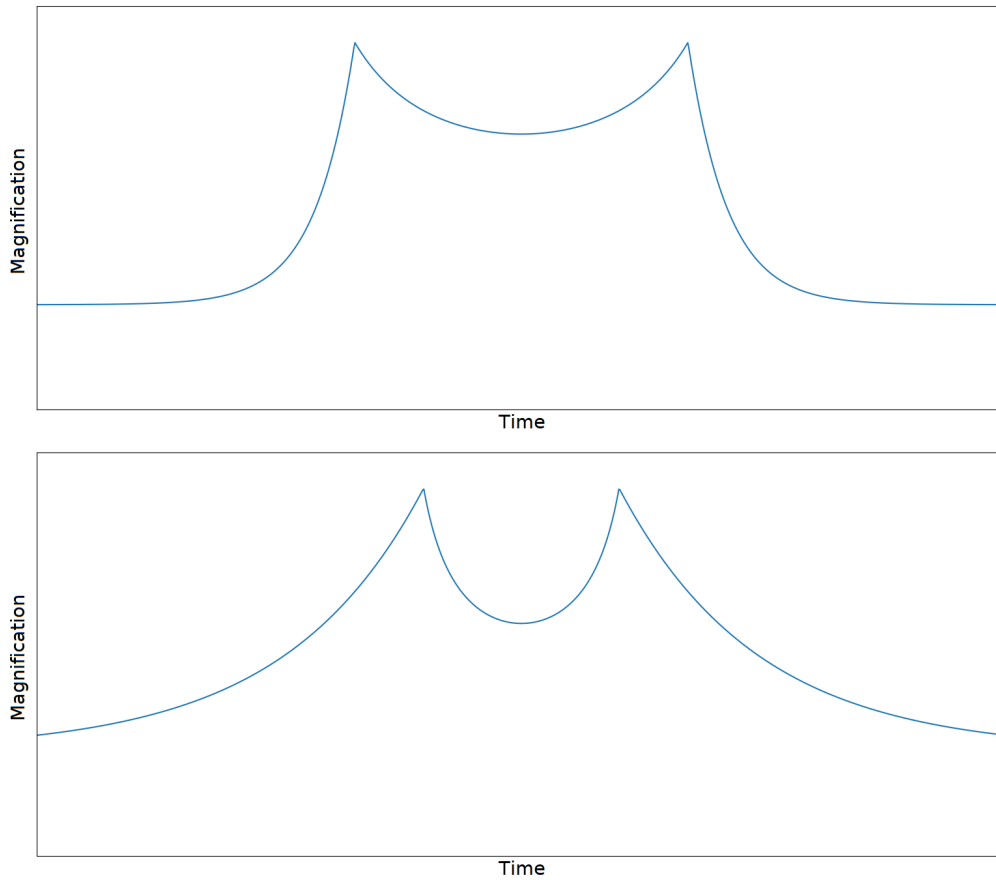
Figure 1.6: Curves emphasising the two types of U-shaped events described in V2017A. The top plot shows a FRSD-SRFD 'crater' type event and the bottom plot shows a SRFD-FRSD 'volcano' event.

1.2.3.1   Gravitational Lensing

Schneider et al. (1992) give a description of a general Gravitational Lensing system, which is summarised in fig. 1.7. Consider the source sphere S_s, i.e. a sphere with radius D_s (i.e. the distance to the source), centred on the observer O, and, the corresponding deflector sphere S_d with radius D_d, i.e. distance to the centre of the lens L. In addition, consider the observed sphere S_o, which is the apparent "sky" of the observer. On S_o the source would have apparent angular position β,

if it were the case that the light rays from the source S were not influenced by the gravitational field of the deflector. The lens, however, does bend the light rays, so the straight line SO is no longer a physical ray path. Rather, the light rays are curved near S_d. One such ray $SI'O$ is drawn along with its approximation SIO, consisting of the two asymptotes of the real ray, SI and IO. The angle $\hat{\alpha}$ between these asymptotes is the deflection angle caused by the matter distribution L. The observer would therefore see the source at position θ.
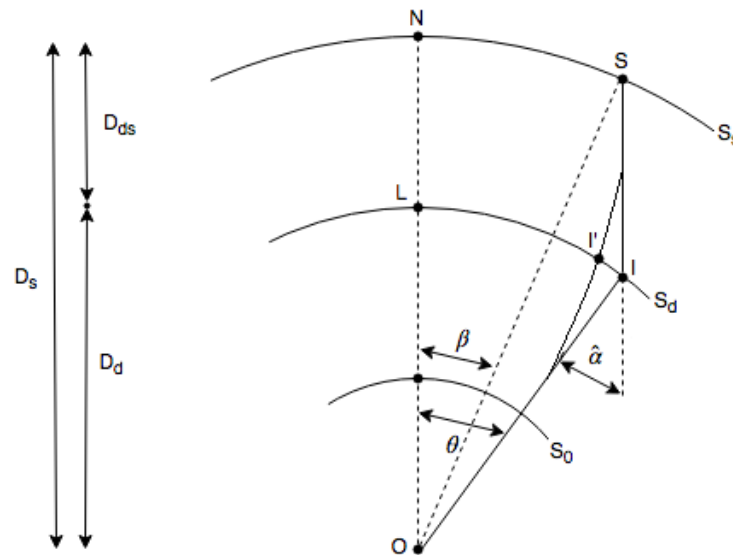
Figure 1.7: A General GL system; the 'centre' of the lens is at L, and the line through L and the observer O is the 'optical axis'. Relative to that, the source S has an undisturbed angular position β. A light ray $SI'O$ from the source is deflected by an angle $\hat{\alpha}$, so that an image of the source is observed at position θ. (Recreated from Schneider et al. (1992), p30, Fig. 2.2.)

In an astrophysical context, for all of the cases we are interested in, the deflection angles are very small. Therefore, only a small cone around the optical axis OL needs to be considered in general. In this case, we can replace the three spheres in our lensing description with their corresponding tangent planes (the source plane,

lens plane and observed plane). This is depicted in fig. 1.8. The separation of the light ray from the optical axis, *LI*, can be described as a two-dimensional vector ξ in the lens plane. Again, since the deflection angle $\hat{\alpha}$ is small, the distinction between I and I' is unnecessary, so the angles θ and $\hat{\alpha}$ can be described as angular vectors in the tangent plane S_o.
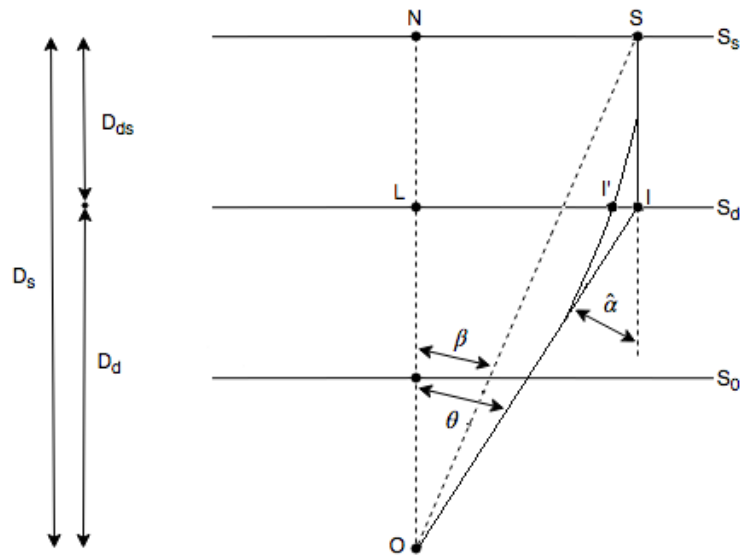
Figure 1.8: A modification of the system shown in fig. 1.7, where the spheres are each approximated by their tangent planes due to the deflection angles being small.

From this description, we can derive a relation between the source position (described by the unlensed position angle β) and the positions of the images $\theta = \xi/D_d$ of the source (the lens equation):

$$\beta = \theta - \frac{D_{ds}}{D_s}\hat{\alpha}(\xi) \tag{1.1}$$

or, in terms of the distance $\eta = D_s\beta$ from the source to the optical axis:

$$\eta = \frac{D_s}{D_d}\xi - D_{ds}\hat{\alpha}(\xi) \tag{1.2}$$

It is often useful to rewrite this, however, in dimensionless form. If we define a length scale ξ_0 in the lens plane and a corresponding length scale $\eta_0 = \xi_0 D_s/D_d$ in the source plane. Then, we can define the dimensionless vectors

$$x = \frac{\xi}{\xi_0} \; ; \; y = \frac{\eta}{\eta_0}. \tag{1.3}$$

With these definitions, we can rewrite eq. (1.2) as

$$y = x - \alpha(x), \tag{1.4}$$

where

$$\alpha(x) = \frac{D_d D_{ds}}{\xi_0 D_s}\hat{\alpha}(\xi_0 x). \tag{1.5}$$

In the case of an axially symmetric lens, the matter within a disc of radius x located at the centre of mass contributes to the deflection at the point x as if it were located at that centre, and the matter outside does not contribute, in a similar manner to gravitational forces of spherical mass distributions in three dimensions. Schneider et al. (1992) show in chapter 8 that

$$\alpha(x) \equiv \frac{m(x)}{x}. \tag{1.6}$$

where $m(x)$ is the dimensionless mass within a circle of radius x, related to the mean surface mass density inside the circle $\bar{\kappa}(x)$ by

$$m(x) = x^2 \bar{\kappa}(x). \tag{1.7}$$

This allows us to update eq. (1.4) to

$$y = x - \frac{m(x)}{x}. \tag{1.8}$$

In addition to describing a general lensing situation, Schneider et al. (1992) provide a number of gravitational lensing models, which form the basis of the SAV identification procedure being developed. The simplest of these models is the Schwarzschild lens. This is a lens formed by a single "point mass" M. In this case, we have $m(x) = 1$, giving the lens equation

$$y = x - \frac{1}{x}. \tag{1.9}$$

Further to this, we can obtain what is known as the 'Chang-Refsdal lens'. This involves the perturbation of the gravitational field of a point mass, for example if a star in a galaxy acts as a lens, the field of the galaxy will perturb that of the star. As a two length-scale problem, the lensing by stars is generally referred to as microlensing, whereas the lensing by the galaxy as a whole is distinguished as macrolensing. In the case of the Chang-Refsdal lens, we obtain an updated version of eq. (1.9). For a point mass M, again using the length scale ξ_0, the lens equation becomes

$$y = x - \frac{x}{|x|^2} - \begin{pmatrix} \kappa_c + \gamma & 0 \\ 0 & \kappa_c - \gamma \end{pmatrix} \cdot x \tag{1.10}$$

where "·" denotes matrix multiplication and κ_c and γ are the surface mass density and the shear (a measure of the distortion of images by the lens), respectively, of the macrolens at the location of the point mass. As discussed in section 1.2.3, there are two types of U-shaped features - volcanoes and craters. As explained in V2017A, binary lenses typically yield a six-cusped caustic pattern which has sufficient complexity to generate a large range of light curve features including both crater and volcano type SAVs. Here, a 'caustic pattern' describes the 'envelope' of light rays projected onto the source plane. The patterns feature 'folds' (lines) and 'cusps' (points) which separate regions of different image multiplicity. Sources located in the vicinity of the folds/cusps will be highly magnified.

Keeping this in mind, we are therefore looking for a lensing model involving two masses. We can generalise the lens equation from the Chang-Refsdal model to account for an arbitrary number of point masses. Consider a reference mass M, and N point masses with mass m_i and position x_i measured in units of ξ_0. Equation (1.10) then becomes:

$$y = \begin{pmatrix} 1 - \kappa_c + \gamma & 0 \\ 0 & 1 - \kappa_c - \gamma \end{pmatrix} \cdot x - \sum_{i=1}^{N} \frac{m_i}{|x - x_i|^2}(x - x_i). \qquad (1.11)$$

In the case of a binary mass system, we obtain

$$y = x - \frac{x}{|x|^2} - q\frac{x - d}{|x - d|^2} - \begin{pmatrix} \kappa_c + \gamma & 0 \\ 0 & \kappa_c - \gamma \end{pmatrix} \cdot x. \qquad (1.12)$$

as stated in V2017A, where our two point-like objects are of mass 1 and q, located respectively at $0 = [0,0]$ and $d = [d,0]$. In this case we are employing the usual

dimensionless units, where all angles are expressed in units of the Einstein Radius, θ_E.

1.3   This Thesis

The remainder of this thesis focuses on concepts surrounding Symmetric Achromatic Variability and Bayesian Inference. Chapter 2 introduces Bayesian statistics, followed by Bayesian inference. We then explore modern computational methods for performing Bayesian inference, with a particular focus on nested sampling. The subsequent chapters all consist of my own novel research. Chapter 3 then introduces an automated approach for identifying SAV in blazar light curves using nested sampling. I first propose a model capable of capturing the long- and short-term variations seen in the light curve of a lensed blazar source, then describe a fitting procedure which uses nested sampling to fit the proposed model to data. I then build upon this fitting procedure to describe a method for simulating artificial blazar light curves using real data from the OVRO 40m telescope. The chapter closes with a discussion of the results of applying the fitting procedure to real data. Chapter 4 features an exploration of a 'transdimensional' nested sampling approach in which the dimensionality of the problem is not fixed and the number of model parameters is allowed to vary freely. I introduce a framework for exploring a transdimensional space, then introduce some toy problems which allow us to analyse the current approach. Finally, chapter 5 describes an ongoing campaign to observe the blazar J1415+1320 using e-MERLIN. The observations are carried out at both C- and L-band, with the hopes of observing a future SAV event at a lower frequency than in the OVRO 15GHz light curves. This would provide very

strong evidence as to whether these events are truly achromatic, which would have significant implications for the validity of the SAV hypothesis.

Chapter 2

Bayesian Inference

Inference is a process central to our daily lives. As events occur day-to-day, we take in information about what we see and hear, and, using past experiences, we reach conclusions which help us interpret the world around us. If one suddenly notices smoke inside a building, we might naturally conclude that there is a fire and take actions accordingly. Dark clouds accumulating overhead might drive us to take an umbrella to work. As scientists, we are more specifically concerned with making inferences about data. Having completed an experiment and collected results, are our hypotheses affected? If so, how? Bayesian Inference provides us with a statistical framework by which we can answer these questions in a meaningful and intuitive manner. This chapter introduces Bayesian statistics, tools and methods, including a description of two popular approaches to Bayesian Inference: Markov Chain Monte Carlo and nested sampling.

2.1 Frequentist vs Bayesian Statistics

When discussing probability, it is useful to do so within the context of one of the two rivaling schools of thought which exist within the field of statistics, Bayesians and frequentists. Both paradigms are well-founded within the rules of probability theory, with the key difference between them being how they view the definition of probability. A frequentist would understand the probability of an event occurring as

> "The relative frequency of the event in the limit of an infinite number
> of independent trials."

This definition fits nicely with the problems we often encounter when first introduced to probability and statistics in school. For example, if one was to roll a fair 6-sided die a sufficiently large number of times, one would find that roughly $\frac{1}{6}$ of all rolls would result in a 6, with this relative frequency moving closer to being exactly $\frac{1}{6}$ as the number of rolls increases.

While this definition seems perfectly intuitive in these situations, what if we wanted to assess the probability that it will rain tomorrow? The frequentist definition implies that this question should be framed as "What is the relative frequency of rain tomorrow in the limit of an infinite number of tomorrows?". Suddenly this definition does not feel as intuitive, since it is immediately clear that this situation is unrepeatable. Tomorrow will only occur once, so are we now considering some infinite number of alternate realities in order to determine the relative frequency of this event? Instead, it might be more meaningful to consider e.g. the relative frequency of rain across all previously recorded days in which the atmospheric conditions (wind speed, air pressure, etc.) were exactly as they are expected to

be tomorrow. Alternatively, we could find the relative frequency of rain across all days with tomorrow's date from previous years. These approaches allow us to use historical data to provide a prediction of our desired probability, but within the frequentist framework this will only ever be an estimate and we cannot give a true answer to the question.

In addition to being unable to handle unrepeatable events, Trotta (2008) provides additional arguments as to why the frequentist definition is unsatisfactory. Firstly, the definition is circular, in the sense that it assumes that repeated trials have the same probability of outcomes, which is exactly the definition of probability we wanted to define in the first place. Additionally, the definition only holds for an infinite sequence of repetitions, which will we will never have in practice. This leads to awkward situations where complicated corrections have to be made to account for a small sample size.

These issues give way to the Bayesian school of thought which does not suffer from these problems. Bayesian statistics gives us a different definition of probability:

"A measure of the degree of belief in a proposition."

With this definition, we can obtain a solution to both of the problems posed above. Given a fair dice, we can determine the probability of rolling a 6 as our belief that such an event will occur. Since we believe this to be a fair dice, each result is equally likely to appear, so we can assign each result an equal probability of $\frac{1}{6}$. When asked "Will it rain tomorrow?", we are no longer worried about the question being unrepeatable. It is perfectly possible to quantify one's degree of belief that it will rain, but this raises an important point: Bayesian statistics is viewed by

some as being subjective, since one individual's degree of belief in an event may be different than another's. However, as explained by Sivia & Skilling (2006), this is not truly subjectivity: it is simply the case that one's degree of belief is conditional on the information at one's disposal, and that conditioning should be stated explicitly.

2.2 Bayes' Theorem and Marginalisation

One critical aspect of the Bayesian approach to probability is that one's belief in a given proposition is not necessarily fixed. A person driving to an appointment may expect to arrive on time, based on when they set off, but upon seeing a sign for roadworks along their route, they might begin to anticipate that they are going to be late. We often change our minds in light of new information, and we therefore need some way to express how our belief in a proposition might change once new data has been acquired.

In addition to providing an excellent introduction to Bayesian data analysis, Sivia & Skilling (2006) show that we achieve the sum and product rules as a consequence of following the axioms of probability (Cox 1946). The sum rule is given as:

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

where $\overline{X}$ denotes the proposition that X is false, the vertical bar '|' means 'given' (i.e. everything to the right of this symbol is taken as being true), and I is the relevant background information, since there is no such thing as an absolute prob-

ability. Conditioning on I is often omitted from calculations to reduce algebraic cluttering, but it is important to acknowledge its existence. The product rule states:

$$P(X,Y|I) = P(X|Y,I)P(Y|I), \tag{2.2}$$

where the comma is read as the conjunction 'and'. The 'and' conjunction is commutative, i.e. $P(X,Y|I) = P(Y,X|I)$, so by applying the product rule we get:

$$P(X|Y,I)P(Y|I) = P(Y|X,I)P(X|I), \tag{2.3}$$

which we can rearrange to give Bayes' theorem:

$$P(X|Y,I) = \frac{P(Y|X,I)P(X|I)}{P(Y|I)}. \tag{2.4}$$

This simple formula describes how one's prior belief in X (given the background information) is updated given some new information Y. The quantities in Bayes' theorem each have formal names. $P(X|I)$ is called the *prior*, since it expresses our knowledge about X prior to considering the new information Y. This is augmented by the *likelihood* $P(Y|X,I)$ to give the *posterior* $P(X|Y,I)$. This calculation is weighted by the *evidence* $P(Y|I)$, which acts as a normalisation constant and does not depend explicitly on X. If we expand $P(X,Y|I)$ using eq. (2.2), we get:

$$P(X,Y|I) = P(Y,X|I) = P(Y|X,I)P(X|I). \tag{2.5}$$

If we add the expression $P(X,\overline{Y}|I)$ to both sides, while similarly noting that, $P(X,\overline{Y}|I) = P(\overline{Y}|X,I)P(X|I)$, we get:

$$P(X,Y|I) + P(X,\overline{Y}|I) = \left[P(Y|X,I) + P(\overline{Y}|X,I)\right] P(X|I). \tag{2.6}$$

The sum rule in eq. (2.1) tells us that the quantity in square brackets is equal to unity, so we are left with:

$$P(X|I) = P(X,Y|I) + P(X,\overline{Y}|I). \tag{2.7}$$

This equation tells us that the probability that X is true, irrespective of whether or not Y is true, is equal to the sum of the probability that both X and Y are true and the probability that X is true and Y is false.

Rather than having a proposition Y, with its negative counterpart $\overline{Y}$, we could instead have a set of alternative possibilities: $Y_1, Y_2, \cdots, Y_M = \{Y_k\}$. If this is a mutually exclusive and exhaustive set of possibilities, i.e. exactly one of the M propositions must be true and the others must all be false, then we can generalise eq. (2.7) to give:

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

We can extend this further by taking eq. (2.8) to the continuum limit, where Y is now taken to be a continuous variable which has a value in some given range:

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

This result is known as marginalisation, which is a very useful tool for Bayesian Inference, as it enables us to 'marginalise over' parameters we are uninterested in, removing them from consideration.

2.2.1   A Bayesian Example

Here we will consider an example of how Bayes' theorem may be applied to a problem, as a way of consolidating the above information. Consider a disease which is known to affect 0.1% of the population. Say there is a test which is known to have a sensitivity (True Positive rate) of 85% and a specificity (True Negative rate) of 95%. Let D be the event that a given person has the disease and let T be the event that, after being tested, the person's test result comes back positive. We therefore know:

$$P(D) = 0.001,$$

$$P(T|D) = 0.85,$$

$$P(\overline{T}|\overline{D}) = 0.95.$$

Making use of the sum rule from eq. (2.1), we also therefore know:

$$P(\overline{T}|D) = 0.15,$$

$$P(T|\overline{D}) = 0.05.$$

Let's say a patient is given the test, but it is *a priori* unknown whether they have the disease. If the test comes back positive, how confident should we be (i.e. what is our degree of belief) that the patient actually has the disease? We would express this quantity as $P(D|T)$ and using Bayes' theorem from eq. (2.4), we have:

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)}.$$

The only term on the right hand side which we don't yet have an expression for is $P(T)$, but by marginalising over D (eq. 2.7) and applying the product rule using and (eq. 2.2), we find:

$$\begin{aligned} P(D|T) &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\overline{D})P(\overline{D})} \\ &= \frac{0.85 \cdot 0.001}{(0.85 \cdot 0.001) + (0.05 \cdot 0.999)} \\ &\approx 0.017. \end{aligned}$$

At first glance this result may seem peculiar. The test seems fairly accurate, so why does a positive result yield such a low probability of actually having the disease? The result does make sense after more thought, however. Consider a population of $1,000,000$ people. We would expect only 1000 of these people to have the disease, and only 850 of those to test positive for it. Conversely, there are $999,000$ people who do not have the disease, out of whom $49,950$ would test positive. This means that we would expect $50,800$ people to test positive for the disease, with only 850 of those people actually having it, or $\sim 1.7\%$.

Due to the subjective nature of Bayesian statistics, it is important to note that this calculation only holds true for the given prior. Our prior belief that the patient has the disease might have been different if we knew more about their medical history, for example. In fact, Bayes' theorem even allows us to consider the case where someone has already been administered one test, and, having tested posit-

ive, is tested once more. Assuming the two tests are independent from one another, we can use the posterior from the previous calculation and use it as the prior for the new calculation. One way to think of this is that the previous positive test result is absorbed into the background information *I* (which we are currently omitting from our calculations, as explained previously). The probability of having the disease after testing positive a second time is therefore:

$$
\begin{aligned}
P(D|T,T) &= \frac{P(T|D,T)P(D|T)}{P(T|D,T)P(D|T) + P(T|\overline{D},T)P(\overline{D}|T)} \\
&= \frac{0.85 \cdot 0.017}{(0.85 \cdot 0.017) + (0.05 \cdot 0.983)} \\
&\approx 0.227.
\end{aligned}
$$

So upon receiving a second positive test result, the probability of having the disease increases from $\sim 1.7\%$ to $\sim 22.7\%$. This stands to reason, as two independent, positive results would naturally increase our belief that the person has the disease over just a single positive result.

2.3   Parameter Estimation and Model Selection

Scientists are often charged with constructing a model $\mathcal{M}$, usually defined by some parameters Θ, in order to explain some data $\mathcal{D}$. By applying the Bayesian techniques introduced in this chapter, we can use the data to make inferences about a given model and its parameters. This type of analysis is typically divided into two types of problems: parameter estimation and model selection.

2.3.1 Parameter Estimation

Parameter estimation is a very common data analysis problem. Given some data and a model describing it, what can we learn about its parameters? In this context, Bayes' theorem (eq. 2.4) defines a relationship whereby our initial knowledge of the model parameters is updated in light of the data:

$$\mathcal{P} = P(\Theta|\mathcal{D},\mathcal{M}) = \frac{P(\mathcal{D}|\Theta,\mathcal{M})P(\Theta|\mathcal{M})}{P(\mathcal{D}|\mathcal{M})} = \frac{\mathcal{L} \times \pi}{\mathcal{Z}}. \tag{2.10}$$

The posterior $\mathcal{P} = P(\Theta|\mathcal{D},\mathcal{M})$ constrains the model parameters after taking the data into account. It is dependent on the three quantities: the likelihood $\mathcal{L} = P(\mathcal{D}|\Theta,\mathcal{M})$, which is the probability that a given choice of parameter values produced the data; the prior $\pi = P(\Theta|\mathcal{M})$, our initial belief in the choice of parameter values; and the evidence $\mathcal{Z} = P(\mathcal{D}|\mathcal{M})$, a normalisation constant which is the probability of observing the data, conditioned on the model, but irrespective of the parameter values. The evidence is computed by:

$$\mathcal{Z} = P(\mathcal{D}|\mathcal{M}) = \int P(\mathcal{D}|\Theta,\mathcal{M})P(\Theta|\mathcal{M})\mathrm{d}\Theta = \int \mathcal{L}(\Theta)\pi(\Theta)\mathrm{d}\Theta. \tag{2.11}$$

Here, we marginalise out the parameters, hence the alternative name for this quantity, the *marginal likelihood*.

2.3.2 Model Selection

Another common problem encountered by scientists occurs when more than one model exists which could explain some data. In this case, we are interested in

determining our relative belief in each model, given the observed data. In order to do so, we again make use of Bayes' theorem (eq. 2.4). For a given model, $\mathcal{M}_i$:

$$P(\mathcal{M}_i|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M}_i)P(\mathcal{M}_i)}{P(\mathcal{D})}. \tag{2.12}$$

On its own, the probability of (i.e. our degree of belief in) a given model doesn't really mean much, but when considering several models $\{\mathcal{M}_1, \mathcal{M}_2, \cdots\}$, the degree of belief in each model is given by:

$$P(\mathcal{M}_i|\mathcal{D}) = \frac{\mathcal{Z}_i \pi_i}{\sum_j \mathcal{Z}_j \pi_j}. \tag{2.13}$$

Here, π_i indicates the prior probability of the model $\mathcal{M}_i$ before considering the data. In order to compare which of two models $\mathcal{M}_i$ and $\mathcal{M}_j$ *a posteriori* describes the observed data better, we can find the posterior odds ratio:

$$\mathcal{P}_j^i = \log\left(\frac{P(\mathcal{M}_i|\mathcal{D})}{P(\mathcal{M}_j|\mathcal{D})}\right) = \log\left(\frac{\mathcal{Z}_i}{\mathcal{Z}_j}\right) + \log\left(\frac{\pi_i}{\pi_j}\right). \tag{2.14}$$

We often have no reason to believe *a priori* that one model is any more probable than another, in which case their prior probabilities would be equal and the final term in eq. (2.14) would disappear, leaving us with Bayes factor, a more commonly used metric:

$$\mathcal{B}_j^i = \ln\left(\frac{\mathcal{Z}_i}{\mathcal{Z}_j}\right). \tag{2.15}$$

Bayes factor is a useful way of denoting our degree of belief in one model relative to another. If it is positive, we favour $\mathcal{M}_i$ over $\mathcal{M}_j$ and vice versa if it is negative. Jeffreys (1961) provides an often-cited scale for interpreting the Bayes factor,

Table 2.1: The "Jeffreys' scale" as reproduced from Trotta (2008). This scale indicates how much or how little $\mathcal{M}_i$ is favoured over $\mathcal{M}_j$, as seen in the right-most column. The probability column refers to the posterior probability of the favoured model, assuming non–committal priors on the two competing models, i.e. $\pi_i = \pi_j = 1/2$ and that the two models exhaust the model space, $P(\mathcal{M}_i|\mathcal{D}) + P(\mathcal{M}_j|\mathcal{D}) = 1$.

$|\log \mathcal{B}^i_j|$	Probability	Strength of evidence
< 1.0	< 0.750	Inconclusive
1.0	0.750	Weak evidence
2.5	0.923	Moderate evidence
5.0	0.993	Strong evidence

illustrated in table 2.1. While this scale can be useful for comparing the relative strength of two or more models, it should not be viewed in isolation as a definitive measure of the performance of a model. It is perfectly possible that another, undiscovered model exists which much better describes the data, and simply calculating Bayes factor between known models might lead one to believe one of those models is best instead.

2.3.3   Kullback-Leibler divergence and Occam's razor

Occam's razor is a principle which can be stated as "Accept the simplest explanation that fits the data" (MacKay 2003). This principle is naturally incorporated into the Bayesian evidence, which penalises models for unnecessary complexity. Hergt et al. (2021) provide a clear example of this. Consider a Gaussian likelihood, with mean μ and variance σ^2 and having a single parameter $x \in [x_{\min}, x_{\max}]$ with a uniform prior. The Bayesian evidence can then be written

$$\mathcal{Z} = \mathcal{L}(\mu) \times \frac{\sigma\sqrt{2\pi}}{x_{\max} - x_{\min}}. \tag{2.16}$$

Here, the first right-hand term is the maximum likelihood, which will only in-

crease with additional parameters. The second term incorporates the ratio of posterior to prior uncertainty. The posterior uncertainty σ is generally smaller than the prior uncertainty $(x_{\max} - x_{\min})$, meaning that the second term penalises the given model for each additional parameter, and thus Occam's razor is upheld. A more complex model will only be favoured if it significantly improves the maximum likelihood, outweighing the Occam penalty. Furthermore, Hergt et al. (2021) demonstrate that an estimator of this Occam penalty can be found in the Kullback-Leibler divergence. KL-divergence (sometimes referred to as the relative entropy) is a measure of the overall compression from the prior to the posterior distribution. In this context it is defined as:

$$\mathcal{D}_{KL} = \int \mathcal{P}(\Theta) \ln\left(\frac{\mathcal{P}(\Theta)}{\pi(\Theta)}\right) d\Theta. \tag{2.17}$$

Due to the use of the natural logarithm, it is measured in units of *nats*, the base e equivalent of bits. The relationship between KL-divergence and the Occam penalty becomes clearer when rewriting the log-evidence as

$$\ln\left(\int \mathcal{L}(\Theta)\pi(\Theta)d\Theta\right) = \int \mathcal{P}(\Theta) \ln \mathcal{L}(\Theta) d\Theta - \int \mathcal{P}(\Theta) \ln\left(\frac{\mathcal{P}(\Theta)}{\pi(\Theta)}\right) d\Theta, \tag{2.18}$$

which, as the authors point out, becomes quite straightforward when going from right to left and making use of eq. (2.10). The first term on the right side of the equation encapsulates the fit of the model parameters, whereas the second is the KL-divergence. This is analogous to eq. (2.16), showing that the KL-divergence describes the Occam penalty.

2.4 Existing Computational Methods

When performing Bayesian inference calculations, we are typically concerned with quantities which are defined by integrals over the posterior distribution, such as means, variances or posterior intervals for a quantity of interest. These integrals are usually unfeasible to compute (or even approximate), especially when the problem is high dimensional. If one were to generate a set of samples $\Theta_1, \Theta_2, \cdots$ from the posterior distribution, it is possible to estimate posterior means using the sample mean and so on (Kass et al. 1998). The problem then becomes finding the most efficient way to sample the posterior distribution. There is a wealth of literature detailing methods for tackling this problem, but in this section I will focus on two popular approaches: Markov Chain Monte Carlo and nested sampling.

2.4.1 Markov Chain Monte Carlo

A very popular approach for generating samples is Markov Chain Monte Carlo (MCMC). This class of methods involve conducting a random walk through the parameter space to produce samples forming a Markov chain. A Markov chain is an ordered sequence of random variables $X = X_0, X_1, \cdots X_{n-1}$ where the probability of a particular state for one variable in the chain depends only on the state from the previous step (and not on the states reached in any of the other steps). The goal of MCMC is to construct a Markov chain with a stationary distribution equal to the posterior distribution. Let T be the transition probability matrix for the chain X, that is, $T_{ij} = P(X_{n+1} = j | X_n = i)$ where i and j are states within the state space. Let ψ be a column vector with elements giving the probability that the chain is currently in each state in the space $\psi_i = P(X_i)$. In principal, ψ can be

seen as a probability distribution, however here we are considering a discrete state space and so we represent it as using a vector. ψ is said to be *stationary* if

$$T^\intercal \psi = \psi. \tag{2.19}$$

Note that π is often used to denote the stationary distribution of a Markov chain, however ψ has been substituted to avoid confusion with the prior distribution. By applying transition matrix $T^\intercal \psi$, we update each element ψ_i to give the updated probability of being in that state in light of the probabilities in the transition matrix. If at any step along the Markov chain we find that eq. (2.19) holds, then we know that further applications of T will not alter ψ, which is the stationary distribution for this chain. That is not to say that the state cannot change between steps, but that the probabilities of reaching each state remain the same between each subsequent step. To ensure that the chain will converge to a stationary distribution, the chain must be *ergodic*, meaning it satisfies two conditions: *irreducibility* and *aperiodicity*. Irreducibility means that a sequence of (non-zero probability) transitions exists from any state to any other state. Aperiodicity means the system does not return to the state at fixed intervals. A sufficient (but not necessary) condition to ensure ergodicity is *detailed balance*, meaning that for every pair of states, the probability of transitioning from one state to the other is the same probability of making the reverse transition. Note that π is often used to denote the stationary distribution of a Markov chain, however ψ has been substituted to avoid confusion with the prior distribution. By applying transition matrix $T^\intercal \psi$, we update each element ψ_i to give the updated probability of being in that state in light of the probabilities in the transition matrix. If at any step along the Markov chain

we find that eq. (2.19) holds, then we know that further applications of T will not alter ψ, which is the stationary distribution for this chain. That is not to say that the state cannot change between steps, but that the probabilities of reaching each state remain the same between each subsequent step. To ensure that the chain will converge to a stationary distribution, the chain must be *ergodic*, meaning it satisfies two conditions: *irreducibility* and *aperiodicity*. Irreducibility means that a sequence of (non-zero probability) transitions exists from any state to any other state. Aperiodicity means the system does not return to the state at fixed intervals. A sufficient (but not necessary) condition to ensure ergodicity is *detailed balance*, meaning that for every pair of states, the probability of transitioning from one state to the other is the same probability of making the reverse transition.

Arguably the most common MCMC algorithm in use is the Metropolis-Hastings algorithm (Hastings 1970). The algorithm is designed to construct a chain of samples with transition probabilities such that the chain will converge towards the target distribution (in our case the posterior distribution) as its stationary distribution. By generating a sufficient number of samples, we can therefore be assured that the samples will be correctly drawn according to the posterior. The initial state is determined by choosing an arbitrary point x_0 from the parameter space and a (biased) random walk is conducted to generate the samples which form the steps in the Markov Chain.

To derive the MH algorithm, we first start with the condition of detailed balance. If we let $P(x'|x)$ be the probability of transitioning from state x to x' and let $\mathcal{P}$ be our target distribution, the posterior, detailed balance asserts that

$$P(x'|x)\mathcal{P}(x) = P(x|x')\mathcal{P}(x'), \tag{2.20}$$

which can be re-written as

$$\frac{P(x'|x)}{P(x|x')} = \frac{\mathcal{P}(x')}{\mathcal{P}(x)}. \tag{2.21}$$

We split the transition into two sub-steps; the proposal step and the accept-reject step. We define some proposal distribution $q(x'|x)$ which gives the probability of proposing a transition from state x to state x'. Typically this is often chosen to be a simple distribution, such as a Gaussian, but the exact distribution is left for the analyst to decide on, and is treated as a 'black box' by the MH algorithm. We then define an acceptance distribution $\alpha(x',x)$, which gives the probability of accepting the proposed state x'. We can then write

$$P(x'|x) = q(x'|x)\alpha(x',x). \tag{2.22}$$

Combining this with eq. (2.21), we get

$$\frac{\alpha(x',x)}{\alpha(x,x')} = \frac{\mathcal{P}(x')}{\mathcal{P}(x)}\frac{q(x|x')}{q(x'|x)}. \tag{2.23}$$

All that remains is then to choose a suitable acceptance distribution, such that the ratio in eq. (2.23) is satisfied. The commonly used 'Metropolis choice' is given as:

$$\alpha(x',x) = \min\left(1, \frac{\mathcal{P}(x')}{\mathcal{P}(x)}\frac{q(x|x')}{q(x'|x)}\right). \tag{2.24}$$

In the original Metropolis algorithm (Metropolis et al. 1953), the precursor to the MH algorithm, it was assumed that the proposal distribution $q(x'|x)$ was symmetric, i.e. $q(x'|x) = q(x|x')$, and therefore the acceptance distribution in eq. (2.24)

reduces to

$$\alpha(x',x) = \min\left(1, \frac{\mathcal{P}(x')}{\mathcal{P}(x)}\right). \tag{2.25}$$

The 'Metropolis choice' from the MH algorithm is therefore a more general description, but it can be useful to consider this version of the acceptance distribution if the proposal distribution is symmetric. The full Metropolis-Hastings algorithm is therefore given as follows:

Algorithm 1: Metropolis-Hastings Algorithm

Select an initial state x_0;

Set $t = 0$;

while $t < N$ **do**

    Generate candidate state x' according to $q(x'|x_t)$.;

    Calculate the acceptance probability $\alpha(x',x_t) = \min\left(1, \frac{\mathcal{P}(x')}{\mathcal{P}(x_t)} \frac{q(x_t|x')}{q(x'|x_t)}\right)$.;

    Generate a uniform random variable $u \in [0,1]$;

    if $u \leq \alpha(x',x_t)$ **then**

        Accept the new state and set $x_{t+1} = x'$;

    else

        Reject the new state and set $x_{t+1} = x_t$;

    end

    Set $t = t + 1$

end

Here we have set N as the maximum number of iterations allowed. This is a straightforward approach to terminating the algorithm and it is left up to the analyst to determine an appropriate N for their problem. There are alternative termin-

ation criteria, for which the reader is directed to e.g. Roy (2020).

MCMC methods allow us to generate posterior samples as a way of approximating the posterior distribution, which is suitable for parameter estimation problems. On the other hand, these methods do not provide us with the Bayesian evidence, which is critical for model selection problems. There are methods for estimating either the evidence from the resultant MCMC chains (Chib & Jeliazkov 2001, Heavens et al. 2017), or Bayes' factor (Marin & Robert 2009, Lartillot & Philippe 2006), however in practice these methods are usually computationally inefficient and often inaccurate. This is primarily the case because the likelihood typically peaks in a very small region of the prior, with likelihood values away from the peak being orders of magnitude smaller. This results in MCMC moving quickly away from the extreme regions of the prior volume and spending much more time near the likelihood peak. These low-likelihood regions (most of the prior volume) are therefore undersampled and their contribution to the evidence integral won't be accurate (Feroz 2008). This is not a problem for nested sampling, another Monte Carlo algorithm which reverses the priorities of MCMC, focusing on computation of the evidence, while providing posterior samples as a by-product.

2.4.2 Nested Sampling

Nested sampling (Skilling 2006) provides a method for efficiently computing the Bayesian evidence:

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)d\Theta. \qquad \text{(2.11 revisited)}$$
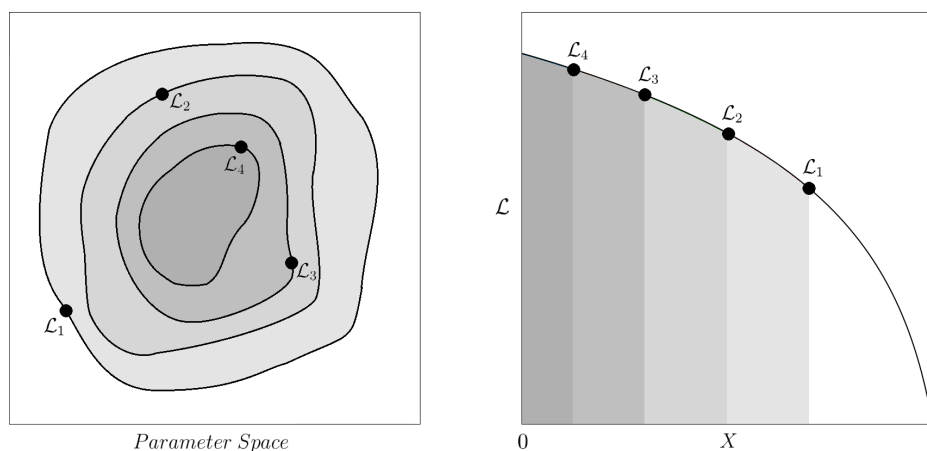
Figure 2.1: Sorted nested likelihood contours enclosing portions of the prior volume. The area under the right-hand curve is equal to the evidence as given by eq. (2.27). Recreated from Skilling (2006).

As indicated previously, numerical analysis quicky becomes impractical for this problem as the dimensionality of Θ increases past a few dimensions. For the nested sampling approach, we instead consider the prior volume X. This quantity can be accumulated from its elements dX in any order, so we define

$$X(\mathcal{L}_*) = \int_{\mathcal{L}(\Theta) > \mathcal{L}_*} \pi(\Theta) d\Theta \qquad (2.26)$$

as the cumulative prior volume covering all likelihood values greater than $\mathcal{L}_*$. For a given likelihood value $\mathcal{L}_*$, this naturally defines an iso-likelihood contour, enclosing a region containing some fraction of the prior, consisting of all coordinates Θ for which $\mathcal{L}(\Theta) = \mathcal{L}_*$. By writing the inverse function of $X(\mathcal{L}_*)$ as $\mathcal{L}(X)$, the evidence can be computed as

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X) dX, \qquad (2.27)$$

a one-dimensional integral over unit range, where $dX = \pi(\Theta)d\Theta$. This process is depicted in fig. 2.1. This transformation removes all complications of gemoetry, topology and dimensionality, making the task of computing the evidence much more straightforward. One can make use of numerical quadrature techniques to find

$$\mathcal{Z} \approx \sum_i w_i \mathcal{L}(X_i), \tag{2.28}$$

where w_i is the quadrature weight associated with the point X_i. For simplicity, this is often taken as $w_i = X_{i-1} - X_i$. At this point, the complex nature of evaluating the evidence integral has just been shifted onto the calculation of the prior volume; given a complicated, high dimensional space, it is not trivial to identify the volume enclosed by a given likelihood contour. Nested sampling deftly deals with this issue by making use of a probabilistic integration technique. We simply draw sequential samples according to the prior and subject to the constraint that the sample must be drawn from within the iso-likelihood contour defined by $\mathcal{L}_*$. Because $dX = \pi(\Theta)d\Theta$, sampling according to the prior means that the samples will have their prior volumes drawn uniformly from the interval $[0, X_*]$, where $X_* = X(\mathcal{L}_*)$ as defined in eq. (2.26). Taking our first sample from somewhere within the entire prior volume, we can evaluate the likelihood at that point, giving us a new iso-likelihood contour from which to sample our next point, and so on. It is expected that the bulk of the posterior occupies a small fraction $X \approx e^{-H}$ of the prior volume (Sivia & Skilling 2006), where H is the information entropy given by:

$$H = \int \mathcal{P}(\Theta) \ln \mathcal{P}(\Theta) d\Theta. \tag{2.29}$$

It is therefore sensible for the sampling to be linear in $\ln X$, rather than in X. We therefore define a sequence of 'shrinkage ratios' $t = t_1, t_2, t_3, \cdots, t_m$, where each t_j is in the range $[0, 1]$ and

$$X_i = \prod_{j=1}^{i} t_j. \tag{2.30}$$

We expect that, on average, each sample will reduce the previous volume by $\frac{1}{2}$. Therefore the expected volume enclosed by the iso-likelihood contour defined by $\mathcal{L}_i$ is $\langle X_i \rangle = \left(\frac{1}{2}\right)^i$. As we continue generating samples, the volume will shrink exponentially towards the high-likelihood region. The initial samples will contribute very little towards the evidence, since their likelihoods will be insignificant until reaching $X_i \approx e^{-H}$. Once the volume shrinks far past this fraction, we once again find that samples will stop contributing to the evidence, since their quadrature weights will diminish significantly.

We can generalise this process for the case where, instead of taking a single sample within each iso-likelihood contour, we can take N samples. By choosing the worst point (lowest $\mathcal{L}$, highest X) as the i'th sample in our sequence, we effectively increase the resolution with which we sample the prior volume. In this case, we have

$$X_0 = 1, \quad X_i = t_i X_{i-1}, \quad P(t_i) = N t_i^{N-1} \text{ in } (0,1), \tag{2.31}$$

with t_i in this case being the largest of N random numbers drawn uniformly from

$[0,1]$. The mean and standard deviation of $\ln t$ are $-\frac{1}{N}$ and $\frac{1}{N}$ respectively. The individual $\ln t$ are independent, so after i steps, we can estimate the (log-) prior volume using $\ln X_i \approx -(i \pm \sqrt{i})/N$. Note that the uncertainties in these volume estimates will propagate through to give some uncertainty to our evidence estimate. Skilling (2006) explores this further. Now that we can statistically estimate the prior volume, we can determine a termination condition. Given that we now have $X_i \approx e^{-i/N}$ and that we know most of the prior mass is contained within the region $X \approx e^{-H}$, we can run the algorithm until the number of steps i significantly exceeds NH, i.e. when most of $\mathcal{Z}$ has been found.

It is not actually necessary to generate N samples each step of the algorithm, since there are $N-1$ left over after deleting the worst one. The N samples are referred to as *live* points, whereas the deleted points are added to the set of *dead* points. At each step, we therefore only need to generate one new point, according to the prior, within the iso-likelihood contour, in order to replace the point being deleted. Finding a method to draw new points subject to these constraints is widely accepted to be the most challenging aspect of nested sampling. There are many ways to approach this problem, however. Several nested sampling variants exist, which distinguish themselves by the method used to sample from the iso-likelihood contour. Perhaps the most common approach involves copying one of the remaining points and performing a random walk using MCMC, while rejecting proposals which would lie outside of the enclosed prior volume. With any such MCMC approach, it is important to ensure that, in addition to maintaining detailed balance (and therefore ergodicity), the chain is long enough that the final point is effectively independent from the starting point, as to not bias the exploration of the enclosed prior volume. Other alternative approaches have been

developed to improve sampling efficiency, while being more robust to parameter degeneracies or distributions with multiple modes. Two popular nested sampling variants are: `MultiNest` (Feroz et al. 2009), which constructs sets of intersecting ellipsoids in an effort to approximate the iso-likelihood contours and then rejection samples from within those; and POLYCHORD (Handley et al. 2015), which generates samples using an MCMC procedure that utilises slice sampling (Neal 2000) at each step.

Whichever method is used for sampling, when the algorithm terminates, the set of dead points can then be used to calculate the evidence using eq. (2.28). Furthermore, as a byproduct we can use the dead points to produce posterior samples (allowing us to perform parameter estimation). Posterior samples can be acquired by randomly sampling from the area $\mathcal{Z}$ under the one-dimensional curve $\mathcal{L}(X)$. This area has already been divided into regions with area $\mathcal{L}_i w_i$, so the posterior probability for a dead point i is given by

$$p_i = \frac{\mathcal{L}_i w_i}{\mathcal{Z}}.$$ (2.32)

Thus, nested sampling allows us not only to calculate an estimate of the Bayesian evidence, but also to sample from the joint posterior distribution over the parameters, making it a powerful computational tool for Bayesian inference problems.

Chapter 3

SAV Identification

When the SAV hypothesis was first introduced (V2017A), ten features were identified in the light curves of seven of the 981 blazar sources investigated. These features were manually identified by eye, as detailed in section 1.2.3. While this was sufficient in identifying sources of interest with regards to this emerging phenomenon, it requires a significant amount of time and effort and is not easily reproducible. A better alternative would be to design an automated procedure which is capable of identifying SAV events within the OVRO 40m data set, with some statistically rigorous way of quantifying our belief in the results. This type of procedure would theoretically be capable of finding evidence of SAV in the OVRO data set which has gone unnoticed during manual inspection of the light curves. In this chapter I provide the details of one such approach, which makes use of Bayesian inference (more specifically nested sampling) to fit a model to OVRO 15GHz light curves and identify any SAV events present.

3.1 Modelling SAV

In order to use Bayesian inference to identify SAV activity, we require a model capable of describing features like those seen in the OVRO 15GHz light curves. Since SAV is hypothesised to occur as a result of gravitational lensing, it is sensible to try and construct a physical model based on the binary mass lensing system suggested in V2017A and described in section 1.2.3.1.

3.1.1 Physical Model

Following the method described in V2017A we can employ eq. (1.12) to simulate a series of 'lensing patterns' using the technique of inverse ray-tracing, which involves follow a large number of rays from the observer to the source plane. We 'shoot' around 10^8 rays onto a regular grid on x. Using eq. (1.12), we then compute the source-plane location of each ray. The rays are then binned on a suitable grid in the source plane. Because lensing preserves surface brightness, the magnification is the ratio of the image area to the source area on the sky. The magnification for each source bin is thus the ratio between the number of rays in the bin and the number of rays in the absence of the lens detection. These 2D bins form an image, where each bin is a pixel and the pixel's colour represents the magnification at that point. Figure 3.1 shows the lensing patterns described in V2017A as "Cupid's Bow" and "Barbed Arrow". By plotting the trajectory of a source component across the source plane, one can then trace out a 'magnification curve' showing the magnification that the emission of that component would undergo as it travels. We can think of this curve as a function of time $\mu(t)$, where μ is the ratio of the flux of the lensed component to the flux of the same component

if it was unlensed.
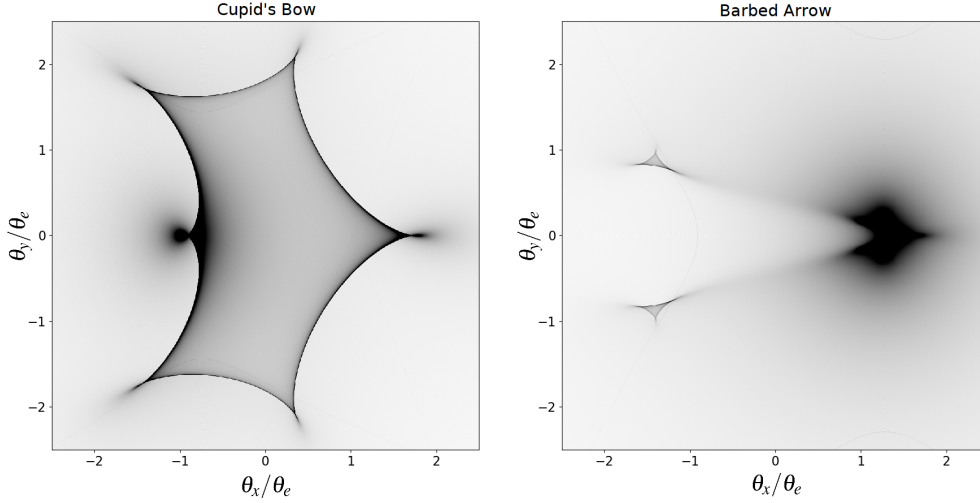
Figure 3.1: 'Cupid's Bow' (left) and 'Barbed Arrow' (right) lensing patterns.

Simulating a binary lens according to eq. (1.12) requires four parameters: the binary mass ratio q, the binary separation d (given in units of the Einstein radius), the external convergence κ, the external shear γ. There are additional variables involved in the inverse ray-shooting simulation such as the maximum/minimum angle of the rays, however these could be fixed pre-runtime. To then include the trajectory of the source component we must include another four parameters: the x and y coordinates of the starting position in the source plane s_x and s_y and the 2D velocity components v_x and v_y. Finally, the source component will have a physical size, even if it can't be resolved by the OVRO 40m telescope and thus is treated as a point source. We therefore convolve the lensing pattern with a 2D Gaussian filter defined by σ, which represents the size of the component. This will have the additional effect of smoothing the resultant magnification curve.

This nine-parameter model is sufficient to generate a magnification curve (as

a function of time) for a given set of parameter values, however this approach has a significant flaw. One could design a likelihood function which, given a set of parameter values and the corresponding magnification curve, would return a measure of how well that curve describes some data, however this means that for every likelihood evaluation in the nested sampling run, an inverse ray-tracing simulation would need to be run. This process is slow and ultimately makes using this model intractable for use with nested sampling. While the model, being physically derived, captures how SAV events are hypothesised to occur, it is not feasible to use it in this case. It is therefore desirable to develop a model which can approximate the magnification curves the physical model produces, but with far less overhead.

3.1.2 Toy Model

It was established in V2017A that we can distinguish two types of U-shaped events: a volcano type and a crater type (see section 1.2.3). In order to be able to correctly identify SAV, we must identify a model flexible enough to capture both of these event types. Obreschkow et al. (2009) introduced a parameterised model for the normalised profiles of HI- and CO-emission lines. The model gives a line profile in terms of the normalised flux density, ψ and uses five parameters: The normalised flux density at the line centre ψ_0; the normalised peak flux density ψ_{max}; the line width between the two peaks w_{peak} ; the line width at the 50-percentile level of the peak flux density w_{50}; and the line width at the 20-percentile level w_{20}. The model is shown diagrammatically in fig. 3.2.
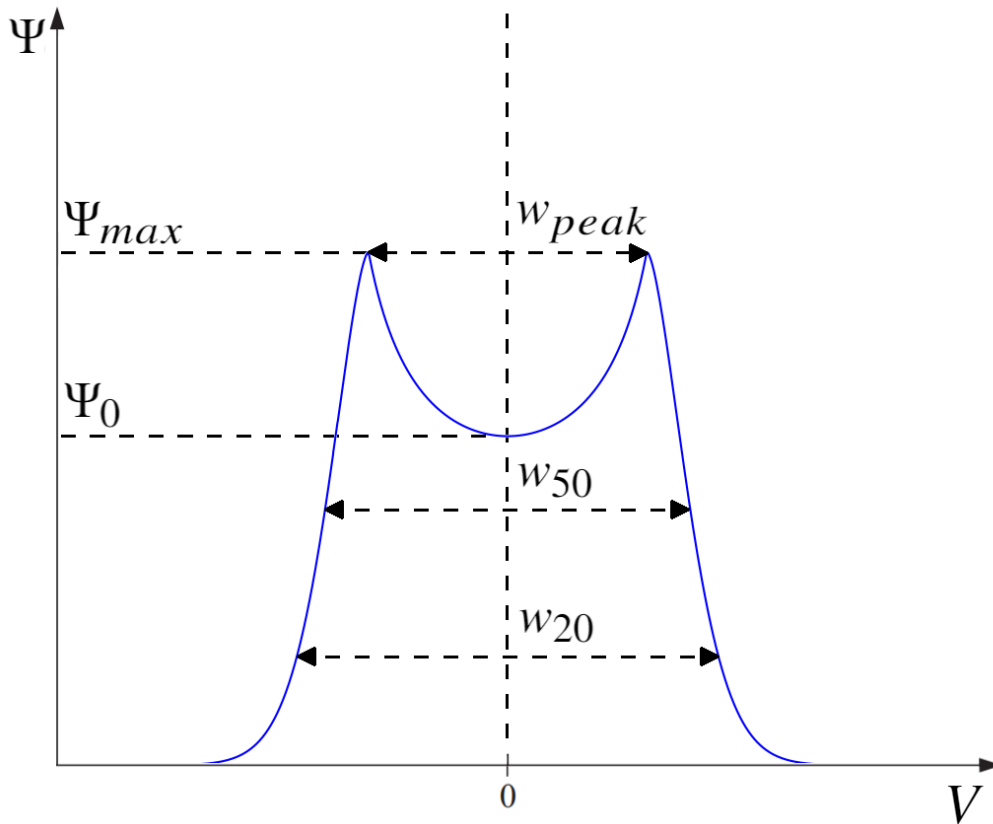
Figure 3.2: Diagram illustrating the five-parameter Obreschkow et al. (2009) model.

While HI- and CO-emission line profiles are entirely unrelated to the problem of fitting to SAV, this model does produce curves with two symmetrically opposed peaks separated by a U-shaped dip. Moreover, the model parameters provide an intuitive sense of what the resultant curve will look like. This means that this model can be adapted quite easily to approximate the magnification curves which inverse ray-tracing would allow us to produce. In fact by tuning the parameters of this model, it is possible to produce both volcano and crater shaped curves, as shown in fig. 3.3. The disadvantage of such an approach is that it is not physically motivated, so identifying SAV activity by fitting a model like this one to some

data would not provide any information about the lensing system. In this case however, once SAV activity has been identified, it may then be possible to follow up with the physically-derived lensing model to perform fits more locally to the discovered events. The posterior on the parameters of the physical model would then be informative in investigating the lensing system.
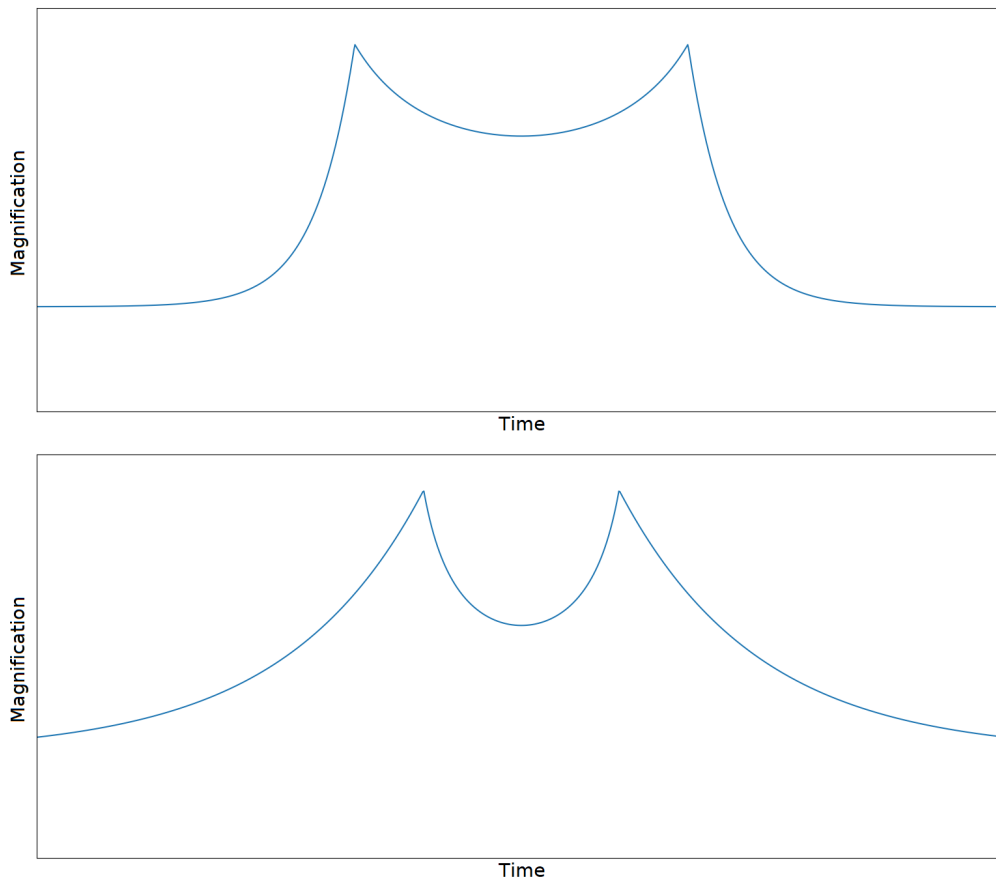
Figure 3.3: FRSD-SRFD 'crater' (top) and SRFD-FRSD 'volcano' (bottom) magnification curves described in V2017A, approximated using the Obreschkow et al. (2009) model.

In its original form, the model uses an analytic function to approximately recover the normalised velocity profiles $\psi(V)$ of emission lines, where the function is given as

$$\psi(V) = \begin{cases} \psi_{\max} \cdot e^{[k_1(|V|-k_3)^{k_2}]} & \text{if } |V| \geq \frac{w_{\text{peak}}}{2} \\[2mm] k_5(k_4 - V^2)^{-0.5} & \text{if } |V| < \frac{w_{\text{peak}}}{2} \text{ and } \psi_{\max} > \psi_0, \\[2mm] \psi_0 & \text{if } |V| < \frac{w_{\text{peak}}}{2} \text{ and } \psi_{\max} = \psi_0 \end{cases} \tag{3.1}$$

where $k_1, \cdots, k_5$, are free parameters which can be inferred from the five model parameters as follows:

$$k_1 = -0.693 \cdot 2.322^{\left[\frac{\ln(w_{50}-w_{\text{peak}})-\ln 2}{\ln(w_{50}-w_{\text{peak}})-\ln(w_{20}-w_{\text{peak}})}\right]},$$

$$k_2 = \frac{0.842}{\ln(w_{20} - w_{\text{peak}}) - \ln(w_{50} - w_{\text{peak}})},$$

$$k_3 = \frac{w_{\text{peak}}}{2},$$

$$k_4 = \begin{cases} \frac{1}{4}\frac{w_{\text{peak}}^2 \psi_{\max}^2}{\psi_{\max}^2 - \psi_0^2} & \text{if } \psi_{\max} > \psi_0, \\[2mm] 0 & \text{if } \psi_{\max} = \psi_0 \end{cases},$$

$$k_5 = \psi_0\sqrt{k_4}.$$

(3.2)

While this model can produce both magnification curve shapes, it needs some modifications to actually produce magnification curves. Firstly, to avoid any confusion, we will rename the $\psi_{\max}$ and ψ_0 parameters to $\mu_{\max}$ and μ_0 respectively, and we replace V with t, since we are now concerned with magnification as a function of time $\mu(t)$, rather than normalised flux density as a function of velocity $\psi(V)$. In addition, the baseline magnification of the curve (away from the U-shaped feature) using the original model would be zero, when the magnification should be unity. This can be fixed by simply adding 1 to $\mu(t)$ at every point

along the curve. The modified function becomes:

$$
\mu(t) = \begin{cases}
\mu_{\max} \cdot e^{[k_1(|t|-k_3)^{k_2}]} + 1 & \text{if } |t| \geq \frac{w_{\text{peak}}}{2} \\
k_5(k_4 - t^2)^{-0.5} + 1 & \text{if } |t| < \frac{w_{\text{peak}}}{2} \text{ and } \mu_{\max} > \mu_0, \\
\mu_0 + 1 & \text{if } |t| < \frac{w_{\text{peak}}}{2} \text{ and } \mu_{\max} = \mu_0
\end{cases} \tag{3.3}
$$

In addition to magnification, the source can also experience demagnification, where the curve dips below unity, which the model does not yet allow for. We can further modify eq. (3.3) to give

$$
\mu(t) = \begin{cases}
\mu_{\max} \cdot e^{[k_1(|t|-k_3)^{k_2}]} + 1 & \text{if } |t| \geq \frac{w_{\text{peak}}}{2} \\
k_5(k_4 - t^2)^{-0.5} \cdot \frac{\mu_{\max}+1}{\mu_{\max}} & \text{if } |t| < \frac{w_{\text{peak}}}{2} \text{ and } \mu_{\max} > \mu_0, \\
\mu_0 + 1 & \text{if } |t| < \frac{w_{\text{peak}}}{2} \text{ and } \mu_{\max} = \mu_0
\end{cases} \tag{3.4}
$$

which preserves the overall shape of the curve, while allowing the U-shaped 'trough' to drop below unity. At this point, a value of e.g. $\mu_{\max} = 2$ would mean the curve peaks at $\mu = 3$, since we added 1 to $\mu(t)$ at all points. It would be more convenient for $\mu_{\max}$ to represent the actual magnification at the peaks, so we then replace $\mu_{\max}$ with $\mu_{\max} - 1$ anywhere it appears in the function definition (including the definitions of $k_1, \cdots, k_5$). Next, we must consider that the magnification curves given by the model are centred at $t = 0$, when in actuality the events found in the OVRO light curves could be centred at any arbitrary point in time (within the span of time that observations have been ongoing). We therefore introduce a sixth parameter t_0 to indicate the time at which the event is centred. Subtracting a constant factor of t_0 from t in our function will translate the curve to be centred at $t = t_0$. One final aspect of the model to consider is that defining a prior distribution over the model parameters, as is required for

nested sampling, could be potentially problematic. More specifically, if one were to define a prior over μ_{max} (for example) for some range of values and then define a prior over μ_0 with and overlapping range, this introduces the possibility that $\mu_{\text{max}} < \mu_0$, a situation for which $\mu(t)$ is undefined. The same problem arises for the 'width' parameters w_{peak}, w_{50} and w_{20}, since we must ensure that the U-shaped feature widens as the magnification tapers off. One tactic for dealing with this problem is to define a likelihood function which simply returns zero when this condition is met. While this approach is functional, there is nothing to stop the nested sampling algorithm from proposing points which would have zero likelihood in this way and would therefore lie outside of the iso-likelihood contour. This could decrease the efficiency of the algorithm, as time is wasted proposing these points. An alternative solution is to reparameterise the model in such a way that the output is preserved, but where we can ensure that overlapping priors will not be a problem. We can replace μ_0 with a new parameter $q = \mu_0/\mu_{\text{max}}$, where $0 < q < 1$. Similarly, we can replace w_{50} with $\Delta w_1 = w_{50} - w_{\text{peak}}$ and replace w_{20} with $\Delta w_2 = w_{20} - w_{50} = w_{20} - w_{\text{peak}} - \Delta w_1$, with the added conditions that $0 < \Delta w_1$ and $0 < \Delta w_2$. This reparameterisation enforces the assertions that $\mu_{\text{max}} \geq \mu_0$ and that $w_{\text{peak}} < w_{50} < w_{20}$, while maintaining an intuitive sense of how the altering the values of the new parameters would affect the resultant magnification curve. The downside of this approach is that we are introducing linear parameter degeneracies, which nested sampling can traditionally struggle with, although this is not a problem for POLYCHORD thanks to its in-built contour whitening. With those modifications, our final magnification model comprises six parameters $\Theta_{\text{mag}} = (\mu_{\text{max}}, q, w_{\text{peak}}, \Delta w_1, \Delta w_2, t_0)$, with magnification curves produced using the function

$$\mu(t) = \begin{cases} (\mu_{\max} - 1) \cdot e^{[k_1(|t-t_0|-k_3)^{k_2}]} + 1 & \text{if } |t - t_0| \geq \frac{w_{\text{peak}}}{2} \\ k_5(k_4 - (t-t_0)^2)^{-0.5} \cdot \frac{\mu_{\max}}{\mu_{\max}-1} & \text{if } |t - t_0| < \frac{w_{\text{peak}}}{2} \text{ and } q < 1, \\ q \cdot \mu_{\max} & \text{if } |t - t_0| < \frac{w_{\text{peak}}}{2} \text{ and } q = 1 \end{cases} \qquad (3.5)$$

where we now have

$$k_1 = -0.693 \cdot 2.322^{\left[\frac{\ln \Delta w_1 - \ln 2}{\ln \Delta w_1 - \ln(\Delta w_1 + \Delta w_2)}\right]},$$

$$k_2 = \frac{0.842}{\ln(\Delta w_1 + \Delta w_2) - \ln \Delta w_1},$$

$$k_3 = \frac{w_{\text{peak}}}{2},$$

$$k_4 = \begin{cases} \frac{1}{4} \frac{w_{\text{peak}}^2 (\mu_{\max}-1)^2}{(\mu_{\max}-1)^2 - (q(\mu_{\max}-1))^2} & \text{if } q < 1, \\ 0 & \text{if } q = 1 \end{cases}$$

$$k_5 = q(\mu_{\max} - 1)\sqrt{k_4}.$$

The final magnification model allows for the approximation of the magnification curves we can attain using inverse ray-tracing, while also being nicely suited for use with nested sampling.

3.1.3 Background Variability

So far we have considered models which generate magnification curves representative of the lensing undergone by a source due to SAV, however these events obviously do not occur in isolation and SAV is not the only form of variability

observed in blazars. To complete our model we must also attempt to account for the typical variability seen in the light curves of unlensed sources.

We firstly consider the long-term variability of sources observed by the OVRO monitoring program. Over the course of several years, the source flux densities can often (but not always) be seen to increase or decrease according to some trend. This trend can sometimes be linear, but is often more complex and would require a higher order polynomial to describe. Generally, however, a cubic function of the form $S(t) = at^3 + bt^2 + ct + d$ is sufficient to capture the long-term variability of OVRO sources. A naive approach would be to incorporate the cubic coefficients a, b, c and d as model parameters, although doing so would make defining a prior distribution quite difficult, as it is not immediately obvious what combination of these parameters would produce a cubic which describes a given light curve's long-term variability. Instead, it is better to consider the (time and flux density) coordinates of the stationary points of the cubic, which are easier to relate to a given light curve. If we only consider cubic curves with two unique stationary points, we have the coordinates of the first turning point (t_1, S_1) and the second turning point (t_2, S_2). Given these coordinates, it is possible to identify the unique cubic curve which passes through (and has zero derivative at) both points. Ultimately the stationary point coordinates are preferable parameter candidates for modelling long-term variability, compared to the cubic function coefficients.

When it comes to shorter-term variability, a common approach is to model blazar light curves as a series of exponential flares (Hovatta et al. 2009, Abdo et al. 2010*b*). We can define a flare using four parameters: the peak flux S_{peak}, the peak time t_{peak}, the rise rate r and the decay rate d. The amplitude of a given flare at time t is then given by

$$S(t) = \begin{cases} S_{\text{peak}} e^{(t - t_{\text{peak}})/r}, & t \le t_{\text{peak}} \\ S_{\text{peak}} e^{(t - t_{\text{peak}})/d}, & t > t_{\text{peak}}. \end{cases} \tag{3.7}$$

While this model is perfectly suitable for describing an exponential flare, there is a problem. For most of (if not all of) the OVRO 15GHz light curves, fitting a single flare would not be sufficient for capturing the observed variability. Several flares will be needed, however the exact number will be different from source to source and that number will not be known *a priori*. This is problematic because traditional nested sampling algorithms expect a fixed dimensionality, whereas this problem would be ideally solved by an approach which allows the number of flares to freely vary (therefore varying the dimensionality of the problem as the algorithm progresses). Chapter 4 explores a novel framework which would facilitate such a 'transdimensional' approach, however for the current problem, we will instead make use of the 'adaptive' method (Higson et al. 2019, Hee et al. 2016, Chua et al. 2018). We term each set of four parameters describing an individual flare as a 'particle'. We then choose a fixed dimensionality which is sufficient to include N particles, in addition to any other model parameters which don't contribute to the particles. In this case, inspection of the OVRO light curves indicates that any of the light curves can be decomposed into no more than 15 exponential flares, so we would then include 60 parameters in our model to describe those 15 flares. We then also include in our model the remaining parameters which describe the cubic background and the magnification curve. Finally, we add in one final parameter N, which represents the interpreted dimensionality of the current parameters. Each live point in a nested sampling run will have some (integer) value for N and we can treat that live point as only having that many particles,

only considering the first *N* particles and ignoring the rest. In this way, *N* is able to vary freely and we can determine the most likely value for *N* (i.e. the most likely number of particles) *a posteriori*.

3.1.4   Noise Modelling

With each OVRO observation, an estimate of the telescope thermal noise is calculated from the variance of the time stream data that go into producing the flux estimate, however this is not the only contribution to the total flux error. Generally there will also be an additional pointing error, which will be approximately constant during a single observation and thus will not affect the thermal noise estimate. It will, however, result in a random error in the flux estimate such that the expectation flux measured is *AS*, where *S* is the source flux and *A* is determined by the pointing calibration measurement carried out just before the source observation. We assume that the value of *A* is independent between subsequent observations, and its distribution will be complicated, however it will result in an additional error between successive data points equal to αS. If we take the (seemingly unlikely) assumption that α is Gaussian distributed, we can sum this additional error with the thermal noise in quadrature to find the total error on each data point $\sqrt{(\alpha S_i)^2 + \sigma_i^2}$. If we include α as a free parameter in our SAV model, we can perform model selection against the same model without α. The Bayesian evidence will indicate which of these two models better describes the data. The additional parameter must provide a significantly better description of the data, otherwise the additional term will naturally incur an Occam penalty and the original model will be preferred. Testing both models, it seems that including the

additional error term does produce a model which is preferred, as indicated by the Bayesian evidence. Typically, we find that the maximum *a posteriori* (MAP) value for α to be around $\alpha = 0.05$. Interestingly, the OVRO data reduction process already includes a term to account for the pointing error, however this result may indicate that it currently gives an underestimate.

With all of this in mind, we have finally constructed a model sufficient for identifying SAV in blazar light curves, defined in terms of the parameters $\Theta = (\Theta_{\text{mag}}, \Theta_{\text{cubic}}, \Theta_{\text{flares}}, \alpha)$, where Θ_{mag} is defined previously, $\Theta_{\text{cubic}} = (t_1, S_1, t_2, S_2)$ and $\Theta_{\text{flares}} = (N, S_{\text{peak},1}, t_{\text{peak},1}, r_1, d_1, \cdots, S_{\text{peak},N}, t_{\text{peak},N}, r_N, d_N)$. The cubic acts as a baseline, which we can then add N exponential flares to by summing their amplitudes with the cubic. We then find the product of that curve with the magnification curve and the final resultant curve can then be compared to the data.

3.2 The Fitting Procedure

With a suitable model to describe the light curve of a source undergoing SAV, we return to the problem of designing a procedure to fit this model to data, with the goal of ultimately identifying SAV within the OVRO 40m data set. Given a blazar light curve which was known *a priori* to feature an SAV event, one could run nested sampling using the model in an attempt to recover the event. Assuming we appropriately defined both a likelihood and prior distribution, nested sampling would return a posterior distribution over the parameters. This allows us to make inferences about the parameter values, such as finding the MAP values for the parameters in Θ_{mag}, which would indicate the most probable position of the SAV event, among other things.

Taking things a step further, it is important to consider that when using this approach with real OVRO data, it would likely be the case that we don't actually know *a priori* whether there might be one event, two events, or no events at all. It is conceivable that one might extend the adaptive model described above to also allow for a variable number of SAV events, however this would increase the complexity of this (already fairly complex) model significantly. As an alternative, we can simply perform multiple nested sampling runs, each time using an altered version of the model which expects no SAV events (so Θ_{mag} has no parameters), one SAV event (Θ_{mag} has five parameters), two SAV events (Θ_{mag} has ten parameters), etc. We can then use the Bayesian evidence computed during each run to perform model selection and determine which model best describes the observed data.

3.2.1   Likelihood and Weighting

We assume the noise on the data to be approximately Gaussian and so we define a likelihood function of the form $\mathcal{L}(\Theta) \propto \exp\left(-\frac{1}{2}\chi^2\right)$, where χ^2 is the standard statistic measuring the misfit between the observed data $O = \{O_1, O_2, ..., O_N\}$ and the data predicted by the model $E = \{E_1, E_2, ..., E_N\}$,

$$\chi^2 = \sum_i^N w_i (O_i - E_i)^2. \tag{3.8}$$

Typically, $w_i = \omega_i = 1/\sigma_i^2$, where σ_i is the observational uncertainty, so observations with high uncertainties have less weight in determining goodness of fit. In this case, however, we are concerned not just with uncertainties, but also the observing cadence. Typically each source in the OVRO 40m catalogue is observed

once every three days, however this is not always possible due to telescope main-
tenance, weather conditions (particularly wind), incomplete/bad observations, etc.
This means, as is common with astronomical data, the light curves are unevenly
sampled. It is very feasible, therefore, that uninteresting areas of the light curve
(i.e. a region away from the SAV events) could be quite densely sampled, whereas
the events themselves may be (at least partially) undersampled, and we should
consider weighting the data points accordingly. Robust weighting (Briggs 1995)
introduces a controllable robustness parameter R, allowing a scalable trade-off
between 'natural' weighting, where data points are weighted by their inverse vari-
ance, and 'uniform' weighting, which gives higher weighting to data points in
sparse regions. While this weighting scheme was originally developed to weight
visibilities in the u-v plane for radio interferometric imaging, in principle there is
nothing to stop it being applied in this case. We first divide our data temporally
into a number of fixed-width windows, and for each window k, we calculate W_k
as the sum of the inverse variances of all data points within that window. The
weights can be calculated using

$$w_i = \frac{\omega_i}{1 + W_k f^2},$$

(3.9)

where

$$f^2 = \frac{(5 \cdot 10^{-R})^2}{\frac{\sum_k W_k^2}{\sum_i \omega_i}}.$$

(3.10)

.

The robustness parameter R can take any value between -2 (close to uniform
weighting) and 2 (close to natural weighting), and can be fixed before the fitting

procedure begins, or even possibly included as a free parameter in the model. In effect, this type of weighting scheme should stop the model from necessarily fitting too closely to dense clusters of data points and allow it more freedom to fit to undersampled regions.

While initial tests of implementing robust weighting as part of this model looked promising, time constraints prevented any proper analysis of the impact it might have on the fitting procedure. Instead, we will simply assume a natural weighting scheme going forward, although the robust scheme could certainly be revisited in the future.

3.2.2 The Prior Distribution

We will naively assume fairly broad, uniform priors over all of the model parameters, since there is currently nothing to suggest that a more specific distribution would be preferable for any of the model parameters. The possible ranges of parameter values are shown in table 3.1. Here $t_{\min}$ and $t_{\max}$ are the time coordinates of the earliest and latest data points respectively, and similarly $S_{\min}$ and $S_{\max}$ are the minimum and maximum flux density coordinates respectively. By making use of these values, we can ensure that the range of possible parameter values is scaled to the data we are fitting to. Some of the parameter ranges have additional terms which allow the parameter values to extend beyond the bounds defined by $t_{\min}, t_{\max}, S_{\min}$ and $S_{\max}$, providing the model some more flexibility. For example, t_1 and t_2 can fall anywhere in the range between $t_{\min}$ and $t_{\max}$, but also within a margin of 5% of that range on either side. Note that this is not the case for t_0, the time coordinate(s) of the centre of the magnification curve(s), be-

cause the individual peaks of the magnification curves are very similar in shape to the exponential flares. This means that if, for example, an SAV event occurred, but observations on the source only began in time to record the second peak, with more than half of the magnification curve lying outside of the span of time defined by the observations, we would not be able to conclude that the variation observed is the end of an SAV event, rather than just an exponential flare. A similar argument can then be made for the case where we see a peak at the very end of the light curve, indicating an event may be ongoing, but more observations would be needed to extend the light curve and conclude whether it is a true event (or at least whether the fitting procedure concludes that this is the case). The minimum value of w_{peak} was chosen as it would require a minimum of five data points to indicate the type of variation we see from SAV events: one point at each of the peaks, one point in the U-shaped dip, and two more points away from the event. This means that, with an average cadence of 3 days in the OVRO data, we would realistically only expect to observe an SAV event which spans around 10 days from peak to peak. Conversely, the maximum value was chosen because if there were some event which had its peaks separated by the entire span of the currently available data, we would not be able to see the peaks in that data, and would have to wait for more observations to be carried out. All other parameter limits were selected to simply give a wide enough range so that if any one of the parameters took a value outside of its range, the model would give a very bad fit to any of the light curves in the OVRO data set.

Table 3.1: Prior ranges for all parameters in the SAV model. Note that while most of the parameters can take any values in their respective prior ranges, N is an integer. Additionally, while the model includes N_{max} flares, the parameters of two different flares will be drawn from the same distribution. Similarly, the model may be extended to include any number of SAV events (or possibly none), but the parameters of two different SAV magnification curves will be drawn from the same distribution.

	Parameter	Prior Range
Θ_{cubic}	t_1	$[t_{\text{min}} - 0.05(t_{\text{max}} - t_{\text{min}}), \; t_{\text{max}} + 0.05(t_{\text{max}} - t_{\text{min}})]$
	t_2	$[t_{\text{min}} - 0.05(t_{\text{max}} - t_{\text{min}}), \; t_{\text{max}} + 0.05(t_{\text{max}} - t_{\text{min}})]$
	S_1	$[S_{\text{min}} - 0.05(S_{\text{max}} - S_{\text{min}}), \; S_{\text{max}} + 0.05(S_{\text{max}} - S_{\text{min}})]$
	S_2	$[S_{\text{min}} - 0.05(S_{\text{max}} - S_{\text{min}}), \; S_{\text{max}} + 0.05(S_{\text{max}} - S_{\text{min}})]$
Θ_{mag}	μ_{max}	$[1, 4]$
	q	$(0, 1]$
	w_{peak}	$[10, \; t_{\text{max}} - t_{\text{min}}]$
	Δw_1	$[1, 500]$
	Δw_2	$[1, 500]$
	t_0	$[t_{\text{min}}, \; t_{\text{max}}]$
Θ_{flares}	N	$[0, 15]$
	S_{peak}	$(0, \; S_{\text{max}}]$
	t_{peak}	$[t_{\text{min}} - 0.05(t_{\text{max}} - t_{\text{min}}), \; t_{\text{max}} + 0.05(t_{\text{max}} - t_{\text{min}})]$
	r	$(0, 1000]$
	d	$(0, 1000]$
	α	$(0, 0.1]$

3.3   Light Curve Simulation

With a procedure in place for identifying the presence of SAV in a given blazar light curve, it is important to test this procedure to determine how well it actually performs (and then tune the procedure accordingly). Unfortunately, there are only a handful of confirmed SAV events which occupy a very small proportion of the set of OVRO 15GHz light curves. It would be desirable, therefore, to find a way to gather/produce some more data in which we could place artificial SAV events to test the fitting procedure's ability to recover those events. Thankfully, we have already developed a model capable of describing the variation seen in the light curve of an unlensed source, and so it should be possible to leverage that to produce a new set of artificial 'OVRO-like' light curves, satisfying our need for more data. It should also be noted that the ability to produce artificial light curves based on the real OVRO data has applications beyond the scope of SAV identification, such as identifying periodicity in flare events or finding correlations between source variability and other indicators e.g. neutrino events. In either of these cases, it would be useful to simulate OVRO-like light curves with no inbuilt periodicity/correlations and then assess the significance of any putitive detection by running the same analysis on the mock data.

We begin by running the fitting procedure with a model which expects no SAV events. In essence, our model now only consists of the parameters in Θ_{cubic} and Θ_{flares}, which capture the non-SAV variability of the source, as well as α to more correctly model the noise. For brevity this model (i.e. with Θ_{mag} removed) will hereafter be referred to as the background model or $\mathcal{M}_0$. More generally, any model which expects N SAV events will be referred to as the N-

SAV model or $\mathcal{M}_N$. We run POLYCHORD to fit the background model to the data from each of the sources in the OVRO catalogue, using a likelihood function as described in section 3.2.1 and the priors described in section 3.2.2. Taking the posterior samples returned by POLYCHORD, along with their corresponding posterior weights, we can generate a curve for each sample and take a weighted average of those curves, resulting in an average curve which describes the data seen in the light curve. In a similar fashion, we can also then find the (weighted) error associated with that curve, allowing us to plot the curve with a confidence interval. An example curve is shown in fig. 3.4. By doing so we can inspect how well the model fits to the data.
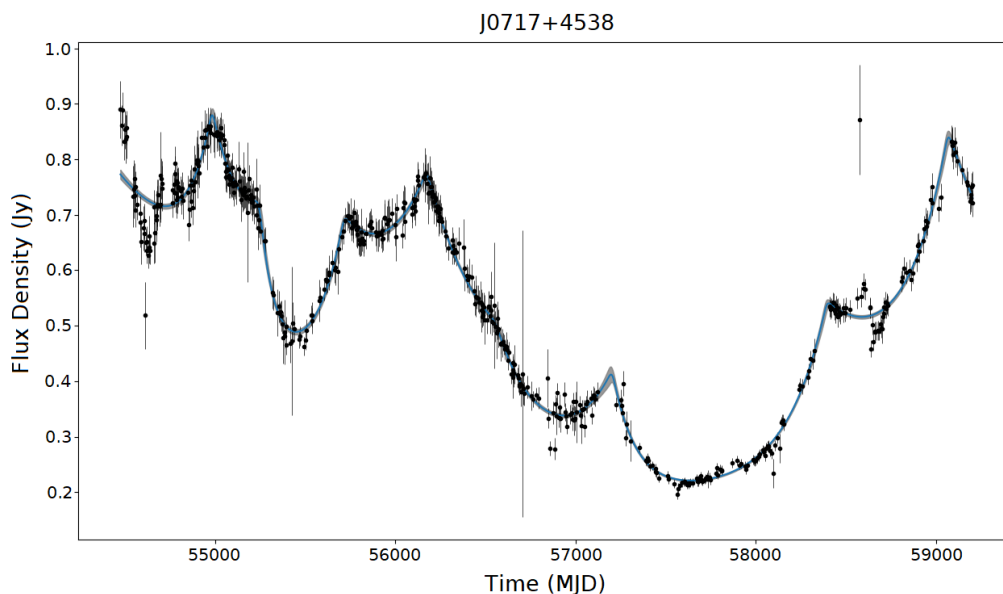
Figure 3.4: Posterior weighted average curve with a 1σ confidence interval produced by fitting the SAV model to the OVRO 15GHz light curve of J0717+4538. Note the confidence interval widens in particular around regions where the data is sparse/missing.

It seems that generally the model provides a good fit to the OVRO light curves, with the associated 1σ confidence intervals mostly being very small. By running

this fitting procedure to a given light curve from the OVRO data set, we can now begin sampling the posterior distribution over the model parameters in order to generate new data which mimics the variations seen in that light curve. We select a sample from the posterior distribution, where the probability of selecting each sample is proportional to the associated posterior weight. From this sample, we then retrieve the values of the four parameters which describe the cubic and use them to produce a cubic curve. We then perform another weighted sample, and find the value of N from that sample, which informs how many exponential flares will be added. Sampling flares from the posterior becomes a little more complicated, however. Each sample consists of multiple flares, where the number of flares is not necessarily consistent between samples. We can get around this by noticing that each of these flares can be described by a single point within the same four-dimensional parameter space (since there are four parameters describing a flare). This 'flattened' space is a concept which is further explored in chapter 4. Using the nested sampling visualisation tool `anesthetic` (Handley 2019), we can re-shape the list of posterior samples, splitting each of the original samples into N new samples, each one taking the values of a different exponential flare. We can then re-weight this flattened list of samples appropriately. We draw N samples from the flattened list (again in a weighted fashion), and for each of the samples, we use the flare parameter values to add a flare to the existing cubic.

The procedure thus far will produce a continuous curve which mimics the variability seen in the OVRO data. In order to produce an artificial light curve from this, we must sample points along the curve and then add some noise to offset the flux values from the generated curve. The OVRO data, as previously stated, is unevenly sampled. We can mimic the cadence of observations from

the OVRO light curve which we originally fit to by finding the amount of time from each observation to the next in the original data, then repeatedly drawing randomly from this set of time differences and sampling that far along the curve. Similarly, we can also mimic the uncertainties from the original light curve by simply taking the uncertainty on a random observation from the original data and use that uncertainty to introduce a gaussian offset in the flux density. We then repeat this for each of the newly sampled points along the curve.

The final outcome of this procedure is a brand new artificial light curve. A handful of light curves produced via this method using J0717+4538 (fig. 3.4) are shown in fig. 3.5.
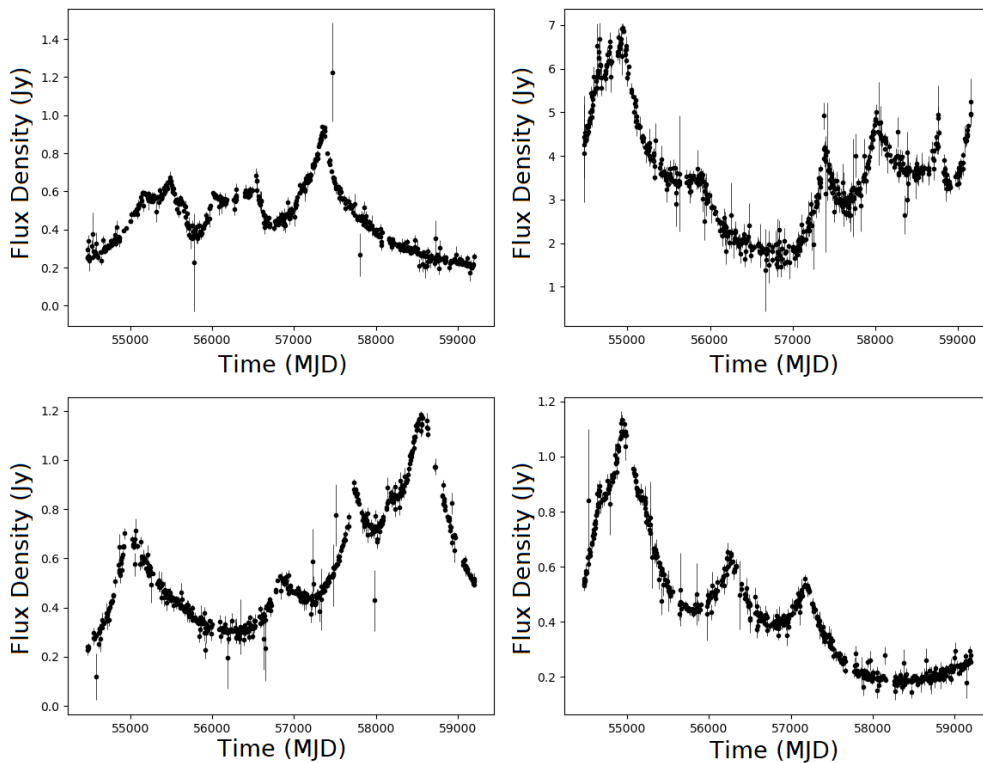
Figure 3.5: Four artificial light curves based on the OVRO 15GHz light curve of the source J0717+4538.

It should be noted that two other members of the OVRO collaboration have implemented their own distinct methods of simulating artificial light curves using real OVRO light curves as seeds. These three methods are being analysed to determine the quality of the simulated light curves they produce (i.e. how similar they are to the real OVRO light curves). At the time of writing, however, this analysis is still ongoing and there are no results available to discuss.

3.4   Results

In order to test the ability of the fitting procedure described in section 3.2, we need to run it with some data. The ultimate goal would be to run the procedure over the entire set of OVRO 15GHz light curves, however we would first need to confirm that the procedure works correctly and determine how reliably it can identify SAV events. Unfortunately time pressures have not allowed such analysis, so instead we focus on demonstrating the results of running the current fitting procedure on both simulated and real OVRO 15GHz light curves. The testing framework presented here is robust and demonstrates how we can analyse the procedure's performance and find ways to improve it in future.

3.4.1   Simulated data

In order to determine how effective the fitting procedure is at identifying true SAV events, we would require a reasonably large number of light curves which are known to display SAV. There are only seven such SAV candidates within the OVRO data set (V2017A), however given the procedure for simulating 'OVRO-like' light curves described in section 3.3, it is possible to quickly generate a large

set of artificial light curves to test the procedure with instead. Given a simulated light curve, one could place in any number (including zero) of artificial events using the magnification curve model, either using the toy magnification model, or the physically-motivated lensing model). One can then run the fitting procedure over the light curve, which hopefully would successfully recover all of the events without falsely identifying some other variation as SAV activity. Because the number of events in the curve is *a priori* known, this process could then be repeated, allowing us to produce a confusion matrix detailing the overall true/false positive/negative rates, which could be interpreted to determine the performance of the fitting procedure. This would also form an excellent starting point for tuning/modifying the procedure to improve its performance. Unfortunately, while generating the artificial light curves is fairly quick, running the fitting procedure is not, meaning this kind of analysis has not been possible in the available time frame. Instead, a demonstration of running the fitting procedure with some simulated data is included here to give some insight into the procedure's potential performance.
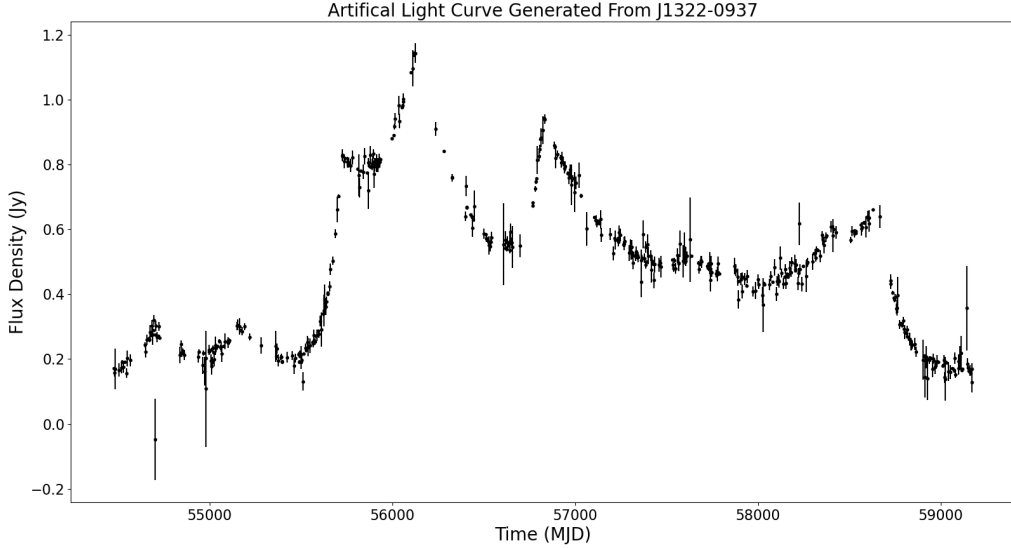
Figure 3.6: Artificial light curve generated using J1322-0937 as a seed (before any artificial SAV events are placed within the data).

We select J1322-0937 randomly from the catalogue of OVRO sources and the corresponding OVRO 15GHz light curve was used as the seed to generate an artificial light curve, shown in fig. 3.6. The 'toy' magnification curve model was then used to place an event into the light curve where $\mu_{\mathrm{max}} = 0.625$, $q = 0.438$, $t_0 = 57640$, $w_{\mathrm{peak}} = 300$, $\Delta w_1 = 200$ and $\Delta w_2 = 300$.

We run the fitting procedure on the artificial light curve, testing three models: the background model, a 1-SAV model and a 2-SAV model, which expect zero, one and two SAV events respectively. Taking the evidence estimates produced by POLYCHORD, we can find the posterior probability for each model, $P(\mathcal{M}_N|\mathcal{D})$, using eq. (2.13). The results of doing so are shown in fig. 3.7. It is clear to see that the 2-SAV model is strongly favoured over the other two models, which might lead us to conclude that there are (at least) two SAV events present in the data, however we know that only one event was manually placed in the light curve, so

this result is indicative of a false positive (type 2 error). In practice, upon seeing $P(\mathcal{M}_N|\mathcal{D})$ peak at $N = 2$, we should keep running the model with higher and higher values for N (i.e. expecting more and more SAV events), until we find an N-SAV model which is not preferred over the one before it, indicating that including more parameters no longer improves the fit enough to offset the Occam penalty. In this case, we stop at $N = 2$, as it is found to be enough to indicate the current capabilities of the fitting procedure.
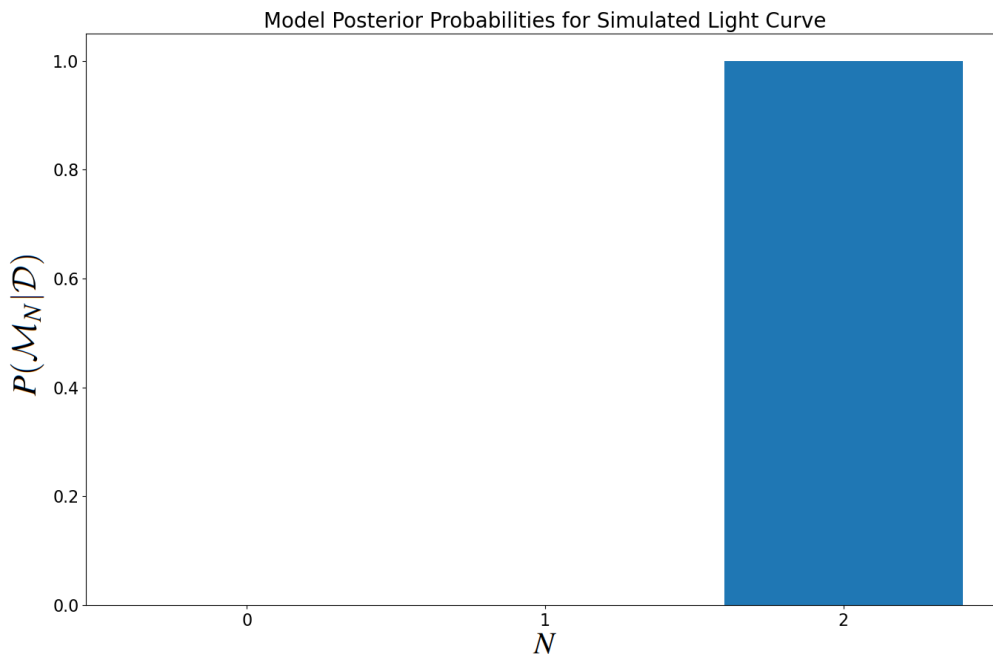
Figure 3.7: Bar plot indicating the posterior probability of each model given the data from the artificial light curve, where model $\mathcal{M}_N$ expects N SAV events. Note that all bars are displayed, but those for $N = 0$ and $N = 1$ are invisibly small.

We can further investigate what went wrong here by following the methods used to create fig. 3.4, and, for each model, producing a weighted average curve from the posterior samples. These curves are shown in figs. 3.8 to 3.10. Note that for the following plots, the yellow highlighted marker indicates the region of the SAV

event that was placed into the simulated curve.
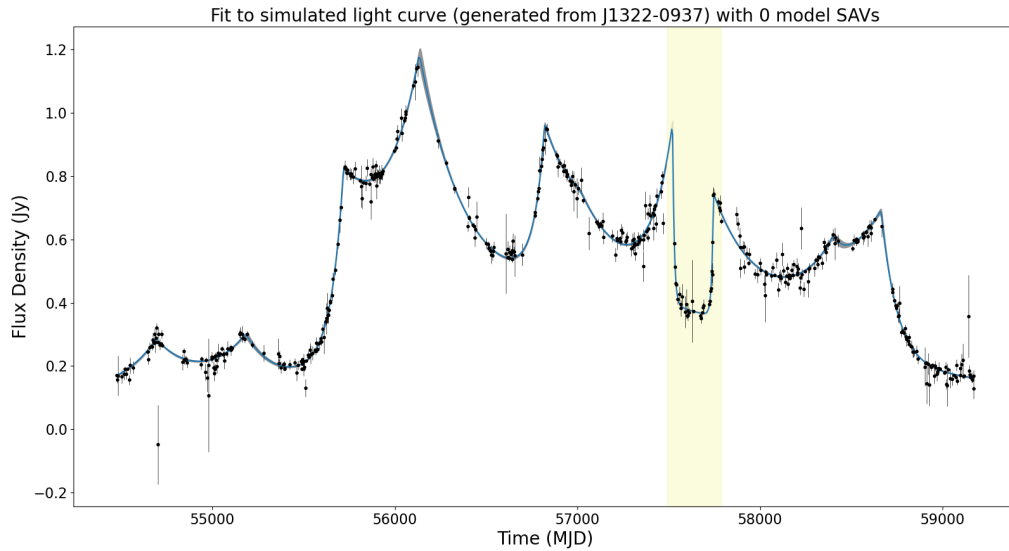
Figure 3.8: Weighted average curve from fitting the background model to the given artificial light curve.
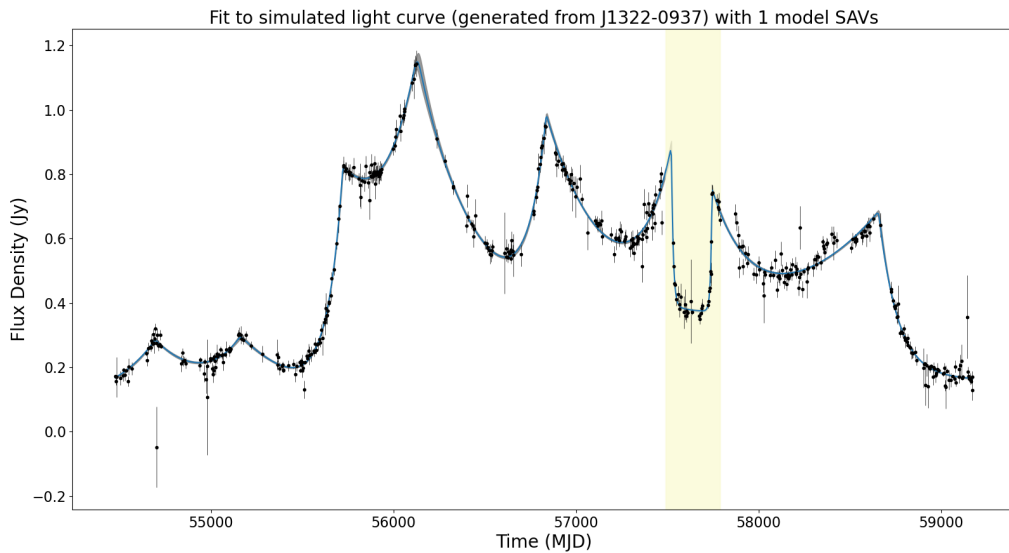
Figure 3.9: Weighted average curve from fitting the 1-SAV model to the given artificial light curve.
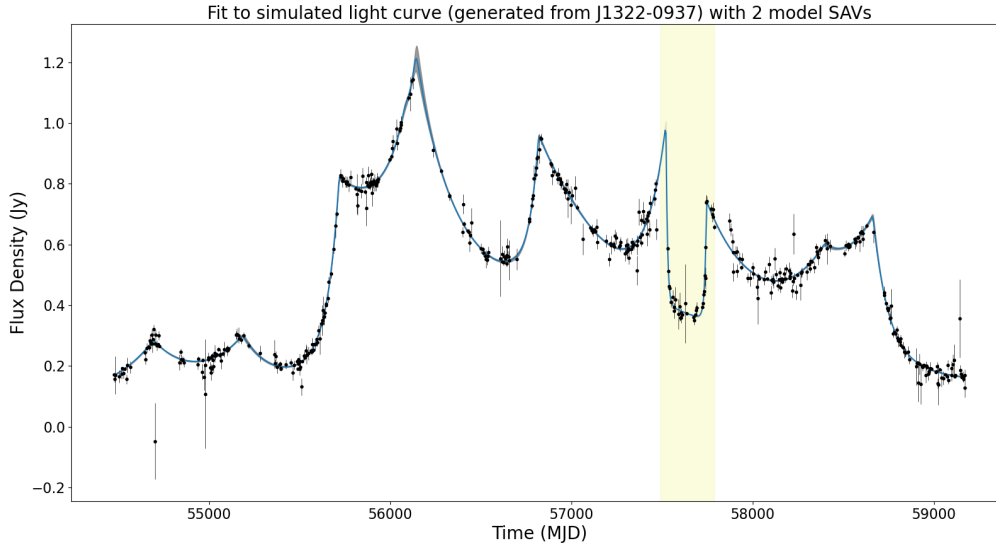
Figure 3.10: Weighted average curve from fitting the 2-SAV model to the given artificial light curve.

Interestingly, the fits each look very good when inspecting by eye, although for the 2-SAV model to be so heavily favoured, the evidence estimate for that run must have been significantly higher than the others. This means that this model must provide a significantly better description of the data (at least as far as our likelihood function is concerned) to offset the Occam penalty of the additional model parameters.

While the fits look good and the evidences suggest the 2-SAV model is preferred, we can investigate further to find exactly which features in the curve the model is reporting as being the result of SAV activity. By marginalising over all model parameters excluding Θ_{mag}, we can produce a similar plot showing the weighted average magnification curve. Obviously the background model does not include a magnification curve, but the results for the 1-SAV and 2-SAV model are shown in figs. 3.11 and 3.13 respectively. We can also inspect the joint posterior

over the parameters in Θ_{mag} by marginalising over the other model parameters
and using `anesthetic` to produce triangle plots, which are shown in figs. 3.12
and 3.14. On the diagonal of these plots, we see the 1D posterior over each para-
meter, with the other cells showing the joint 2D posterior over each parameter
pair. Note that the 2-SAV model has two sets of magnification curve parameters,
so for readability we create separate plots for each.
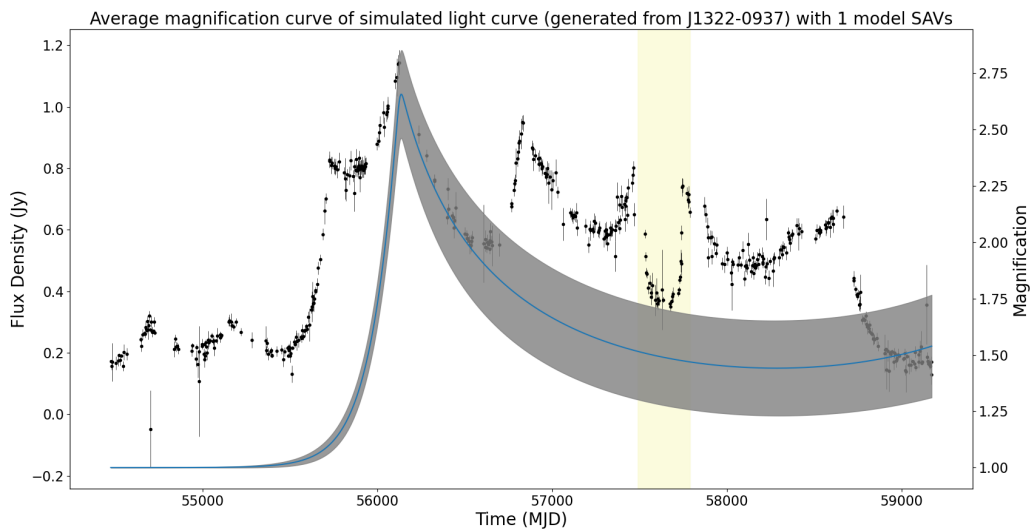
Figure 3.11: Weighted average magnification curve from fitting the 1-SAV model to the
given artificial light curve.
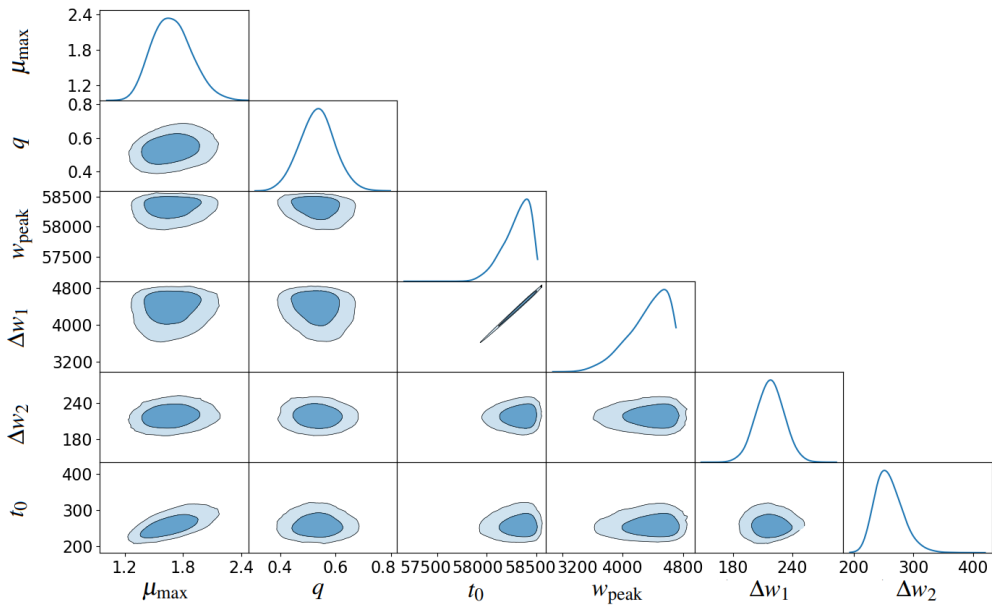
Figure 3.12: Triangle plot illustrating the posterior over the parameters in Θ_{mag} from fitting the 1-SAV model to the artificial light curve.
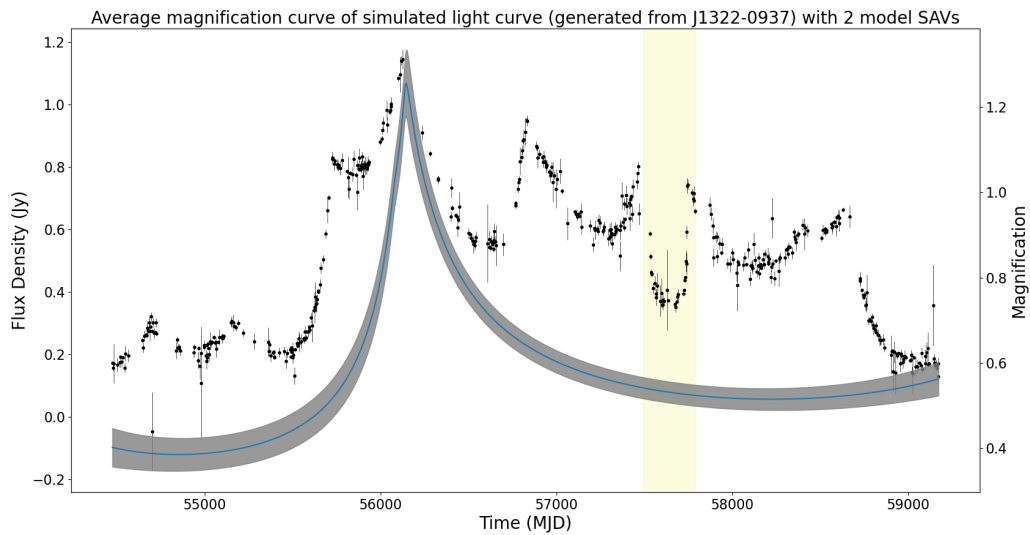
Figure 3.13: Weighted average magnification curve from fitting the 2-SAV model to the given artificial light curve.
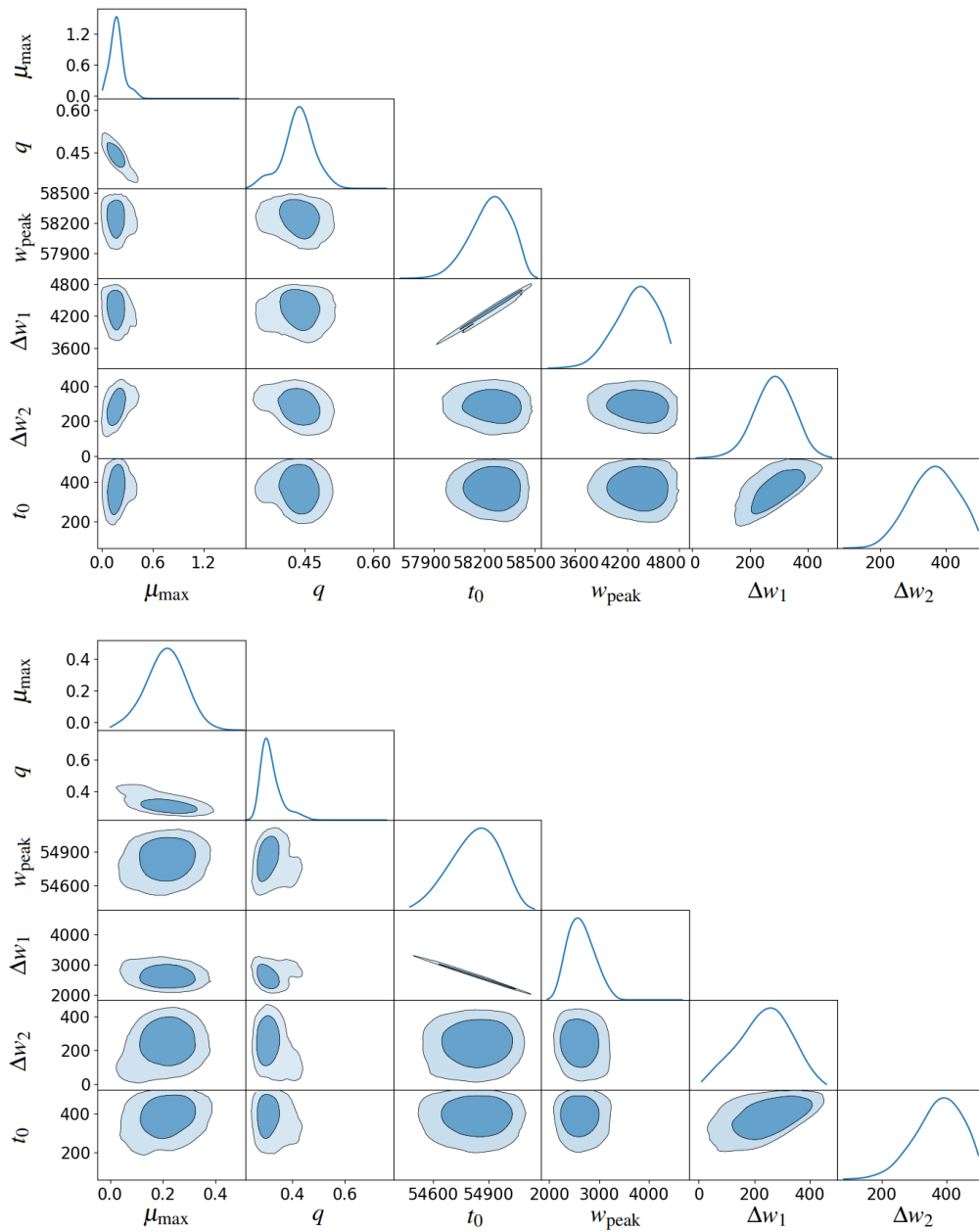
Figure 3.14: Triangle plot illustrating the posterior over the parameters in Θ_{mag} from fitting the 2-SAV model to the artificial light curve. The top plot includes the parameters for the first SAV event and the bottom plot includes the parameters for the second SAV event.

It seems that, in both cases, the magnification curves are very broad and only cap-

turing a single peak, indicating that they have been fit to some long-term variation in the light curve. The 'true' SAV event is not identified at all. If these are the fits that the procedure has determined to best describe the data, then the fitting procedure may need to be modified, as it has failed at identifying the SAV activity in this source. Of course, these results should be taken with a grain of salt, as it would be unwise to measure the fitting procedure's performance from fits to a single artificial light curve. Given more time, we could run this analysis over many such light curves, which would give a much better idea of the overall performance, and allow us to quantify our belief that the procedure is correctly identifying SAV.

While the fitting procedure does not run as well as hoped on a full simulated light curve, there are ways to explore its current limits. By generating some simpler data (fewer features, more pronounced SAV activity, etc.), it is possible to explore what the procedure can and cannot handle. This would require be a long, manual process of repeatedly generating data, performing the fitting and then analysing results, but it would give great insights into how the procedure can be improved. The example given below is designed to illustrate the procedure running on the most complicated data set which has yielded a correct result.

We begin by producing a cubic with parameter values $t_1 = 55000$, $t_2 = 59000$, $S_1 = 0.3$ and $S_2 = 0.9$. We place a single SAV event with parameter values $\mu_{max} = 1.422$, $q = 0.4433$, $t_0 = 56035$, $w_{peak} = 524.5$, $\Delta w_1 = 178.4$ and $\Delta w_2 = 278$. Finally we add in a single flare with parameter values $S_{peak} = 0.5$, $t_{peak} = 56789$, $r = 100$ and $d = 200$. Following the same procedure as before, we sample the resultant curve according to the cadence distribution of a random source and similarly sample the errors from another random source. This process gives us a simulated curve with a single SAV placed on a rising slope, with a single flare placed just to

the right. The spacing between the peaks of the SAV event has been selected to be very close to the spacing between the right-most SAV event and the flare peak. This creates two similar U-shaped features in the curve, making it difficult for the fitting procedure to identify which is the correct event. The results of running the procedure on this data set are shown in fig. 3.15, fig. 3.16 and fig. 3.17 for the background, 1-SAV and 2-SAV models respectively.
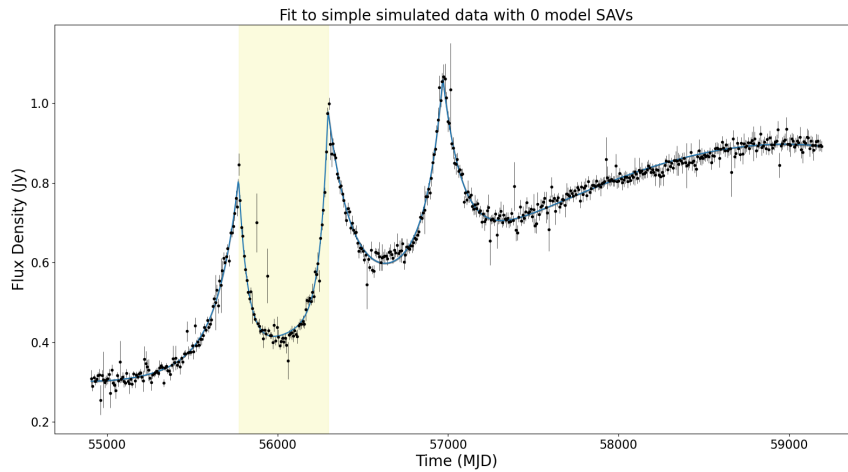
Figure 3.15: Weighted average curve from fitting the background model to the simple artificial light curve.
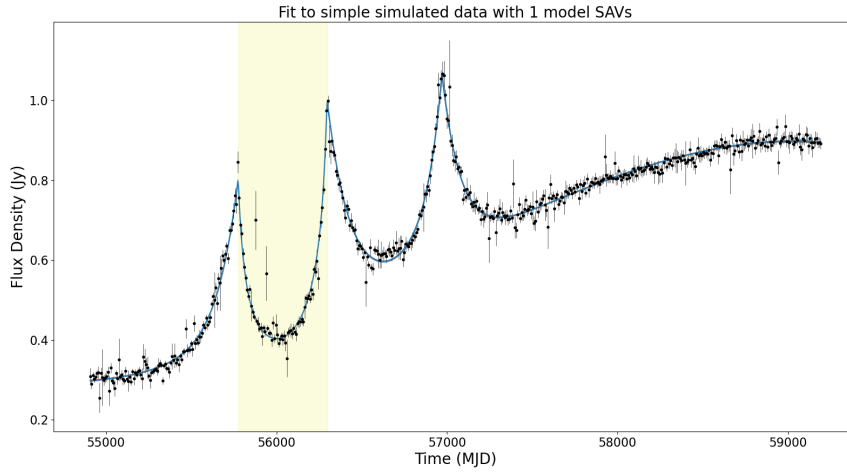
Figure 3.16: Weighted average curve from fitting the 1-SAV model to the simple artificial light curve.
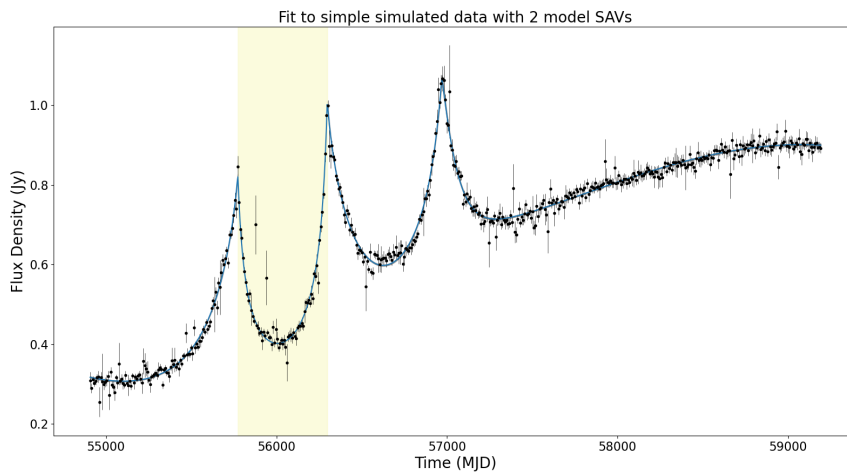
**Figure 3.17:** Weighted average curve from fitting the 2-SAV model to the simple artificial light curve.

In each case, we see a good fit to the data which is almost identical between models. In the cases of the 1-SAV and 2-SAV model, we can once again produce weighted average magnification curves from the posterior samples, which

are shown in fig. 3.18 and fig. 3.20. The triangle plots illustrating the joint posterior over Θ_{mag} are shown in fig. 3.19 and fig. 3.21
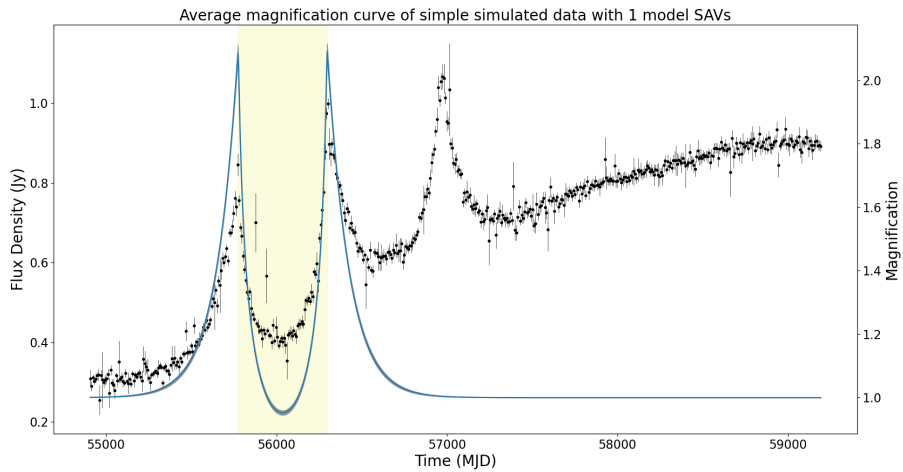
Figure 3.18: Weighted average magnification curve from fitting the 1-SAV model to the simple artificial light curve.
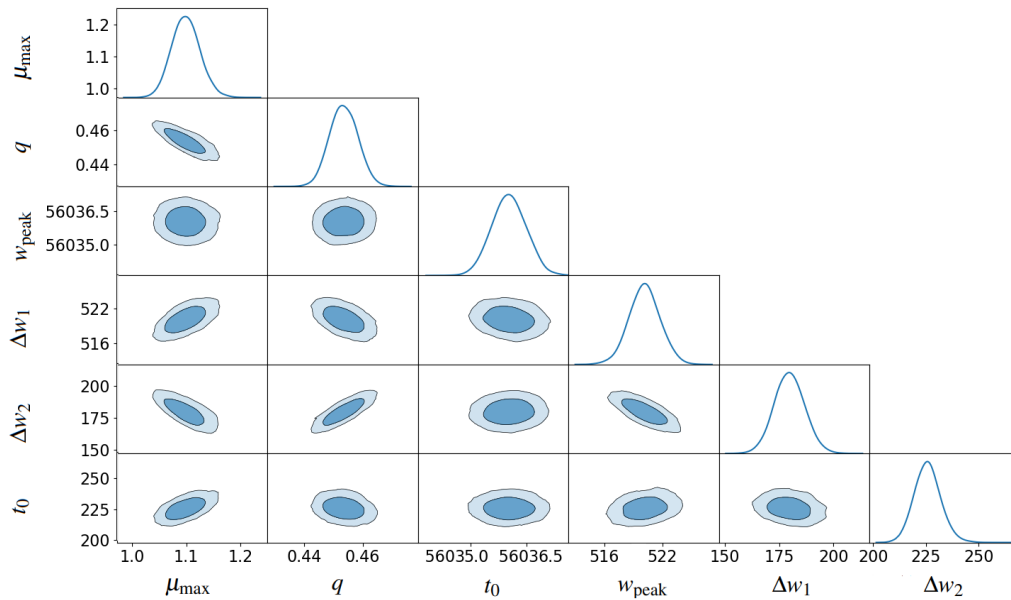
Figure 3.19: Triangle plot illustrating the posterior over the parameters in Θ_{mag} from fitting the 1-SAV model to the simple artificial light curve.
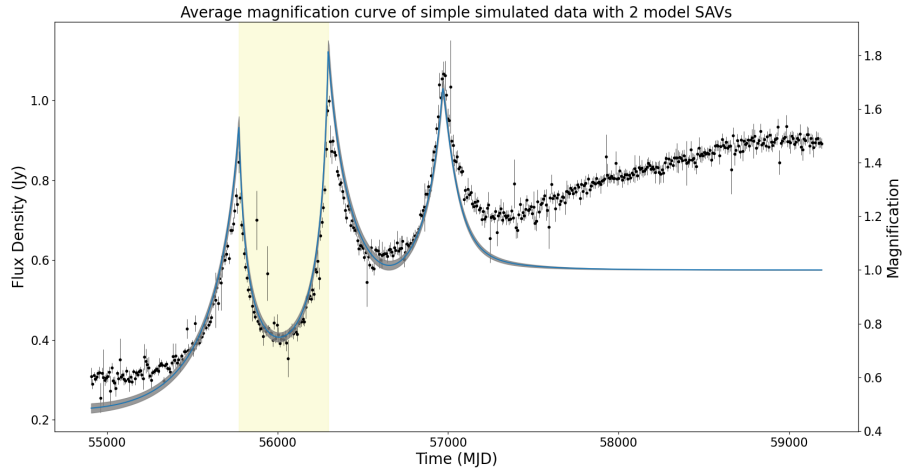
Figure 3.20: Weighted average magnification curve from fitting the 2-SAV model to the simple artificial light curve.
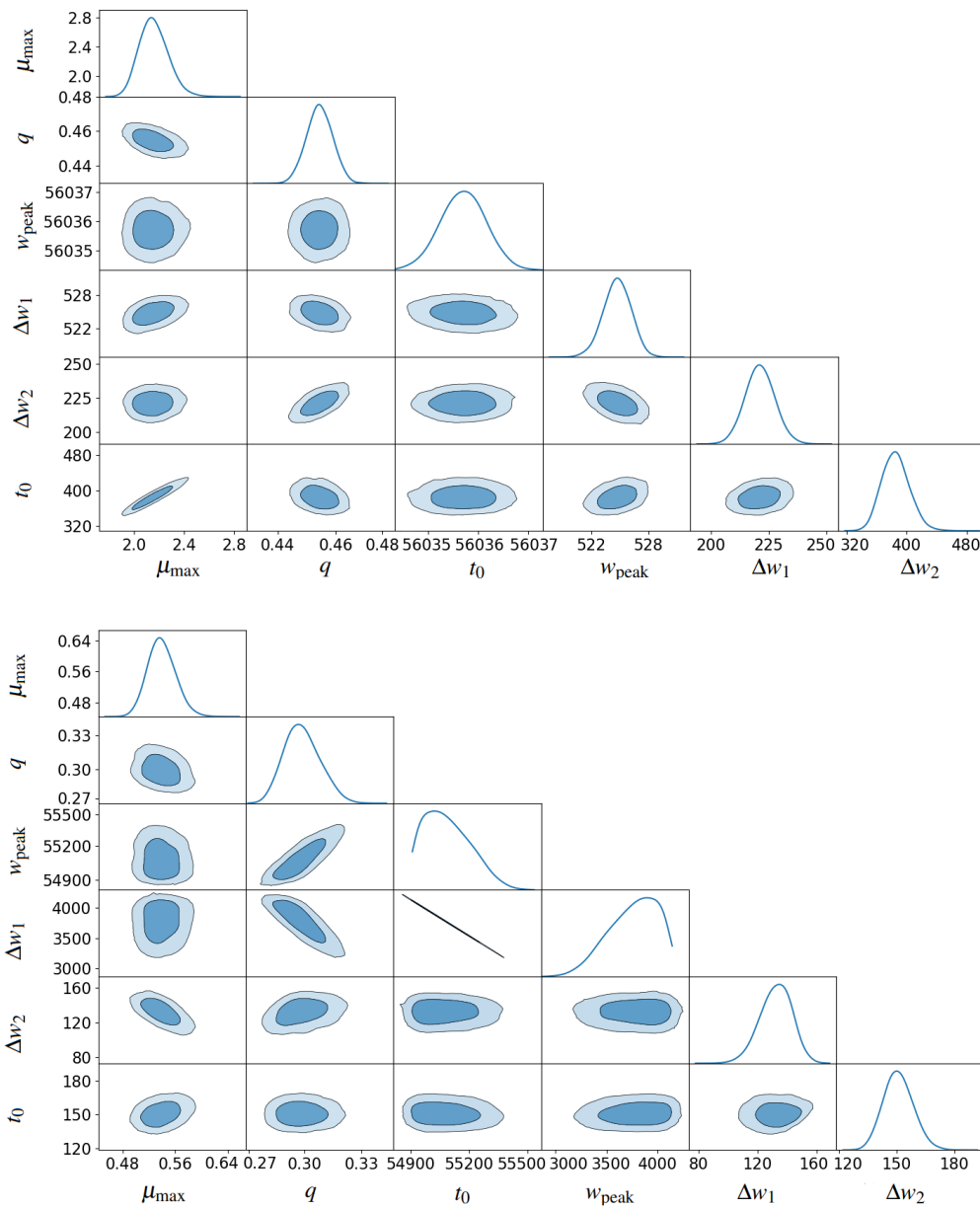
Figure 3.21: Triangle plot illustrating the posterior over the parameters in Θ_{mag} from fitting the 2-SAV model to the simple artificial light curve. The top plot includes the parameters for the first SAV event and the bottom plot includes the parameters for the second SAV event.

It is evident that the procedure finds the two U-shaped features well, with the 1-SAV model clearly preferring the left-most feature. We can then see in figure

fig. 3.22 that the posterior model probabilities indicate that the 1-SAV model is preferred in this case. This is exactly what would be expected from a working fitting procedure. A single artificial SAV event was placed in the data, and *a posteriori* the results indicate that a 1-SAV model best describes that data, as the background model lacks complexity and the 2-SAV model suffers from an Occam penalty for introducing additional parameters without significantly improving the fit. Ultimately this demonstrates that there is certainly potential for the fitting procedure introduced in this chapter. While it is not yet able to handle more complicated data sets, this example demonstrates that the framework built here is sound.
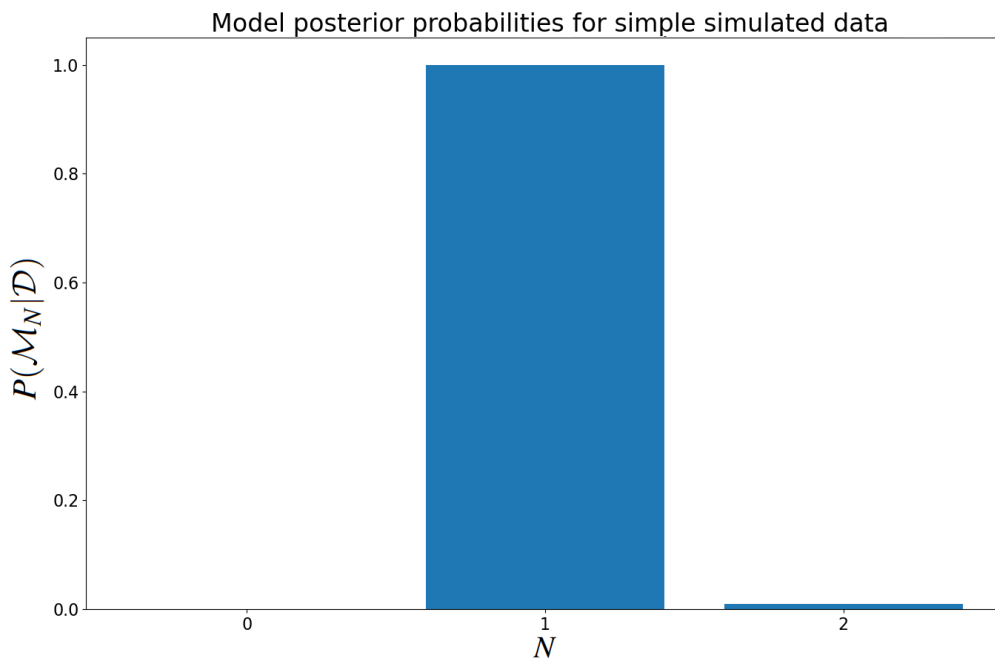
Figure 3.22: Bar plot indicating the posterior probability of each model given the data from the simple artificial light curve, where model $\mathcal{M}_N$ expects N SAV events.

3.4.2 OVRO data

To get an idea of what the current fitting procedure does/does not identify to be SAV events, we would ideally run the procedure over the entire set of OVRO 15GHz light curves, but ultimately this was not possible in the time available. Instead, it is still a useful exercise to focus on one OVRO source and see whether or not the procedure identifies any SAV. We select J1033+56051, one of the original SAV candidates (V2017A), with one event previously identified, centred at around MJD 56469. It should be noted here that just because a candidate SAV event has been identified in this source, we cannot take our procedure's ability to recover that event as a measure of how successful it is at identifying SAV in general, since current evidence cannot definitively tell us that the candidate event is truly a result of SAV, and there may be other variability in the light curve which has resulted from SAV that has been overlooked thus far. Nevertheless, it is still interesting to see whether the fitting procedure does recover that event, or find any others in the same light curve.

We run the procedure over the OVRO 15GHz light curve of J1033+6051, testing the same three models we used for the artificial light curve. Once again, we used the evidence estimates to find the posterior probability for each model $P(\mathcal{M}_N|\mathcal{D})$, as shown in fig. 3.23.
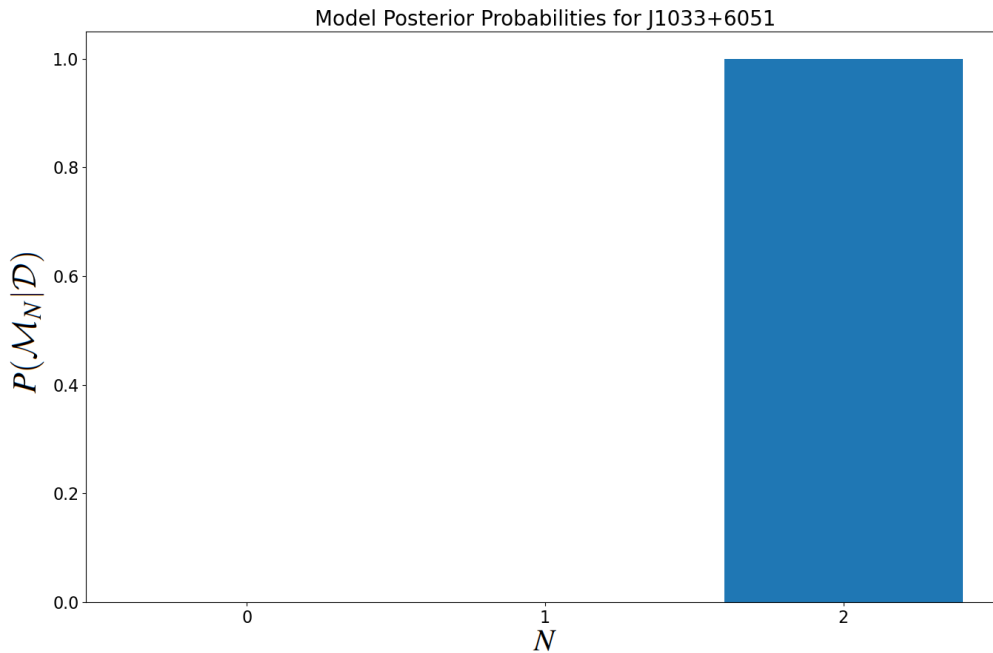
Figure 3.23: Bar plot indicating the posterior probability of each model given the data from J1033+6051 OVRO 15GHz light curve, where model $\mathcal{M}_N$ expects N SAV events. Note that all bars are displayed, but those for $N = 0$ and $N = 1$ are invisibly small.

Once again, the 2-SAV model is strongly favoured over the other two models, meaning our procedure is reporting the presence of at least two SAV events. This may well be possible; there may be an event which was overlooked originally, or that had not yet been observed within the data available in V2017A. Once again, we produce a weighted average curve for each model to inspect the fits (figs. 3.24 to 3.26). This time, the candidate SAV event reported in V2017A is highlighted in yellow.
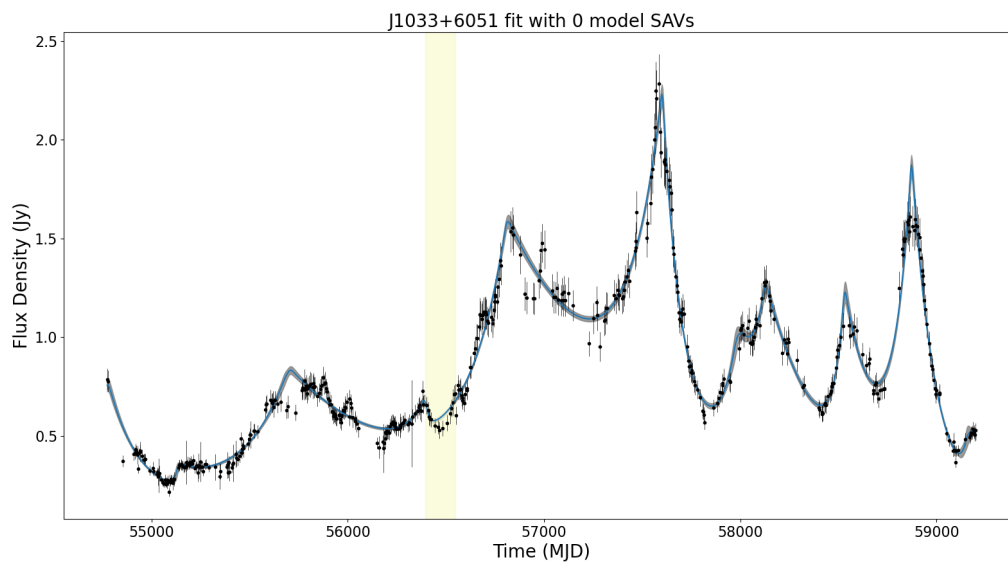
Figure 3.24: Weighted average curve from fitting the background model to the the OVRO 15GHz light curve of J1033+6051.
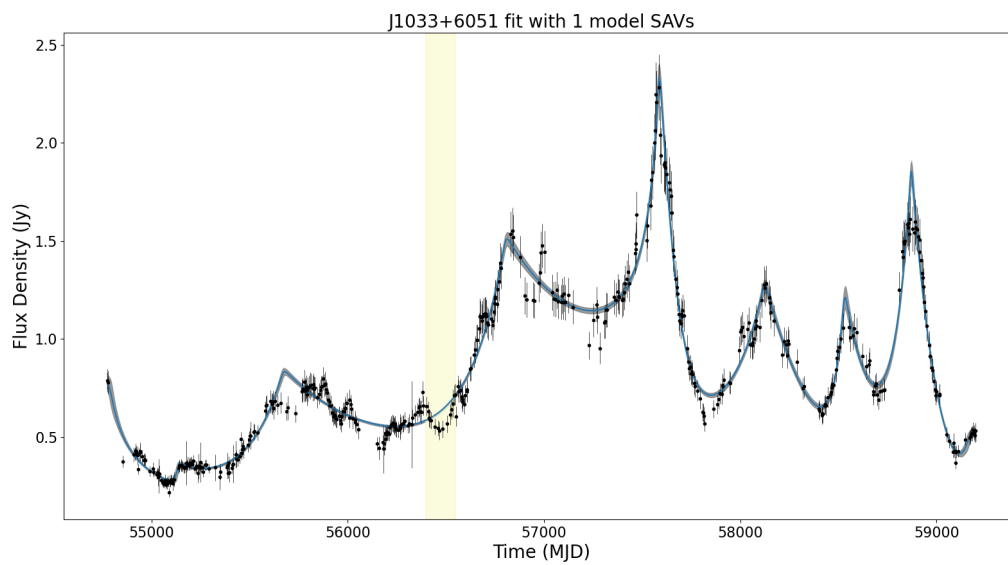
Figure 3.25: Weighted average curve from fitting the 1-SAV model to the OVRO 15GHz light curve of J1033+6051.
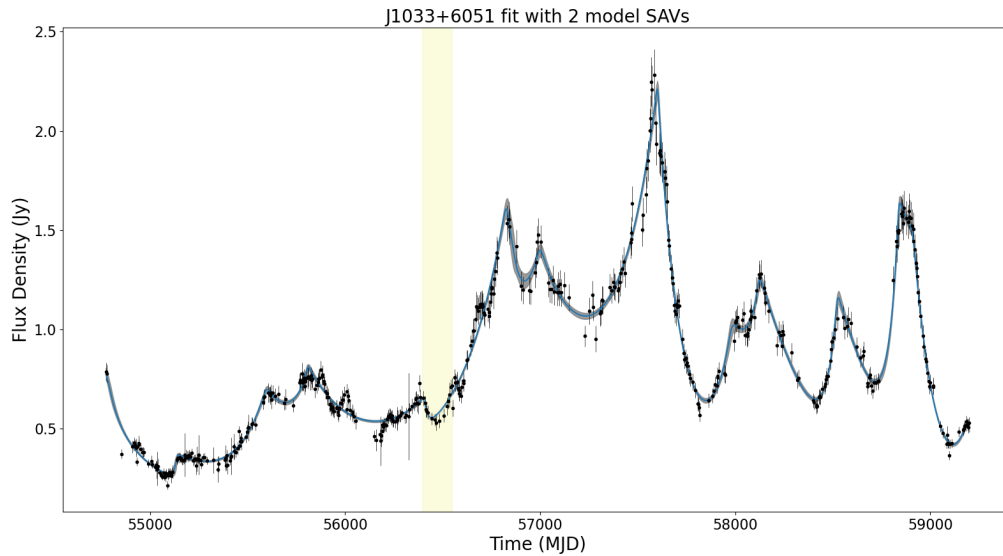
Figure 3.26: Weighted average curve from fitting the 2-SAV model to the OVRO 15GHz light curve of J1033+6051.

Once again the fits appear to be rather good, but this doesn't tell us anything about the variability that the procedure has identified as SAV. Weighted average magnification curves for the 1-SAV and 2-SAV models are shown in figs. 3.27 and 3.29 respectively, and their respective triangle plots are shown in figs. 3.28 and 3.30.
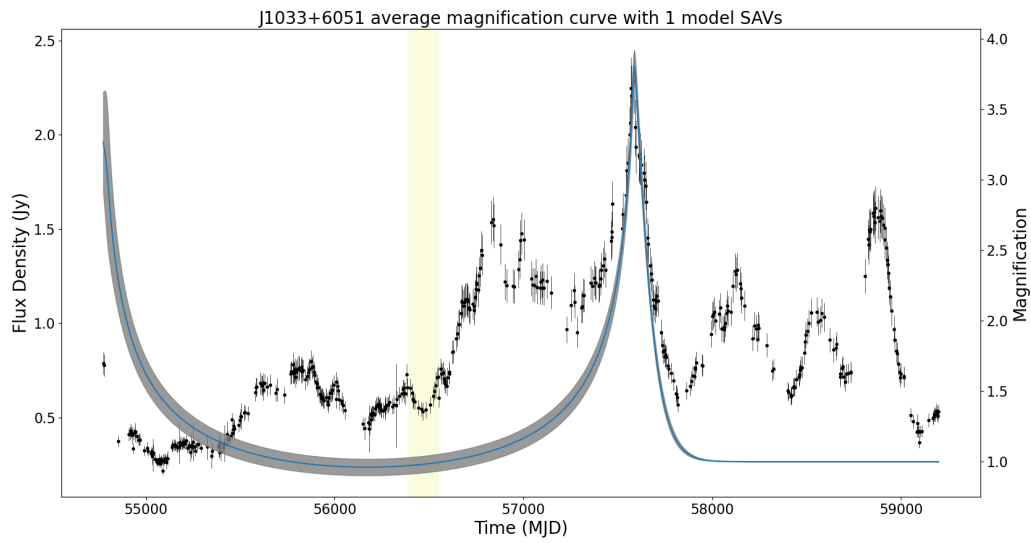
Figure 3.27: Weighted average magnification curve from fitting the 1-SAV model to the OVRO 15GHz light curve of J1033+6051.
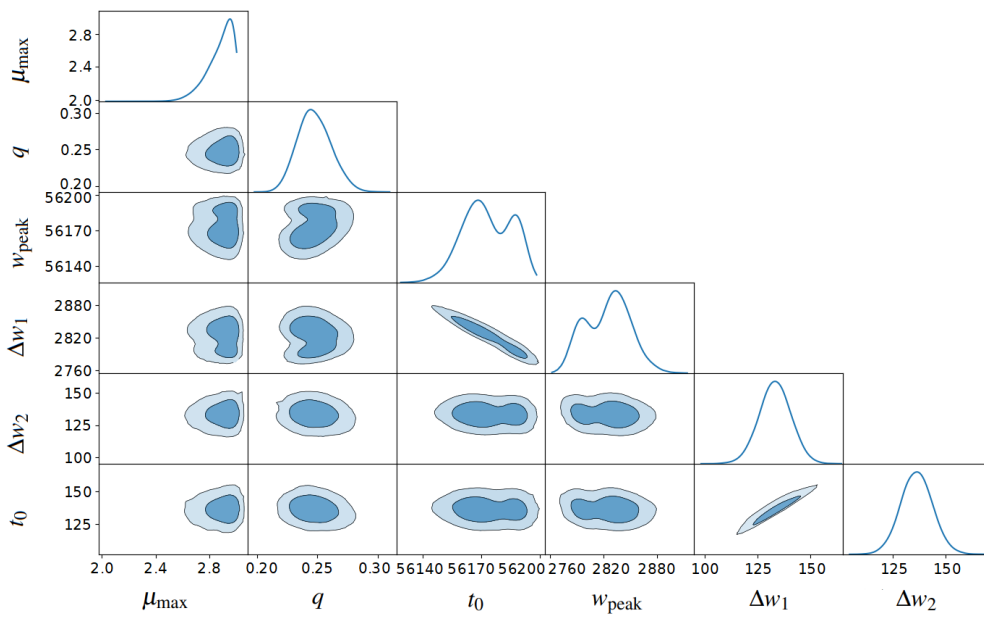
Figure 3.28: Triangle plot illustrating the posterior over the parameters in Θ_{mag} from fitting the 1-SAV model to the OVRO 15GHz light curve of J1033+6051.
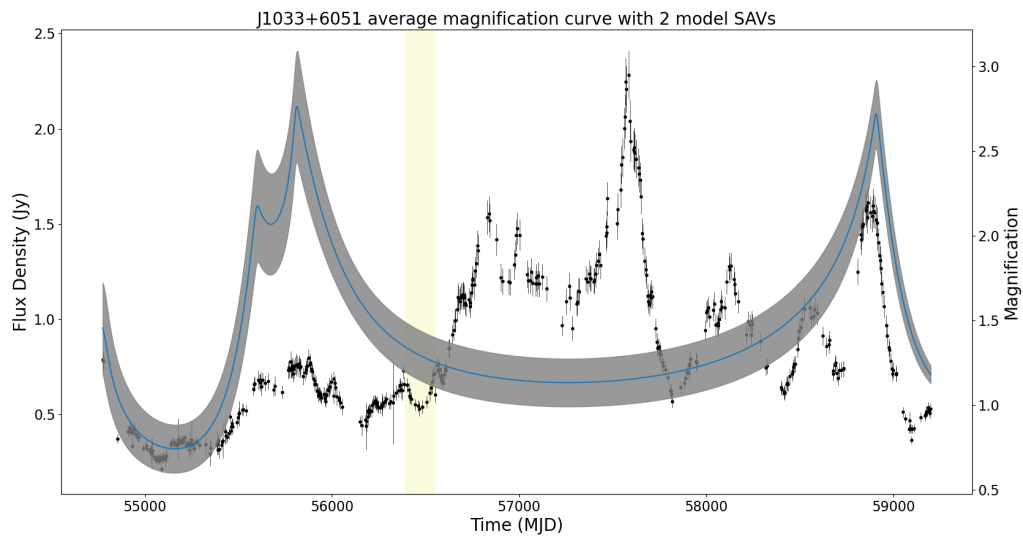
Figure 3.29: Weighted average magnification curve from fitting the 2-SAV model to the OVRO 15GHz light curve of J1033+6051.
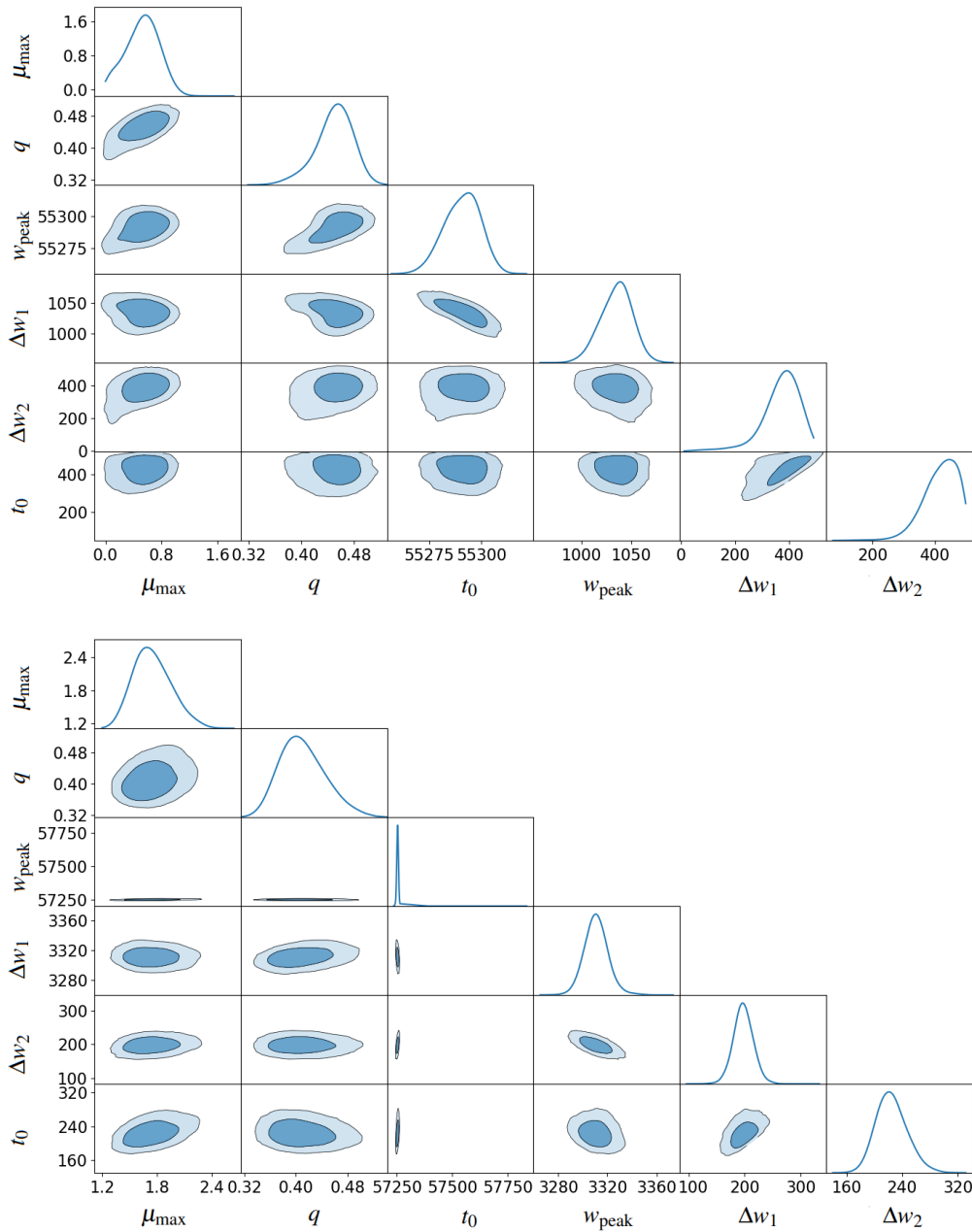
Figure 3.30: Triangle plot illustrating the posterior over the parameters in Θ_{mag} from fitting the 2-SAV model to the OVRO 15GHz light curve of J1033+6051. The top plot includes the parameters for the first SAV event and the bottom plot includes the parameters for the second SAV event.

Just like with the artificial light curve, we see that the procedure has fit very broad

magnification curves, which is fixed to a single peak in the data, and is used by the model to fit to some long-term variations. Without running a full analysis on artificial light curves, as described previously, it is not possible to make any definitive statements about the procedure's ability to recover true SAV activity, meaning these fits could be indicative of real events, although it does seem quite unlikely. It would be possible (although very slow) to follow this up using the physically motivated magnification curve model, to see if the events found correspond to a physically viable lensing scenario, but that is beyond the scope of the work done here.

3.4.3 Discussion

The framework introduced in this chapter certainly goes a decent way towards fitting an SAV model a given light curve, with the existing procedure providing very good fits to the target data. In its current form, the procedure does have some clear shortcomings, but overall the results are promising enough that with more work, the procedure could very possibly be refined to successfully and reliably identify SAV events.

In future, the priority would certainly be to investigate why models with multiple, very broad magnification curves are preferred in these preliminary tests. It seems quite possible that the model has been allowed to be too flexible thus far, and would need to be restricted in future. The first step in doing so would likely be to modify the priors over the model parameters. The current priors were generally well-motivated, but they were also selected to give very naive assumptions about the model parameters. Thinking with a Bayesian mindset, we have now gathered

more data by investigating these early results, which has updated our prior assumptions about the model parameters and their relationships. For example, now that we know the model finds very good fits with broad magnification curves, it may be wise to impose a log-uniform prior on w_{peak}, which would still allow wide magnification curves when necessary, but generally would shift the bulk of the prior volume towards lower values of w_{peak}, meaning we are more likely to fit a narrower magnification curve. Further investigations would likely reveal other opportunities to update the original model. In any case, the fitting procedure should be run with several more light curves before making any decisions about what changes should be made.

In addition to the fitting procedure, the work described here has led to the development of a novel approach for simulating artificial light curves as described in section 3.3. These light curves could prove beneficial to a number of projects involving other OVRO collaborators, in addition to playing an important role in any future development of this fitting procedure. Given more time, it would likely prove very useful to perform an analysis of the performance of the fitting procedure over a large set of simulated light curves with placed events (as previously discussed). Being able to quantify how the model performs with a wide range of data is an essential step towards the goal of reliably identifying new events within the OVRO data set. We could also then be more confident in refining the SAV model, as any changes would be reflected in the model performance. There are many possible changes which could be made: modifying the model priors; adding/removing parameters to/from the model; further exploring the use of a robust weighting scheme; or a number of other possible additions.

Finally, rather than having to run the procedure with multiple different models

and manually identify which is most probable *a posteriori*, it might well be beneficial to incorporate the number of SAV events into the model as a free parameter. This could be done with the adaptive method which has been put to use here for the exponential flares, but a separate 'N' could be included for the number of SAV events. Alternatively, chapter 4 explores a possible transdimensional nested sampling framework, which would suit this kind of problem perfectly.

Chapter 4

Transdimensional Nested Sampling

There exists a subset of Bayesian inference problems where there are N components in some signal, where N is *a priori* unknown. This type of problem could arise in many fields, but is fairly prevalent in astronomy, e.g. finding the number of sources within some noisy astronomical image, or identifying transient events within some time-series data. Each component can be individually modelled as a 'particle' described by K parameters from Θ and the components can then be combined in whatever way suits the problem (e.g. summation) to complete the model. In this case, the dimensionality of the problem is not fixed. Using classical nested sampling techniques, these types of problems would usually be solved by using model selection to compare N_{max} models $\{\mathcal{M}_N\} = \mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_{N_{\text{max}}}$, where each model expects N particles and therefore Θ_N is $N \cdot K$ dimensional. As discussed in section 2.3.3, the evidence naturally encapsulates Occam's razor and therefore model selection should allow us to identify the posterior distribution for the number of particles.

While this approach does provide a solution to these types of problems, it

requires N_{max} separate nested sampling runs, meaning it is computationally inefficient, especially if N_{max} is large. An alternative approach is the so-called 'adaptive' nested sampling method (Higson et al. 2019, Hee et al. 2016, Chua et al. 2018). It was this approach which was used in chapter 3 for SAV identification and light curve simulation. The adaptive method uses a single model which incorporates N as a free parameter, as well as a further $N_{\text{max}} \cdot K$ parameters which for the N_{max} particles. We then modify our likelihood function $\mathcal{L}(\Theta)$ to consider only the first N particles and ignore the rest. The rest of the nested sampling run can then continue as normal and we can recover a joint posterior $\mathcal{P}(\Theta)$ from the posterior weights. By marginalising over the other parameters, we can then find the posterior on N and identify the most likely number of components in the problem. This approach requires only a single nested sampling run, but maintaining $N_{\text{max}} \cdot K$ parameters for each live point when many of those parameters will be ignored is an inefficient use of compute resources (both time and memory), again, especially when N_{max} becomes large. This is even more evident in cases where we have multiple types of particles, each with its own N.

We therefore ideally want to find an approach in which N (and therefore the dimensionality of the space) is allowed to vary freely, but we aren't maintaining parameters which aren't going to be used by the algorithm. Reversible Jump Markov Chain Monte Carlo (Green 1995) is an MCMC algorithm originally designed to deal with inference problems where the dimensionality is not fixed, therefore allowing model selection to be performed. While on its own, the algorithm is prone to the same shortcomings as any other MCMC approach (inability to deal with parameter degeneracies and multiple modes), it does provide a useful framework involving 'jumps' which change the dimension of the model,

which can be adapted for use with nested sampling. This was the approach of Brewer (2014), who applied this framework to create a trans-dimensional variant of the earlier 'diffusive nested sampling' (DNS) (Brewer et al. 2009). This is a powerful approach, however DNS relies on multiple tuning parameters, which complicate the method and may necessitate multiple preliminary runs or analytical work to correctly tune. Furthermore, as elaborated by Brewer & Foreman-Mackey (2016), it is not easy to implement models using DNS. Proposal distributions must be constructed *ad hoc* for each given sampling problem, which requires intimate knowledge of both DNS, as well as the intricacies of that particular problem.

This chapter explores the idea of a generalised trans-dimensional nested sampling variant which uses a similar reversible jump framework, as well as slice sampling *à la* POLYCHORD. I then present the results of testing the algorithm in its current state on a toy problem, before discussing the future development of this approach.

4.1   Exploring a Transdimensional Space

As with most variants of nested sampling, the key distinction of this approach is the method in which the space is explored when drawing new live points. As with any MCMC nested sampling method, we are concerned with ensuring that the chains are uniformly distributed within the given iso-likelihood contour. If each new live point begins as a copy of an existing live point (which is already uniformly distributed), then by evolving the copied point using a Markov chain that satisfies detailed balance i.e. $P(x'|x) = P(x|x')$, we can guarantee that the resultant live point will also be uniformly distributed.

4.1.1 The Particle Space

Normally, nested sampling is focused on exploring the parameter space, where a given point Θ is a vector with length equal to the dimensionality of the space. In the problems being considered here, however, Θ is instead a vector with $NK + 1$ dimensions made up of N particles, each with K parameters, i.e. $\Theta = (\theta, N)$, where $\theta = (\theta_1, \theta_2, \cdots, \theta_N)$ and each θ_i is a single K-parameter particle. It is then helpful to shift our thinking to consider a K-dimensional space where each of the N particles which make up Θ can exist within the same space. In this way we have effectively 'flattened' θ from an NK-dimensional parameter space into a K-dimensional particle space. Importantly, this means that any particle in this space is directly comparable to another. This approach does require some consideration about how to deal with live points, however. For example, how does one generate an initial set of live points? Traditionally, the live points will be drawn from the prior distribution (in the parameter space), however in the transdimensional case, each live point will have a different number of parameters. A simple way of dealing with this is to implement a distinct prior on N, $P(N|\mathcal{M})$, and then sampling N for each initial live point. For each point, we can then sample N distinct K-dimensional points from the the prior distribution (defined in the particle space here) and concatenate them together with N to give the full live point with $NK + 1$ dimensions. It should be noted here that defining a separate prior on N does not cause any problems for Bayesian Inference. Because N is a free parameter in this scenario, the distribution we're looking to acquire is the joint posterior on θ and N:

$$P(\Theta|\mathcal{D},\mathcal{M}) = P(\theta,N|\mathcal{D},\mathcal{M}) = \frac{P(\mathcal{D}|\theta,N,\mathcal{M})P(\theta,N)}{P(\mathcal{D}|\mathcal{M})}, \qquad (4.1)$$

where, from the product rule we know that

$$P(\theta,N) = P(\theta|N,\mathcal{M})P(N|\mathcal{M}). \qquad (4.2)$$

This shows that our two distinct priors are really just two parts of the joint prior on θ and N. The prior on N could be any distribution and so may well be unbounded, e.g. a logarithmic prior which make smaller Ns more likely, with larger Ns becoming asymptotically less probable. Alternatively, the distribution may be bounded on either side by some $N_{\min}, N_{\max} \in \mathbb{N}$, $N_{\min} <= N_{\max}$. For simplicity, the rest of this section will assume a bounded prior on N.

With all of that in mind, the first step is to define exactly how to transition from one point to another in a transdimensional space. There are two types of transitions to consider: transitions where the dimensionality stays the same and transitions where it changes.

4.1.2 Transitions with matching dimensionality

When proposing a transition to a new point with the same dimensionality as the point being evolved, the process is much the same as with regular nested sampling, only now we can leverage the particle space if needed. For this reason, any existing method for exploring the space will suffice. Slice sampling is an MCMC algorithm which allows samples to be drawn from a given distribution. It was introduced as a method to sample from a given posterior distribution, but was shown by Handley et al. (2015) to be ideally suited for sampling uniformly within
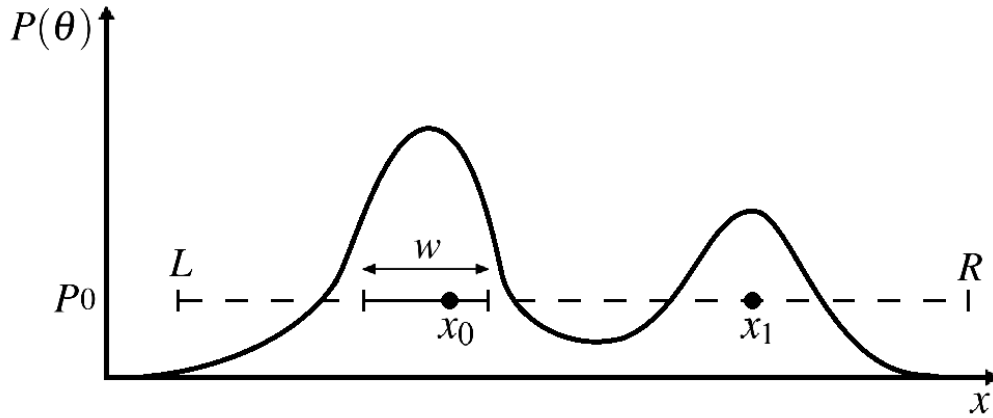
Figure 4.1: One dimensional depiction of slice sampling. Reproduced from Handley et al. (2015)

an iso-likelihood contour. The method also scales well with dimensionality, and naturally satisfies detailed balance, making it a suitable method in this case.

The procedure begins by choosing a 'slice' at a given probability P_0 from the target distribution. One then samples uniformly from the region defined by $p(\theta) > P_0$. This is done by first selecting a point x_0 from the region (i.e. $P(x_0) > P_0$). A line can then be drawn through x_0, horizontally across the slice. In the one dimensional case (shown in fig. 4.1), this is straightforward, however with a multivariate distribution, there are infinite such lines. In this case, a direction $\hat{u}$ can be chosen randomly from some other distribution. Bounds L and R are set at x_0, then repeatedly 'stepped out' by some step size w in opposite directions down the line until they land outside of the region enclosed by $P(\theta) = P_0$. A new point x_1 is then sampled within these bounds. If x_1 is not in the slice ($P(x_1) < P_0$), L or R is replaced with x_1, ensuring x_0 is still within the slice, and sampling is attempted again.

In the nested sampling case, the slice is analogous to an iso-likelihood contour,

where $P_0 = L_*$ and x_0 is the live point being evolved. All that is then required is to find some w and $\hat{u}$. In order to choose a direction $\hat{u}$, one could select randomly from some distribution, however an alternative approach becomes apparent when considering the nature of the flattened particle space. We can find a direction $\hat{u}$ by choosing two random particles from the flattened space (i.e. we can select particles from any existing live point) and subtracting one particle from the other (being careful not to choose the same particle twice so we don't produce a zero vector). If we repeat this N times (where N here is taken from x_0), we can concatenate the resulting vectors together to create our directional vector $\hat{u}$. One benefit of this approach, compared to randomly selecting a direction, is that the algorithm should be better at exploring spaces with degenerate parameters. Since $\hat{u}$ is generated from the existing live points which are known to be within the iso-likelihood contour, it is more likely to be directed along the length of the degenerate contour, allowing it to explore the full contour rather than getting stuck locally. Moreover, the choice of step size w becomes easier. Ideally, w should be roughly the typical width of the slice. If it is too small, the number of 'stepping out' steps increases making the process less efficient, and we risk ignoring other modes from the distribution. If it is too large, we're likely to overstep the actual bounds of the slice by a large amount, and it will likely take many attempts at sampling and shrinking the bounds back in, again decreasing efficiency. Since $\hat{u}$ is produced from existing live points, its magnitude is likely to be of the same order as the width of the slice. We can therefore set $w = a|\hat{u}|$, where a is drawn randomly from $(0,1]$.

4.1.3 Transitions with varying dimensionality

Proposing a transition to a point x_1 with different dimensionality to x_0 is more complicated, although in principal many things stay the same. Crucially, as long as detailed balance is maintained, the algorithm will be guaranteed to uniformly sample from within the iso-likelihood contour. The difficulty arises because the nature of transitioning up/down in N is fundamentally different than exploring a space with fixed dimensionality. Firstly, N is discrete, meaning that approaches like slice sampling would not work (without modification). Secondly, and more importantly, any transdimensional nested sampling algorithm must have mechanisms for the addition and removal of parameters to/from the parameter vector Θ, but those mechanisms must still maintain the detailed balance constraint. When considering transitions with fixed dimensionality, this constraint is easier to satisfy, since it is possible to have a single procedure for proposing new points, whereas adding/removing parameters requires two separate procedures. Here I introduce two such mechanisms for 'stepping up/down' in N. Although it isn't immediately clear whether these methods satisfy detailed balance, the assumption going forward will be that they do, with an analysis in section 4.2 reviewing whether this assumption holds. For now, in both cases, only transitions with a step size of 1 will be considered, e.g. stepping from $N = 4$ to $N = 5$ and vice versa.

4.1.3.1 Stepping up in N

When stepping up in N, we are looking to extend the parameter vector Θ in such a way that the proposed point x_1 is within the region enclosed by the iso-likelihood contour. The goal is to produce an additional K-parameter particle and append that

to Θ. A naive approach would be to repeatedly generate some parameter values from the prior distribution until a combination is found which places the x_1 within the iso-likelihood contour. This would be rather inefficient, however, since the posterior mass is usually confined to a fairly small region of the prior volume. This would be especially noticeable as the algorithm progressed and the iso-likelihood contour shrank. A better alternative would be to generate some parameter values which are informed by the existing live points. By making use of the particle space, we can copy the values from an existing particle chosen randomly from one of the other live points (with all particles having equal probability) and use those values to extend Θ. By doing so, we are leveraging information we already have about the iso-likelihood contour from the existing live points. Since the particle being copied clearly does a good enough job of describing some component of the data (at least well enough for that live point to fall inside the iso-likelihood contour), it is more likely that using those parameter values would provide a valid proposal when compared to drawing from the prior.

4.1.3.2 Stepping down in N

Stepping down in N involves removing a K-parameter particle θ_i from Θ and, again, proposing a point x_1 which is valid (i.e. within the iso-likelihood contour). The simplest way to do this is to randomly select a particle from Θ, with equal probability for each particle, and delete it. It is easy to conceive of alternative methods which involve aggregating the existing parameters, (e.g. selecting two particles and replacing them with a single particle containing the averages of their parameter values), however it is also easy to see how such methods might introduce bias when exploring the space. The method of deleting particles is also

conceptually closer to the reverse of copying particles, making it easier to consider detailed balance.

4.1.4 Proposing Transitions

With appropriate procedures in place for each type of transition, the next step is to consider how a transdimensional nested sampler should use these transitions to explore the space. For a traditional nested sampling algorithm using MCMC to evolve live points, the goal is to end up with a new point x_1 sampled uniformly from the iso-likelihood contour which has gone through enough steps to sufficiently de-correlate it from the original live point. Too many steps adds nothing but additional, unnecessary computation. A series of $\sim \mathcal{O}(NK)$ steps should be enough to sufficiently de-correlate the two points. In the case of a transdimensional sampler, however, we must also take additional steps to allow N to vary. A sufficient approach which would allow us to fully explore the space would be to perform NK slice sampling steps (i.e. keeping N fixed), then make one more step choosing to either increase N by 1, decrease N by 1, or keep N the same, and following the appropriate procedure. This procedure can then be repeated a handful of times to be absolutely sure the new point is de-correlated from the original point, while ensuring the transitions provide enough mobility to explore the space.

The process of choosing whether to step up/down or keep N the same is best done probabilistically, however this raises the question of what the associated probabilities should be with each type of transition. This is easier to think about if we consider that this can be viewed as a Markov chain, where the states corres-
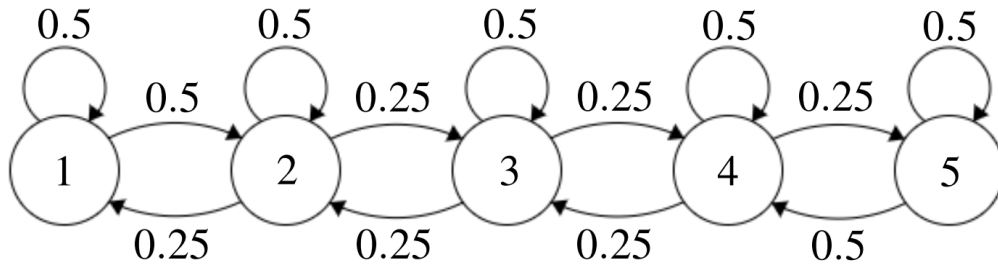
Figure 4.2: State-transition diagram of a naive approach to choosing transitions with $N_{\min} = 1$ and $N_{\max} = 5$.

pond to the current value of N for the live point being evolved, and we have a certain probability of transitioning to each other state each time we choose whether or not to step up/down in N (where these probabilities can be 0). Consider the case where we naively decide to assign a 50% chance to keep N the same and a 50% chance to change it. If we do decide to change N, we then have a further 50% chance to step up and a 50% chance to step down. This is of course unless $N = N_{\min}$ or $N = N_{\max}$, in which case the chance of respectively stepping down/up should be 0%, so we can assign a 100% chance of stepping the other way. This is best depicted by the state-transition diagram in fig. 4.2.

It becomes clear from this diagram, if it wasn't already, that there may be a problem with this naive approach. To rule out any bias in the approach, we want to create Markov chain in which there is an equal probability of being in any given state. Put formally, the stationary distribution ψ of this chain should be uniform. We can identify the actual stationary distribution for this chain by setting some initial values for ψ:

$$\psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \\ \psi_5 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}. \tag{4.3}$$

This vector currently describes the probability of the chain being in each state, assuming a uniform prior on N. By repeatedly applying the transition matrix T, in which each element $T_{i,j}$ is the probability of transitioning from state i to state j, ψ should eventually reach a stationary distribution such that $T^\mathsf{T}\psi = \psi$. Applying the transition matrix once gives us

$$T^\mathsf{T}\psi = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.25 & 0.5 & 0.25 & 0 & 0 \\ 0 & 0.25 & 0.5 & 0.25 & 0 \\ 0 & 0 & 0.25 & 0.5 & 0.25 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$$

$$\tag{4.4}$$

$$T^\mathsf{T}\psi = \begin{bmatrix} 0.15 \\ 0.25 \\ 0.2 \\ 0.25 \\ 0.15 \end{bmatrix}.$$

This vector represents the probability of the chain reaching each state after a single transition. From having an equal probability of the chain, it is now less likely for

the chain to have reached one of the 'edge' states. Repeated applications of the transition matrix find that the distribution ψ becomes stationary when

$$
\psi = \begin{bmatrix} 0.125 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.125 \end{bmatrix}, \tag{4.5}
$$

demonstrating even more pronounced edge effects. It seems that the naive system for proposing transitions in N has a natural bias away from the edge states. Fortunately, the framework is now in place to identify a new set of transition probabilities which do not impose the any bias in this way. Consider a more general Markov chain, again constructed so that each state is a different value of N. This is depicted in fig. 4.3 and the associated transition matrix would be

$$
P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N_{\max}} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N_{\max}} \\ & \vdots & \ddots & \vdots \\ p_{N_{\max},1} & p_{N_{\max},2} & \cdots & p_{N_{\max},N_{\max}} \end{bmatrix}. \tag{4.6}
$$

Note that for brevity of notation, we have assumed $N_{\min} = 1$, whereas in principal this can easily be extended to any $N_{\min}$ and $N_{\max}$. Our goal is to again identify a stationary distribution ψ, where we now have
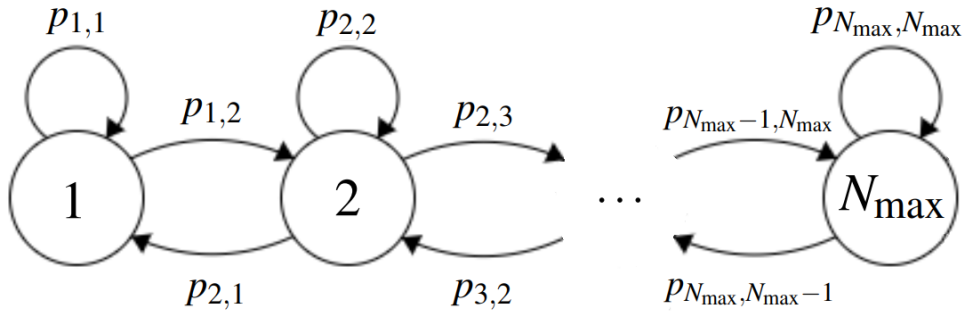
Figure 4.3: A more general state-transition diagram showing only transitions between adjacent states for visual clarity.

$$
\psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{N_{\max}} \end{bmatrix}.
\tag{4.7}
$$

Since this is the stationary distribution, we again have $T^\mathsf{T}\psi = \psi$, which gives the following system of equations:

$$
\pi_1 = \pi_1 p_{1,1} + \pi_2 p_{2,1} + \cdots + \pi_{N_{\max}} p_{N_{\max},1}
$$

$$
\pi_2 = \pi_1 p_{1,2} + \pi_2 p_{2,2} + \cdots + \pi_{N_{\max}} p_{N_{\max},2}
$$

$$
\vdots
$$

$$
\pi_{N_{\max}} = \pi_1 p_{1,N_{\max}} + \pi_2 p_{2,N_{\max}} + \cdots + \pi_{N_{\max}} p_{N_{\max},N_{\max}}
$$

$$
\tag{4.8}
$$

Because we are interested in a uniform stationary distribution, we have

$$\psi = \begin{bmatrix} \frac{1}{N_{\text{max}}} \\ \frac{1}{N_{\text{max}}} \\ \vdots \\ \frac{1}{N_{\text{max}}} \end{bmatrix}, \tag{4.9}$$

so the system of equations in (4.8) reduces to

$$
\begin{aligned}
\frac{1}{N_{\text{max}}} &= \frac{1}{N_{\text{max}}}\left(p_{1,1} + p_{2,1} + \cdots + p_{N_{\text{max}},1}\right) \\
&= \frac{1}{N_{\text{max}}}\left(p_{1,2} + p_{2,2} + \cdots + p_{N_{\text{max}},2}\right) \\
&\vdots \\
&= \frac{1}{N_{\text{max}}}\left(p_{1,N_{\text{max}}} + p_{2,N_{\text{max}}} + \cdots + p_{N_{\text{max}},N_{\text{max}}}\right).
\end{aligned}
\tag{4.10}
$$

This establishes that the condition for a uniform stationary distribution is to simply construct a Markov chain with a transition matrix where all of the columns sum to unity. This all assumes, of course, that a stationary distribution exists for the constructed Markov chain. This adds the additional conditions that the chain must be irreducible and aperiodic, as discussed in section 2.4.1. It is not particularly difficult to construct chains which meet these criteria, however when restricted to a step size of 1 when transitioning in N, it would require extra steps (and probably some additional tuning parameters) to come up with a generalised system for producing transition matrices which followed these rules. Instead, consider the chain depicted in the state-transition diagram in fig. 4.4. Here, we have a chain where transitions are allowed from any state to any other state, including itself, with equal probability. This chain would have a transition matrix where every

element has the same value, $T_{i,j} = 1/N_{\max}^2$. This would, of course, require us to extend our procedures for transitioning up and down in N to allow for a step size of more than 1, which in turn requires an extension of the mechanisms for stepping up/down defined earlier. Let ΔN be the difference between the new and old values of N. If ΔN is positive, we simply repeat the process of stepping up in N (i.e. copying a particle from another live point) ΔN times. Conversely, we can repeatedly delete $|\Delta N|$ particles to step down if ΔN is negative. In either case, we are effectively just repeating the existing method of stepping up/down the desired number of times.

With these extensions, the sampler should be able to fully explore the space, without being biased by the choice of transitions in N. In the context of a transdimensional nested sampling algorithm, each iteration will evolve the live point x_0 by first performing a series of NK slice sampling steps to move around with a fixed N as before, then a new value for N is selected randomly, with equal probability for all N where $N_{\min} \leq N \leq N_{\max}$. The algorithm can then propose a transition to a point with this new N, stepping up/down as appropriate and the proposed point will be accepted if it is within the iso-likelihood contour.

4.2   Results

The algorithm detailed in section 4.1 is capable of taking a model consisting of N K-dimensional particles and performing nested sampling with N as an additional free parameter, computing the evidence and providing samples from the full joint posterior over θ and N. This section details the results of some initial tests of an implementation of that algorithm, firstly with a toy analytical model, then with
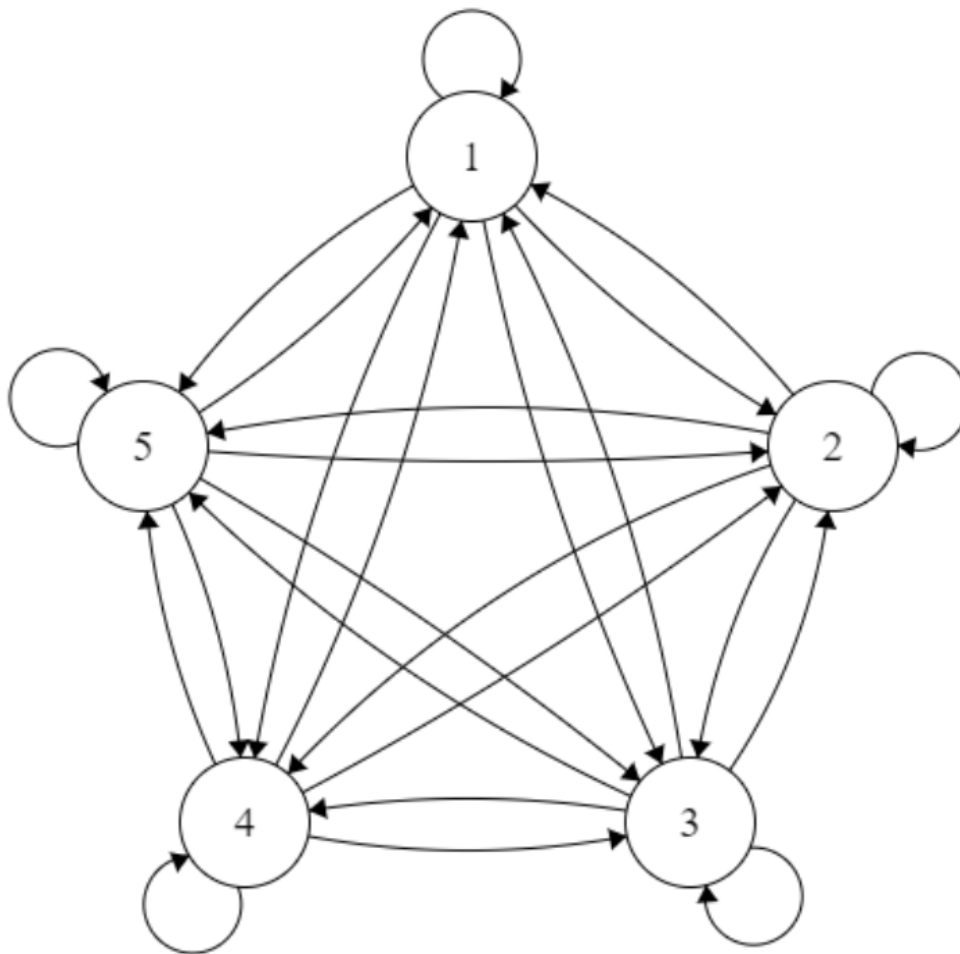
Figure 4.4: A state-transition diagram with five states, where every state has an equal probability of transitioning to any other state (inlcuding itself).

some data.

4.2.1 Analytical Likelihood

One critical assumption made when designing the transdimensional nested sampler was that detailed balance was satisfied by the methods of transitioning up/down in N. This may seem like a wild assumption to make, however by doing so, the rest of the framework of this transdimensional approach could be constructed. It

was established that the process of choosing which N to step to in each iteration of the algorithm should not introduce any bias in N, so any bias that does exist must originate from the procedures by which N is stepped up/down (i.e. copying and deleting particles).

In order to identify any such bias in N, a toy analytical likelihood function was designed. The function gives the joint likelihood over θ and N, and describes a piecewise Gaussian in the parameters, conditional on N, with a mixing amplitude A_N. The function is given as

$$\mathcal{L}(\theta, N) = \frac{A_N N_{\max}(b-a)^{NK}}{\sqrt{|2\pi\Sigma_N|}} \exp\left[-\frac{(\theta - \mu_N)\Sigma_N^{-1}(\theta - \mu_N)}{2}\right], \qquad (4.11)$$

where μ_K is an NK-dimensional vector, Σ_N is an $NK \times NK$-dimensional matrix, and A_N is a scalar defined so that $\sum_N A_N = 1$. In practice, μ_N and Σ_N are arbitrary for the intended purpose of this function, so for simplicity we can set μ_N to be a zero vector, and Σ_N to be the identity matrix.

To understand the design of this function, consider the calculation of the Bayesian evidence:

$$Z = \int \int \mathcal{L}(\theta, N)\pi(\theta, N)\mathrm{d}\theta\mathrm{d}N. \qquad (4.12)$$

Since N is discrete, however, this becomes

$$Z = \sum_N \int \mathcal{L}(\theta, N)\pi(\theta, N)\mathrm{d}\theta. \qquad (4.13)$$

Applying the product rule, we find:

$$Z = \sum_N \int \mathcal{L}(\theta, N)\pi(\theta|N)\pi(N)\mathrm{d}\theta. \tag{4.14}$$

$\mathcal{L}(\theta, N)$ has already been defined, with the other two terms here being our prior on N and the prior on the parameters conditional on N. We can define these priors as:

$$\pi(\theta|N) = \frac{1}{(b-a)^{NK}}, \quad \pi(N) = \frac{1}{N_{\max}}, \tag{4.15}$$

i.e. $\pi(N)$ is a uniform prior on N between 1 and $N_{\max}$, and $\pi(\theta|N)$ is a uniform prior over the parameters between bounds a and b (where $a < b$), given some N. Inserting these into eq. (4.14) and reducing, we're left with:

$$Z = \sum_N \int \frac{A_N}{\sqrt{|2\pi\Sigma_N|}} \exp\left[-\frac{(\theta - \mu_N)\Sigma_N^{-1}(\theta - \mu_N)}{2}\right]\mathrm{d}\theta. \tag{4.16}$$

Because A_N is not dependent on θ, we can take it outside of the integral. The function being integrated is now a normal distribution over θ, so we know that the integral will equal unity, and since we have defined our A_N to also sum to unity, we have

$$Z = \sum_N A_N \int \mathcal{N}(\mu_N, \Sigma_N)\mathrm{d}\theta = \sum_N A_N = 1. \tag{4.17}$$

From Bayes' rule, it then follows that the joint posterior over θ and N is

$$\mathcal{P}(\theta, N) = \mathcal{L}(\theta, N)\pi(\theta, N). \tag{4.18}$$

Again using the product rule, we also know that

$$\mathcal{P}(\theta,N) = \mathcal{P}(\theta|N)\mathcal{P}(N).$$ (4.19)

Rearranging using eqs. (4.18) and (4.19), we obtain

$$\mathcal{P}(N) = \frac{\mathcal{L}(\theta,N)\pi(\theta,N)}{\mathcal{P}(\theta|N)}.$$ (4.20)

We already know the numerator:

$$\mathcal{P}(N) = \frac{A_N \mathcal{N}(\mu_N,\Sigma_N)}{\mathcal{P}(\theta|N)},$$ (4.21)

and by Bayes' theorem, we have:

$$\mathcal{P}(\theta|N) = P(\theta|D,N) = \frac{P(D|\theta,N)P(\theta|N)}{P(D|N)}.$$ (4.22)

Using our shorthand notation, $P(D|\theta,N) = \mathcal{L}(\theta,N)$ and $P(\theta|N) = \pi(\theta|N)$, so we have

$$\begin{aligned}
\mathcal{P}(\theta|N) &= \frac{\mathcal{L}(\theta,N)\pi(\theta|N)}{\int \mathcal{L}(\theta,N)\pi(\theta|N)\mathrm{d}\theta} \\
&= \frac{A_N N_{\max}\mathcal{N}(\mu_N,\Sigma_N)}{A_N N_{\max}\int \mathcal{N}(\mu_N,\Sigma_N)\mathrm{d}\theta} \\
&= \mathcal{N}(\mu_N,\Sigma_N).
\end{aligned}$$ (4.23)

Returning to eq. (4.21), we find

$$\mathcal{P}(N) = \frac{A_N \mathcal{N}(\mu_N,\Sigma_N)}{\mathcal{N}(\mu_N,\Sigma_N)} = A_N,$$ (4.24)

This results tells us that the analytical likelihood function, with the priors defined
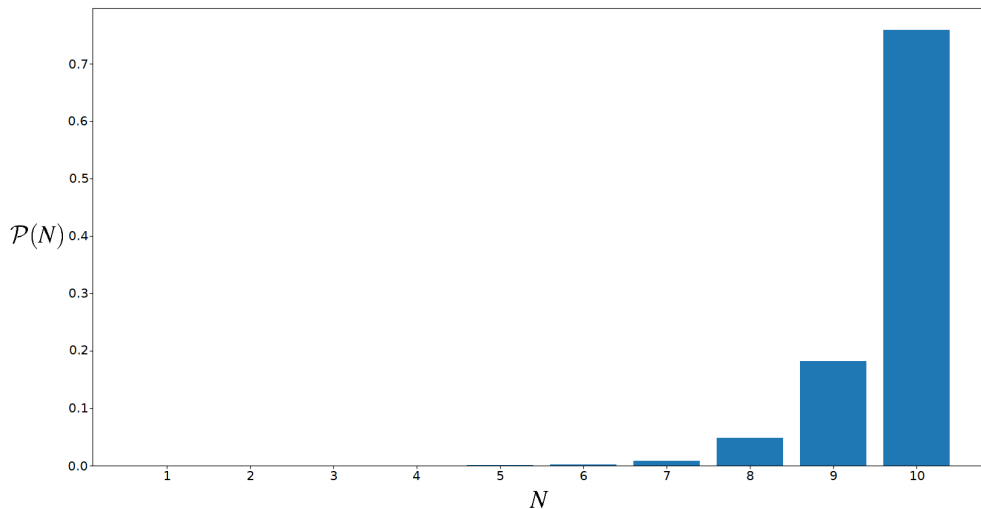
Figure 4.5: Posterior on N where $K = 2$, $N_{\min} = 1$ and $N_{\max} = 10$.

in this way, should produce a posterior on N which is equal to A_N. The trans-dimensional nested sampler already provides us with a full joint posterior on θ and N, so by marginalising over θ, we can easily acquire $\mathcal{P}(N)$. If we initially set $A_N = 1/N_{\max} \; \forall N$, we should expect a uniform posterior on N as a result of running the sampler. If this is not the case, there must be a bias in N. Figure 4.5 shows $\mathcal{P}(N)$ acquired from running the sampler with this analytical likelihood function. Clearly, the sampler in its current form exhibits a strong bias towards high N.

In order to definitively establish the source of this bias, a separate test can be run. By performing separate runs of the algorithm where $N_{\min} = N_{\max} = 1, 2, \cdots, 10$, we remove the possibility of performing steps up/down in N in any of the ten runs. The posterior samples can then be 'merged', weighted by the evidence found from each run. This was done using the nested sampling visualisation tool `anesthetic` (Handley 2019). This merged posterior is shown in fig. 4.6. In this case, the posterior is much more uniform than in the case of a single run. This implies that the strong bias seen towards high N must originate from the proced-
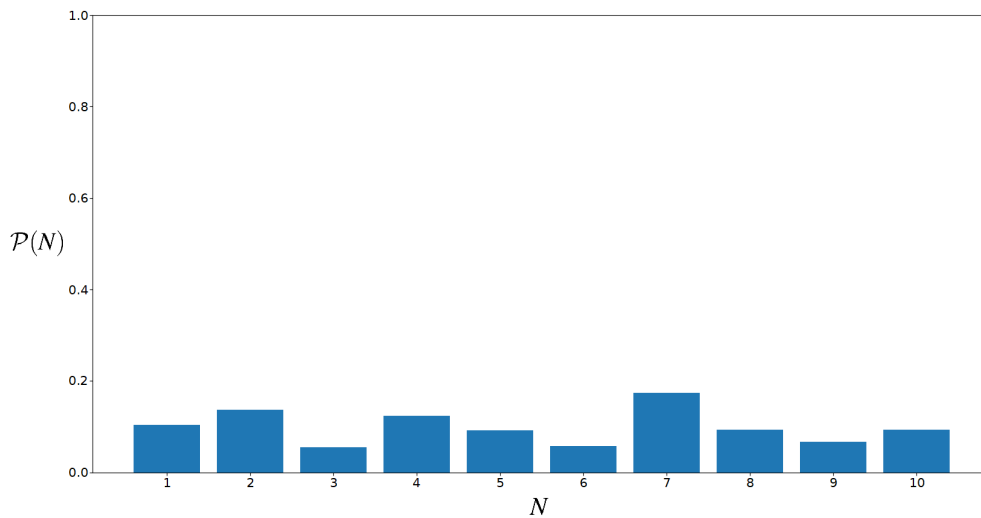
Figure 4.6: Posterior on N produced from ten separate nested sampling runs with $K = 2$ and $N_{\mathrm{min}} = N_{\mathrm{max}} = N$

ures for stepping up/down in N. More specifically, because on average a step up in N is equally likely to be proposed as a step down in N, it must be the case that a proposed step up is, on average, more likely to be accepted than a proposed step down. Since this approach does not make use of acceptance probabilities, this must be because the process of stepping up is more likely to propose a point which lies within the iso-likelihood contour than the process of stepping down.

Interestingly, it seems that given this bias, the posterior on N becomes log-linear. The strength of the bias also seems to scale with K. Using linear regression, a line can be fit to $\log \mathcal{P}(N)$, which indicates the how strong the bias is. As K increases, the line becomes steeper. This is all summarised in fig. 4.7. By finding how this gradient changes with K, we can begin to quantify the bias. While the bias is not yet completely understood, because the bias gets stronger as the prior volume increases, it seems that the cause of the bias may be some sort of volume effect. Identifying exactly how the bias changes with the prior volume (identifying

some functional form for the relationship) would allow us to better understand why the bias occurs, as well as possibly developing a method to counter it.
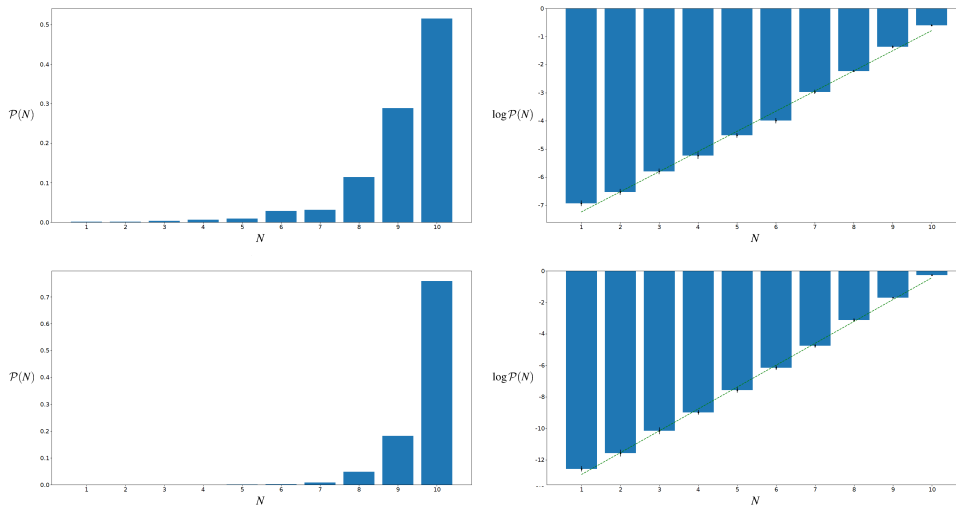
Figure 4.7: Top left: Posterior on N when $K = 1$. Top right: Log-posterior on N when $K = 1$, with line of best fit. Bottom left: Posterior on N when $K = 2$. Bottom right: Log-posterior on N when $K = 2$, with line of best fit.

4.2.2 Testing with Data

Regardless of the bias uncovered using the analytical toy model, it is worth testing how well this transdimensional approach deals with data. Perhaps the simplest test would be to produce some data which can be described by the summation of N 'events', each of which is parameterised by the same basis function taking K parameters. For the purpose of these tests, a Gaussian basis function was used, featuring three parameters describing the height h, width w and central position p of the curve. Depending on the test, each of these parameters can be allowed to freely vary, or one or more of them could be fixed (effectively reducing K).

As a starting point, a curve was generated by first producing a constant back-

ground with a height of zero, then adding three unit-height, unit-width Gaussians, spaced apart as to not overlap with one another. The curve was then sampled uniformly and a small amount of Gaussian noise was added to the sampled data points. The transdimensional sampler was then run on the data, with the model heights and widths fixed at 1, only allowing the position to vary ($K = 1$). A Gaussian (log-)likelihood was used:

$$\log \mathcal{L} = -\frac{1}{2} \sum_i \log(2\pi\sigma_i) - \frac{1}{2}\chi^2, \tag{4.25}$$

where χ^2 quantifies the misfit between the observed data O and the prediction from the model E.

$$\chi^2 = \sum_i \left(\frac{O_i - E_i}{\sigma_i} \right)^2, \tag{4.26}$$

A uniform prior was assumed on $N \sim U(1,10)$, as well as the position parameter $p \sim U(0,25)$. From the sampler output, the maximum a posteriori (MAP) estimate parameter values were found and plotted against the data in fig. 4.8. Parameter estimation using the transdimensional sampler has successfully allowed us to reconstruct the true model for this data set. $\mathcal{P}(N)$ shows a sharp peak at $N = 3$, indicating a strong preference for 3-Gaussian models.
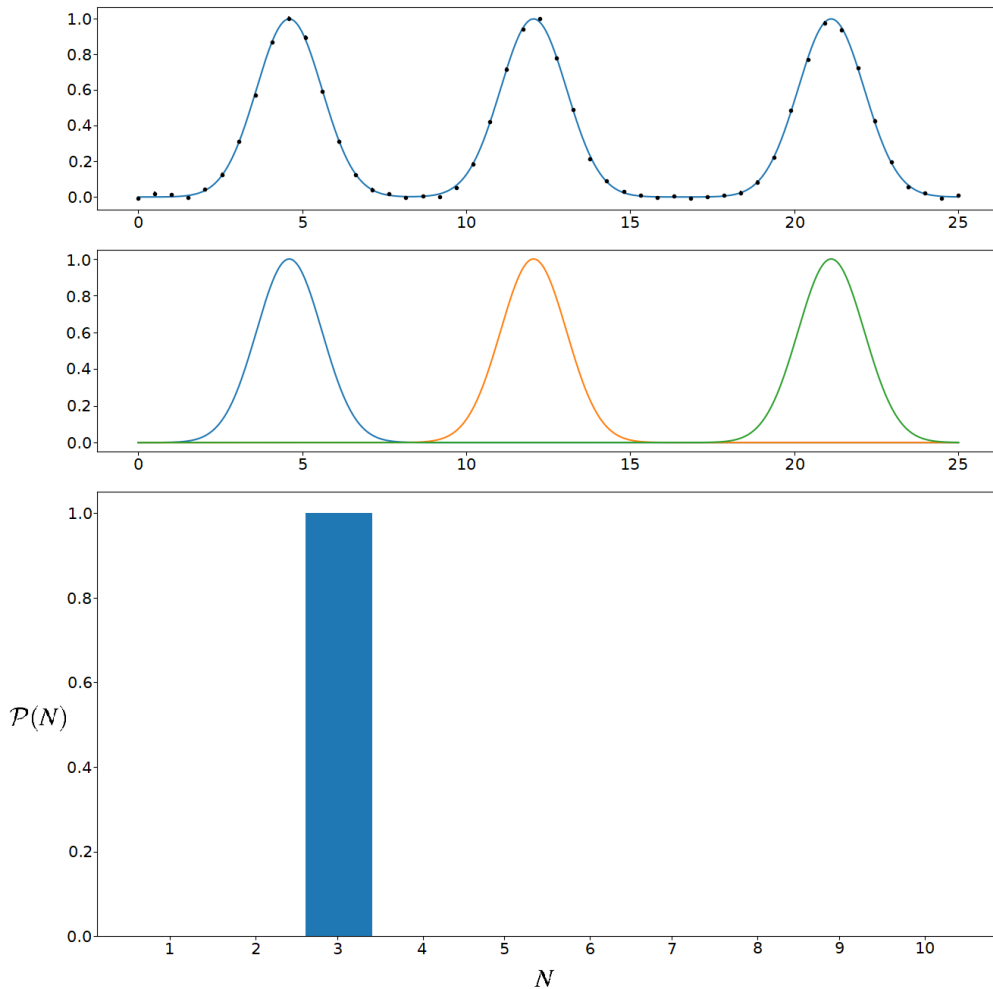
Figure 4.8: Top: Full MAP reconstruction of the Gaussian model. Middle: MAP curve deconstructed to show the three constituent Gaussian basis functions. Bottom: Posterior on N.

Interestingly, if we then also allow the width to vary such that $w \sim U(0.1, 5)$ and now $K = 2$, we find that the results are virtually identical. This is not the case, however, if we keep the width fixed, but allow the height to vary. If we assume a uniform prior on $h \sim U(0.05, 2)$, the MAP estimate parameters suddenly describe a model with four Gaussians (fig. 4.9). The posterior on N also shows a much wider spread in this case. If all three parameters are allowed to vary freely,

the effect is very similar. This is easily explained considering the bias towards higher N found using the analytical likelihood function. The sampler naturally works towards the 'true' posterior, but the in-built bias directs the sampler towards a model with more particles. This means that the sampler incorrectly favours models which stack multiple Gaussians on top of one another. This doesn't happen if the height is fixed, because adding in additional Gaussians in this case would require distorting the curve significantly, impairing the quality of the fit enough to overcome the bias.

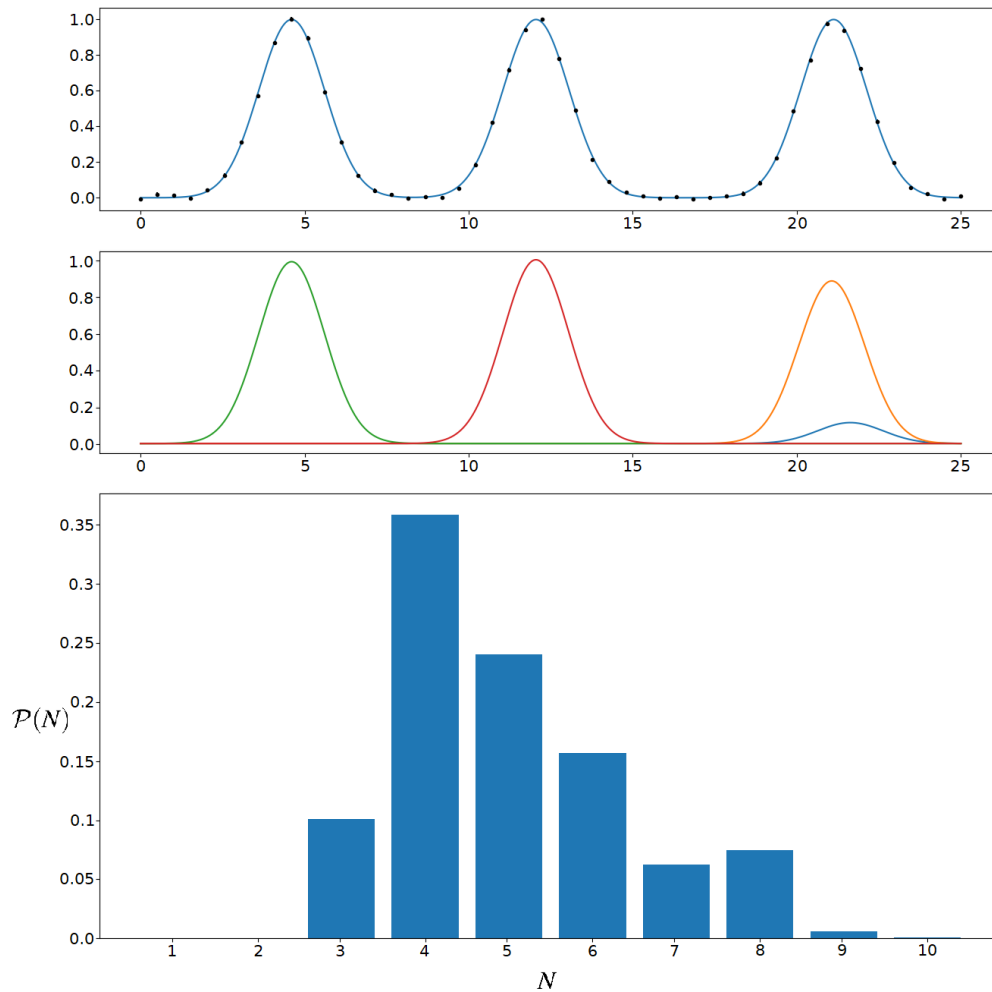**Figure 4.9:** Top: Full MAP reconstruction of the Gaussian model. Middle: MAP curve deconstructed to show the four constituent Gaussian basis functions. Bottom: Posterior on N.

4.3 Conclusions

This chapter has explored a framework for performing transdimensional nested sampling. While this framework has been demonstrated to function, the described implementation features a strong bias towards higher values of N, which would

need to be addressed in future. The bias is not completely understood, but seems to be related to the prior volume. Identifying a functional form to describe this relationship would be invaluable to identifying exactly how this bias arises. Additionally it may be possible to then alter the algorithm to remove the bias. This could simply involve finding alternative methods for stepping up/down in N, or even implementing a weighting scheme on the choice of N when transitioning, or adding an equivalent set of acceptance probabilities dependent on N. In any case, further testing of this implementation, as well as further analysis of the framework as a whole would be enlightening.

Looking beyond the issue of bias, the approach could also be extended in a number of ways. There are a number of features implemented by other popular nested sampling algorithms which would likely improve this approach. For example, using clustering to handle posteriors with multiple modes would likely be a welcome addition. Further than this, it is possible where there are similar problems that this approach cannot yet handle, but may work with future extensions. It is fairly straightforward to consider models which have additional auxiliary parameters which are necessary to describe the data, but aren't included in the particles. A simple example would be adding a parameter to the Gaussian data example from section 4.2 which adds a base height for all of the Gaussians. This parameter only needs to be included once in the model, not repeatedly for each Gaussian. In principal, this extension shouldn't cause any problems. A more complicated extension would be the case where there are multiple types of particle, each with a distinct N and K. In the Gaussian example, this could be adding in a different type of basis function with a different form. The different Ns would then be free to independently vary as the algorithm progressed. The

current code is insufficient to deal with this type of problem, however this kind of extension is certainly not beyond the scope of reason.

Chapter 5

Demonstrating Achromaticity

The discovery of SAV (V2017A) opened a new window on the study of AGN. SAV is very likely due to gravitational milli-lensing by mass condensates of $10^3 - 10^6 M_\odot$ and, if so, it provides a uniquely powerful approach to exploring this difficult to observe mass range. Such a lens is expected from the ΛCDM model, although at time of writing none have been found by subtle deviations in images of sources strongly lensed by galaxies, as far as I am aware. Even finding one definite gravitational lens in this range would have huge implications.

Originally, it was thought that the unusual variations seen in one of these lensing events could be explained by Extreme Scattering Events (ESEs), however this interpretation was critically examined in V2017B and ultimately rejected. One reason for this rejection is that ESEs exhibit a strong frequency dependence, which is not seen in the achromatic events found in the OVRO 40m data. Achromaticity is a defining feature of SAVs, and providing evidence that these events are indeed achromatic would greatly strengthen the argument for SAVs. Alternatively, if evidence can be found which shows that these events *are* chromatic, this would

challenge the gravitational lensing hypothesis and could imply that the events are due to some intrinsic variation in the blazar jets; a scenario which carries its own wealth of astrophysical potential. It is worth noting, however, that even if the SAV hypothesis is correct, we might still expect some chromaticity, since the radio emissions at different frequencies may come from different regions of the jet.

In this chapter I introduce an observing campaign of the source J1415+1320 using e-MERLIN, which was established to gather evidence for achromaticity of the lensed source components. I describe the details of the observations, as well as the motivations behind the campaign. I then finish the chapter by giving an overview and discussion of the data produced so far.

5.1 Monitoring J1415+1320 with e-MERLIN

Ten SAV events were originally identified across seven sources, with J1415+1320 being the exemplar SAV source, having demonstrated multiple SAV events. Because SAVs are expected to repeat, the history of SAV activity in this source makes it the best candidate to observe for future events, which could provide an opportunity to investigate their frequency (in-)dependence. In fact, an as-yet unpublished paper investigating millilensing of this source has identified a 989-day periodicity, which, if real, would allow us to make a prediction that the next window for possible SAV activity begins in August 2022 (personal communication, Readhead 2021). Ideally, it would be wise to monitor each of the seven confirmed SAV sources, however this would dramatically increase the overheads involved in such a campaign, and as will become clear, monitoring J1415+1320 entails certain benefits which would not occur for the other sources.
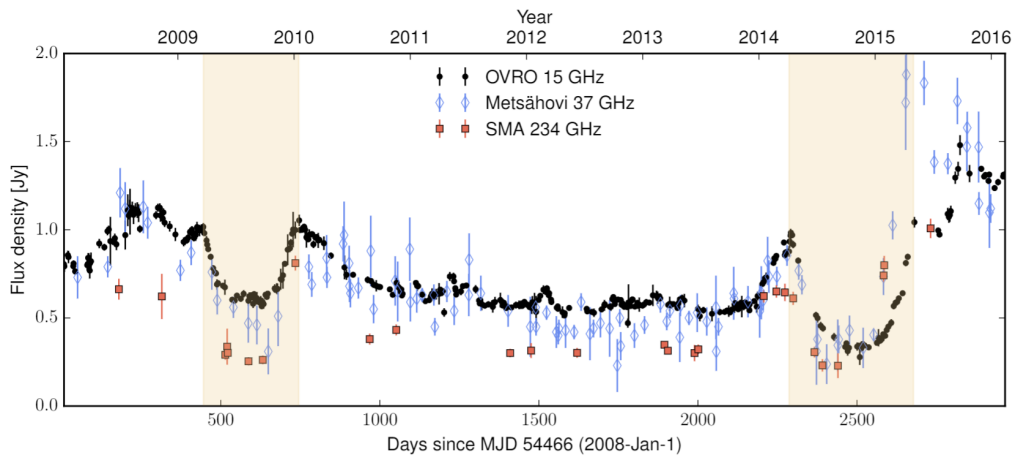
Figure 5.1: Multi-frequency light curves of J1415+1320 showing SAV. Figure from V2017A.

While there has previously been good evidence to suggest that SAV events are achromatic, based on observations of J1415+1320 from the Submillimeter Array (SMA) and the Metsähovi 14m radio telescope, these data sets suffer from low cadence and large error bars, as seen in fig. 5.1. Lower frequencies are also more prone to plasma lensing effects, so it is conceivable that the events are actually frequency-dependent, but the effects have gone unnoticed at the higher frequencies the events have already been observed at. We therefore desire to design a lower frequency monitoring programme which will produce data of comparable quality to that of the OVRO 40m data set.

The enhanced Multi-Element Radio Linked Interferometer Network (e-MERLIN) is an array of seven radio telescopes all situated within the UK, operated by the Jodrell Bank Centre for Astrophysics and the University of Manchester for the Science and Technology Facilities Council. The array offers three observing bands: L-band ($1.23 - 1.74$GHz), C-band ($4.3 - 7.5$GHz) and K-band ($19 - 25$GHz). Comparing this to the OVRO 40m telescope, which has a central frequency of

15GHz, Metsähovi at 37GHz and the SMA at 234GHz, we see that an obsering campaign carried out in both L- and C-band would allow us to explore lower frequencies than have previously been investigated. Furthermore, J1415+1320 is fortuitously located very close to the positions of the standard 3C286/OQ208 calibrator combination used by e-MERLIN, as shown in table 5.1.

Table 5.1: ICRS coordinates of sources

Source	Right Ascension	Declination
J1415+1320	14h 15m 58.818s	$+13° 20' 23.713$"
3C286	13h 31m 08.288s	$+30° 30' 32.962$"
OQ208	14h 07m 00.393s	$+28° 27' 14.695$"

Test data of J1415+1320 taken daily over ten days by the e-MERLIN team in March 2018 at C-band is shown in fig. 5.2. The scatter of flux density when all telescopes are present (which is the case for the majority of the days) is less than 1% rms around the mean of 721 mJy from a thirty minute observation. The four data points with the largest error bars are all from observations when one or more telescopes in the array were missing. This is to be compared with the expected flux variations of $> 40\%$ during an SAV event, assuming that the events are truly achromatic. The source is so bright that we calculate that only ten minutes of observing time is required to make a measurement of sufficient accuracy ($\sim 2\%$). Discussions with the scheduling team at Jodrell Bank indicated that shortening the request any more than this would run into practical issues with time lost to slewing and settling. The short observing times and close proximity of the calibrators mean that these observations could be scheduled to take place approximately once every two days during the array's normal calibration cycles which are carried out during observations for other projects. It was decided that this approach would

be preferable to requesting a longer observation every five days, for a number of reasons:

- Practically speaking this would have the lowest impact on the overall scheduling of the array.

- Producing light curves sampled at both L and C bands every five days would require observations at both frequencies for both J1415+1320 and for the calibrators sources. Often the calibration observations for one of these bands would be required solely for this observation, leading to a significant overhead in total telescope time, without increasing the number of useful on-source measurements. Consulting with e-MERLIN staff, this overhead was estimated as being an hour of additional time for bandpass and absolute flux calibrations, together with the band switching. Therefore the time requested for a programme with a regular five day cadence would give an average request of 16 minutes per day, compared with the five minutes per day requested for the preferred approach.

- The test data showed that on days where one or more of the telescopes were not present, the rms of the flux measurement increased dramatically and there were indications of systematic errors. Should this happen to even two successive observations in a five day cadence, this would lead to significant gaps in the light curve, complicating the analysis and clouding the interpretation. The test observations indicated that the failures of individual telescopes are rapidly corrected, so the proposed higher-cadence observations would be much less prone to that possibility.
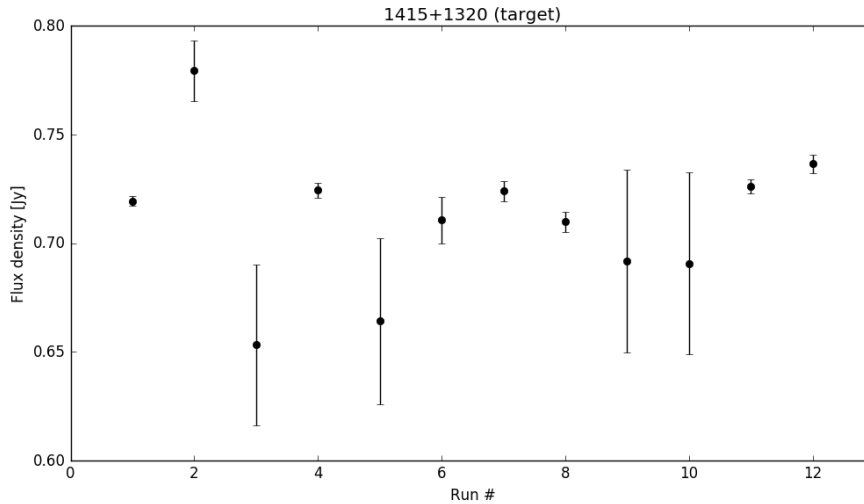
Figure 5.2: Test data from daily thirty minute observations of J1415+1320 with e-MERLIN.

Based on these test observations, a Director's Discretionary Time Proposal was submitted and accepted, allowing observations for the monitoring programme to begin in January 2019, followed by a full observing proposal which was accepted for the following e-MERLIN observing cycle, with observations continuing regularly to the present day (barring a long gap due to a pause in obserations during the COVID-19 pandemic). The programme is planned to continue indefinitely until another SAV event is observed.

5.2 Data and Discussion

At time of writing, 323 observations have been processed between 25th January 2019 and 28th March 2021. Out of these 323 observations, 53 failed, meaning around 83.6% of attempted observations gave some usable data. Of the remaining 270 observations, 178 were carried out in C-band with an average cadence of 3.78

days and 92 were carried out in L-band with an average cadence of 7.21 days. In both cases the gap in observations due to the COVID-19 pandemic was omitted in cadence calculations. Figure 5.3 shows a light curve produced from these data, with the OVRO 40m light curve of the source during the same period for reference.
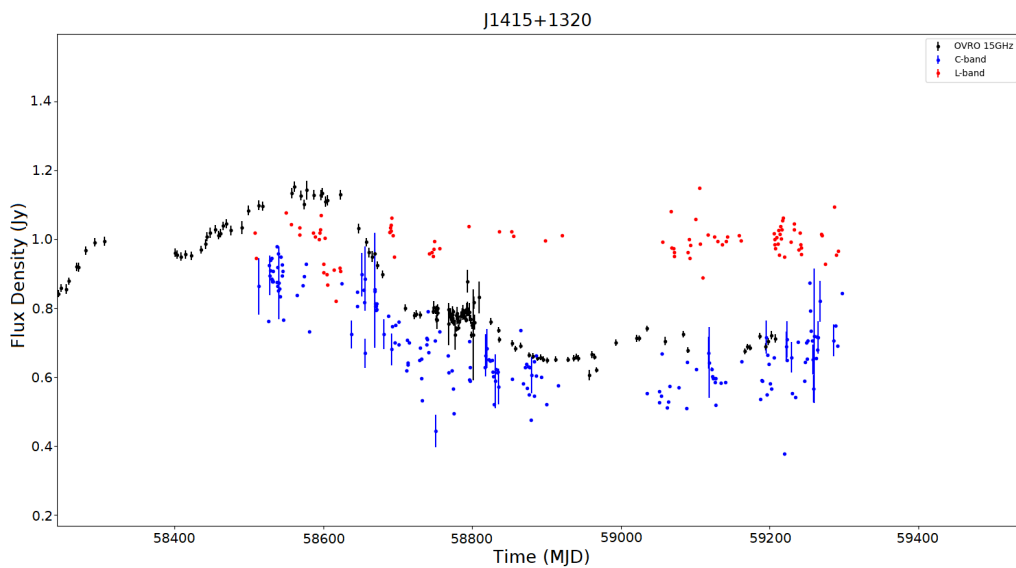
Figure 5.3: e-MERLIN C- and L-band light curves of J1415+1320 plotted against the 15GHz OVRO 40m light curve of the same source.

These light curves should be viewed as preliminary, as they were produced using images generated automatically from e-MERLIN data reduction pipeline. Many of the C-band observations have fairly large error bars, which in many cases may be due to missing antennas. Although the error bars are large on these data points, they are consistent with the surrounding points and so there is no indication of additional systematic errors. The points can therefore be included in any future analysis, but will be significantly down-weighted. Additionally, there is a large amount of inter-day scatter from both bands (even between data points with relatively small error bars) which is currently unexplained. A number of initial checks

have been suggested to identify the cause of this scatter, including examining in detail the data for 1407+2827, the bandpass calibrator for these observations, in addition to calculating the closure phases/ampltitudes to identify possible causes of increased inter-day variability.
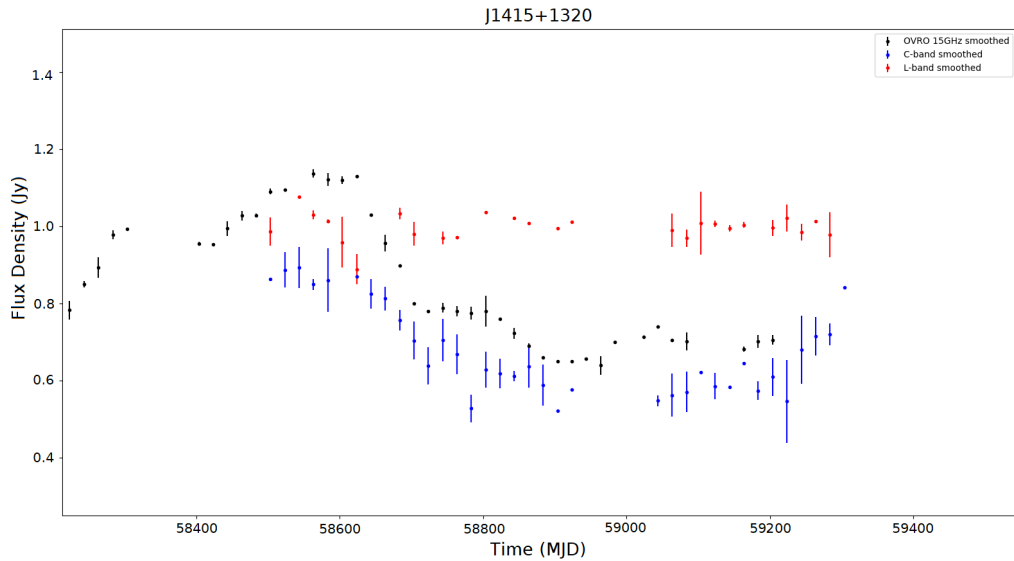

Figure 5.4: Recreation of fig. 5.3, where the C- and L-band e-MERLIN light curves have been smoothed, along with the OVRO 15GHz data. Note that error bars are included for the entire e-MERLIN data set, but some are invisibly small in this plot. More OVRO observations have been carried out past the last observations shown at 59200 MJD (and are expected to continue into the future), although at time of writing these observations have yet not been processed to provide flux estimates.

In an attempt to overcome the issues of the scatter and large error bars, at least until these issues can be investigated further, the light curve was smoothed by binning the data points into fixed 20-day windows and computing the weighted average flux density (where the ith data point is weighted by $w_i = 1/\sigma_i^2$). The result of this smoothing is shown in fig. 5.4. Interestingly, the C-band curve seems to follow the variations of the 15GHz OVRO data fairly closely, whereas the L-band data is much flatter. To investigate this, we find the ratio $R = S_O/S_C$ between

the OVRO and C-band flux densities for every point along the smoothed curves. The result is shown in fig. 5.5, where the error bars for each point are calculated using

$$\sigma_R = R\sqrt{\left(\frac{\sigma_C}{S_C}\right)^2 + \left(\frac{\sigma_O}{S_O}\right)^2}. \tag{5.1}$$

Because the OVRO 15GHz light curve is brighter at all points, the values for R are all greater than unity, scattered around a mean value of $\bar{R} = 1.218$. The quantile-quantile (Q-Q) plot shown in fig. 5.6 indicates that these ratio values are approximately normally distributed around the mean, which tells us that the two light curves do co-vary quite closely, but not exactly. There is one outlying data point in this plot, which corresponds to the point in the smoothed curve at MJD 58783 that is somewhat discrepant from the rest of the data. Futhermore, the scatter plot in fig. 5.7 shows a strong, positive correlation, with the Pearson correlation coefficient having a value of $\rho = 0.941$. We have strong evidence, therefore that these two light curves are correlated with one another.

Figure 5.5: Ratios of OVRO and C-band flux densities from the smoothed light curves. The red dashed line indicates the mean value of R.

Figure 5.6: Q-Q plot showing observed ratios between OVRO and C-band data against predictions from a normal distribution with the same mean and standard deviation.

Figure 5.7: Scatter plot of flux densities from the smoothed OVRO light curve against those from the smoothed C-band light curve. The black dashed line is the best fit to the data found via linear regression.

Repeating the same steps to instead compare the OVRO data with the L-band data, we find that the new ratios are not as well-described by a normal distribution, as demonstrated in fig. 5.9. Because the L-band data is much less variable, we see in fig. 5.8 that the ratios are initially greater than unity, but then drop below as the OVRO curve falls to a lower flux density. The scatter in fig. 5.10 shows very little evidence of correlation, with $\rho = -0.273$.
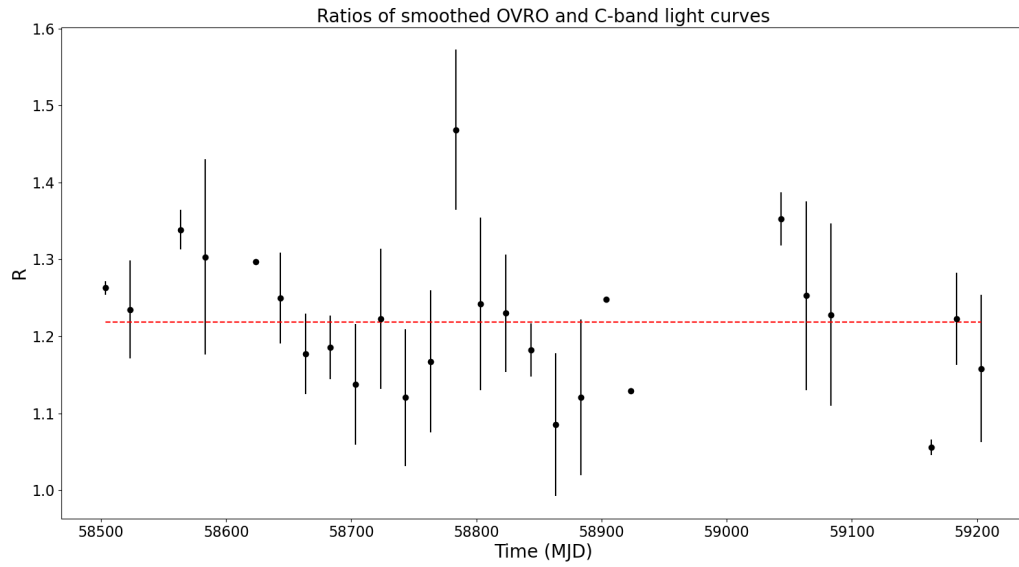
Figure 5.8: Ratios of OVRO and L-band flux densities from the smoothed light curves. The red dashed line indicates the mean value of R.
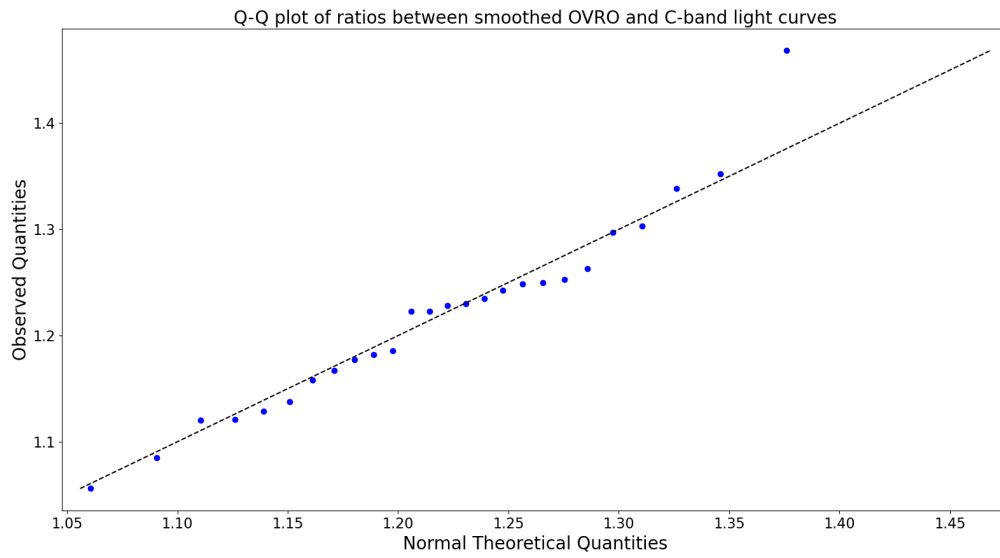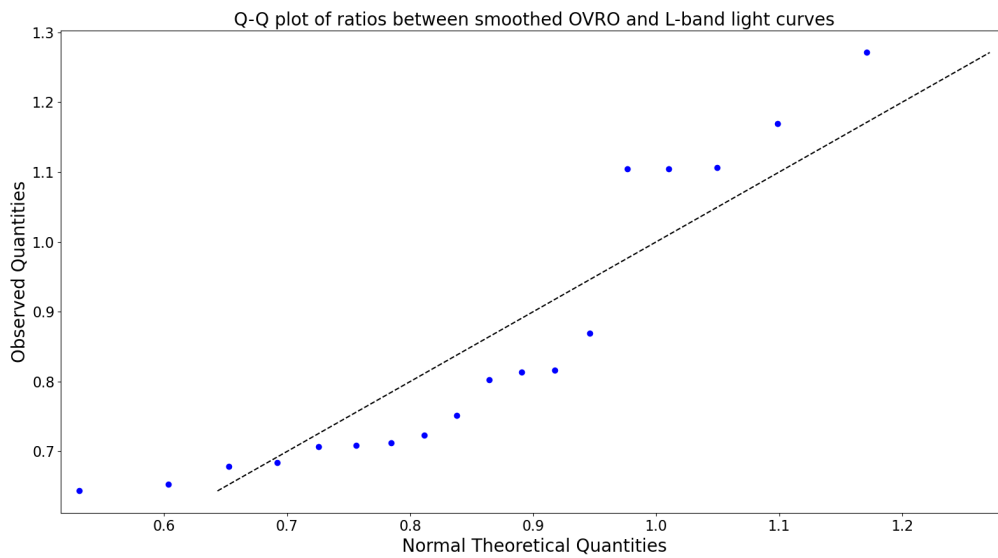

Figure 5.9: Q-Q plot showing observed ratios between OVRO and L-band data against predictions from a normal distribution with the same mean and standard deviation.
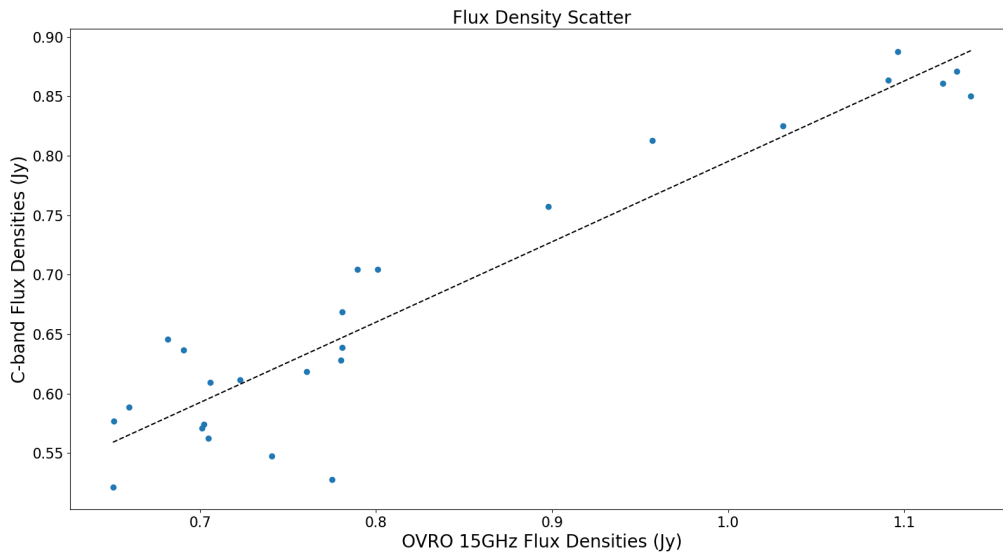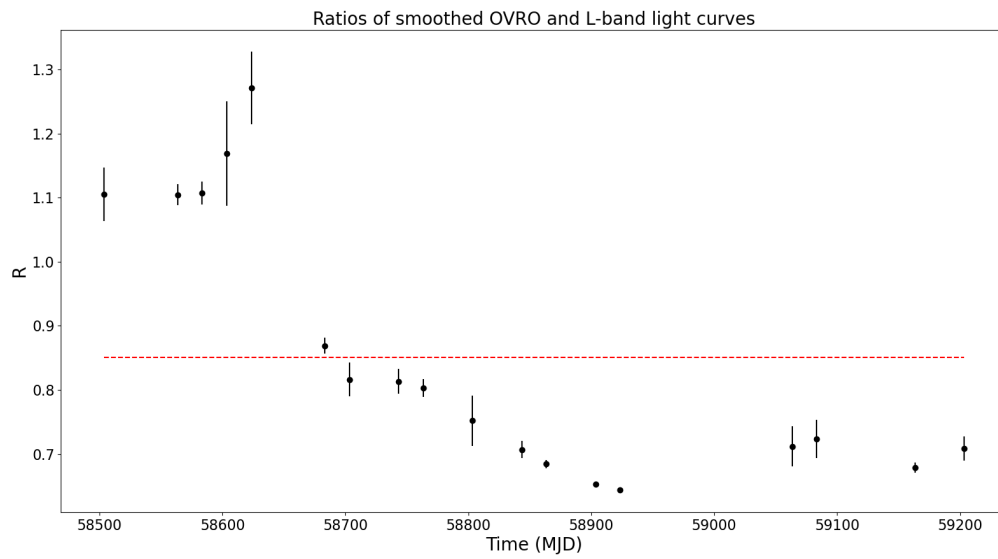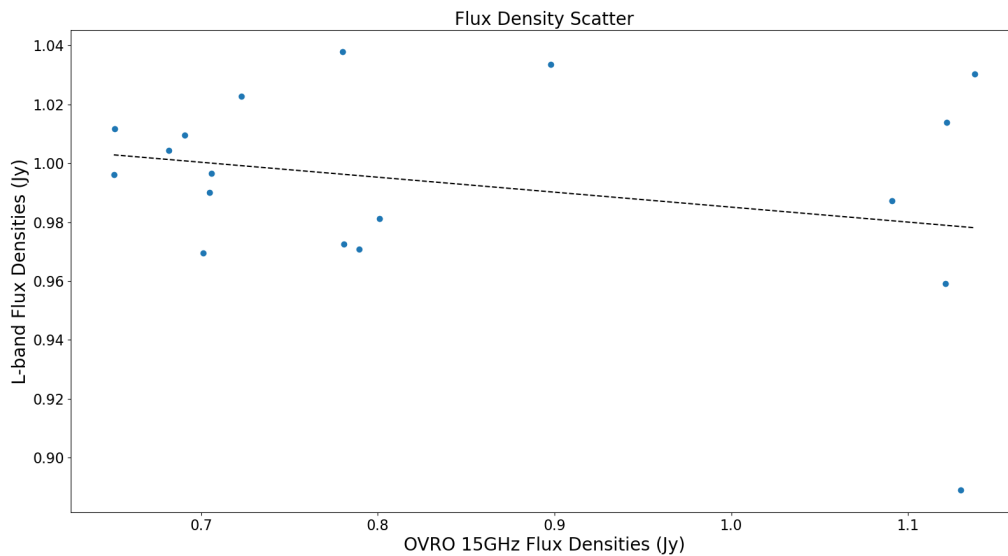
Figure 5.10: Scatter plot of flux densities from the smoothed OVRO light curve against those from the smoothed L-band light curve. The black dashed line is the best fit to the data found via linear regression.

We observe that the flux density rises between the 5GHz C-band observations and the 15GHz OVRO observations, indicating an inverted spectrum. Since there is very strong evidence that, in radio, the emission mechanism for blazars is synchrotron, this implies that at least part of the emitting region is optically thick. This would suggest, however, that the emission would be even fainter at L-band compared to C-band, whereas the flux observed is higher. A plausible explanation for this is that there are multiple sites of synchrotron emission. Firstly, we have a diffuse, optically thin region, which, in order to be bright enough to observe, must be extended and therefore will not be variable. We also have source components consisting of relativistic plasma which travel away from the central engine down the jet. These source components will have some intrinsic variability and will move past the putative gravitational lens, so their emission will show SAV type structures. At C-band we would therefore observe some constant emission

from the diffuse material, in addition to some variable emission from the source components. At lower L-band frequencies, however, the source components are optically thick and so the dominant emission comes from the diffuse region, with very little contribution from the source components.

5.3 Conclusion and Further Work

While this initial inspection of the data has unveiled some possible problems with high inter-day scatter which need to be investigated, applying a weighted smoothing produces data which is stable on timescales short enough to monitor the intrinsic variability of the source. The C-band data mimics the variations seen in the OVRO 40m data well and the observation cadence has been regular enough to expect to capture fairly short term variations of the source, meaning the data should hopefully provide some quality analysis once a new event is uncovered. Overall the data set certainly looks promising, with the monitoring programme continuing indefinitely, ready to capture the next SAV event. It is not a question of *if* another event will occur, but a question of *when*. SAV sources are repeaters and so we simply have to keep observing and wait for an event to become apparent. When attempting to determine whether future events show evidence of achromaticity, having this second data set (in addition to the OVRO data) would allow for an excellent level of analysis, given the cadence of the observations and the general quality of the data. Variations in the OVRO data are mimicked closely by the C-band e-MERLIN data, so if SAV events are indeed achromatic, we should expect to see strong evidence in the comparison of the two data sets. Conversely, discrepancies between the two data sets around the next event could indicate that

the assumption of achromaticity is incorrect. Either possibility would have strong implications for the validity of the SAV hypothesis and time will tell exactly what those implications will be.

Chapter 6

Summary and Future Work

Symmetric Achromatic Variability remains an open and active area of research, still in its infancy. As our knowledge of this phenomenon grows, we will gain insight into exciting new frontiers of astrophysics. Because SAV activity is rare, with only seven blazar sources originally suggested as candidates, it remains a priority to keep searching for more sources which exhibit this unique type of variability. It would therefore be extremely useful to have a reliable, statistically rigorous method for examining blazar light curves for signatures of SAV. In chapter 3 I introduced a framework designed to do just that. I described a model which I have developed to describe the light curve of a source exhibiting SAV, and provided justified reasoning for each of the model's components. I then detailed a fitting procedure which uses nested sampling to fit that model to a given blazar light curve. I showed how this fitting procedure can be utilised in a novel approach for generating artificial blazar light curves, based off real light curves from the OVRO data set. Finally, I showed the results of running the fitting procedure with both real and artificial light curves. My analysis showed that the procedure is

capable of achieving very good fits to the observed data using three different SAV models. However, in both test examples the procedure was unable to recover the *a priori* known SAV event. In its current form, the fitting procedure seems to have a tendency to fit broad magnification curves to long-term variations in the data. The procedure was shown, however, to succesfully identify an SAV event placed in a simplified light curve (one which is not representative of those seen in the OVRO catalogue). In order to gain a better understanding, it would be beneficial to run this kind of analysis over a much wider range of (artificial) light curves, which would allow us to quantify our belief that the events found by the fitting procedure are valid. This was not possible in the given time frame, but it is a priority for future work. Additionally, being able to quantify the performance of the procedure would allow us to directly measure the impact of modifying the SAV model. This would make it much easier to refine and tune the model to achieve demonstrably better results.

Chapter 4 introduced a novel transdimensional alternative to traditional nested sampling. Contrary to existing nested sampling methods, a transdimensional approach would allow for the number of parameters in the model to vary as a free parameters. This is done by defining a set of K parameters called a 'particle', where the model can consist of N particles (each containing their own K parameters), where N is included as a model parameter and K is fixed. I describe a process by which N is able to vary, allowing the algorithm to explore a space where the dimensionality is not fixed. This includes the introduction of a novel concept: the 'flattened' particle space. The approach proved to be functional, successfully managing to fit a toy model to some data, however it was shown analytically to produce a strong bias towards larger N. It is likely that this bias comes from the

mechanisms through which live points are evolved to increase and decrease N, but the exact cause is currently unknown. Finding the source of this bias and eliminating it would be the top priority for future development of this method. A good place to start would be to quantify the bias, investigating how it varies with the model parameters and any hyperparameters. This would hopefully give some insight into why it occurs, and possibly hint at how to counteract it. This could be by developing alternative mechanisms to step up/down in N, or perhaps by adding in some acceptance probability conditional on N. I, along with my collaborators, intend to publish the work we have done on this transdimensional approach in the near future, which will hopefully inspire other researchers with different perspectives to help build upon what we have produced so far.

Finally, chapter 5 details an ongoing campaign to monitor J1415+1320 at L- and C-band with e-MERLIN. It is hoped that the high-cadence light curves produced from these observations will cover the period of the next expected SAV event. Thanks to the lower frequency of these observations compared to the 15GHz OVRO data, this will hopefully provide strong evidence as to whether these events are truly achromatic, which would have significant implications for the SAV hypothesis either way. I presented the L- and C-band light curves produced with the currently available data, which look very promising apart from a few large error bars and a large amount of inter-day scatter which is currently unexplained. Smoothing these light curves allows us to compare them directly with the OVRO light curve. Preliminary analysis indicates a strong correlation between the C-band and OVRO light curves, whereas the L-band light curve is much less variable and is not correlated. With the C-band data following the variations in the OVRO data so closely, we can be confident the next expected SAV

event should show up clearly in both light curves, assuming the SAV hypothesis is correct. The campaign is expected to continue indefinitely, at least until another event is observed in the OVRO light curve.

Bibliography

Abdo, A. A. et al. (2009*a*), 'Bright Active Galactic Nuclei Source List from the First Three Months of the Fermi Large Area Telescope All-Sky Survey', ApJ **700**(1), 597–622.

Abdo, A. A. et al. (2009*b*), 'Fermi/Large Area Telescope Bright Gamma-Ray Source List', ApJS **183**(1), 46–66.

Abdo, A. A. et al. (2010*a*), 'Fermi Large Area Telescope First Source Catalog', ApJS **188**(2), 405–436.

Abdo, A. A. et al. (2010*b*), 'Gamma-ray Light Curves and Variability of Bright Fermi-detected Blazars', ApJ **722**(1), 520–542.

Abdo, A. A. et al. (2010*c*), 'The First Catalog of Active Galactic Nuclei Detected by the Fermi Large Area Telescope', ApJ **715**(1), 429–457.

Ackermann, M. et al. (2011), 'The Second Catalog of Active Galactic Nuclei Detected by the Fermi Large Area Telescope', ApJ **743**(2), 171.

Alcock, C. et al. (1993), 'Possible gravitational microlensing of a star in the Large Magellanic Cloud', Nature **365**(6447), 621–623.

Antonucci, R. (1993), 'Unified models for active galactic nuclei and quasars.', ARA&A **31**, 473–521.

Beckmann, V. & Shrader, C. (2012), The AGN phenomenon: open issues, *in* 'Proceedings of "An INTEGRAL view of the high-energy sky (the first 10 years)" - 9th INTEGRAL Workshop and celebration of the 10th anniversary of the launch (INTEGRAL 2012). 15-19 October 2012. Bibliotheque Nationale de France', p. 69.

Bennett, A. S. (1962), 'The revised 3C catalogue of radio sources.', MmRAS **68**, 163.

Brewer, B. J. (2014), 'Inference for Trans-dimensional Bayesian Models with Diffusive Nested Sampling', *arXiv e-prints* p. arXiv:1411.3921.

Brewer, B. J. & Foreman-Mackey, D. (2016), 'DNest4: Diffusive Nested Sampling in C++ and Python', *arXiv e-prints* p. arXiv:1606.03757.

Brewer, B. J., Pártay, L. B. & Csányi, G. (2009), 'Diffusive Nested Sampling', *arXiv e-prints* p. arXiv:0912.2380.

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

Chib, S. & Jeliazkov, I. (2001), 'Marginal likelihood from the metropolis–hastings output', *Journal of the American Statistical Association* **96**(453), 270–281.
  URL: *https://doi.org/10.1198/016214501750332848*

Chua, A. J. K., Hee, S., Handley, W. J., Higson, E., Moore, C. J., Gair, J. R., Hobson, M. P. & Lasenby, A. N. (2018), 'Towards a framework for testing general

relativity with extreme-mass-ratio-inspiral observations', MNRAS **478**(1), 28–40.

Cox, R. T. (1946), 'Probability, frequency and reasonable expectation', *American Journal of Physics* **14**(1), 1–13.
URL: *https://doi.org/10.1119/1.1990764*

Dermer, C. D., Schlickeiser, R. & Mastichiadis, A. (1992), 'High-energy gamma radiation from extragalactic radio sources.', A&A **256**, L27–L30.

Edge, D. O. et al. (1959), 'A survey of radio sources at a frequency of 159 Mc/s.', MmRAS **68**, 37–60.

Fanaroff, B. L. & Riley, J. M. (1974), 'The morphology of extragalactic radio sources of high and low luminosity', MNRAS **167**, 31P–36P.

Feroz, F. (2008), Bayesian Methods for Astrophysics and Particle Physics, PhD thesis, University of Cambridge.

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**(4), 1601–1614.
URL: *https://doi.org/10.1111/j.1365-2966.2009.14548.x*

Fiedler, R. L. et al. (1987), 'Extreme scattering events caused by compact structures in the interstellar medium', Nature **326**(6114), 675–678.

Goldreich, P. & Sridhar, S. (2006), 'Folded Fields as the Source of Extreme Radio-Wave Scattering in the Galactic Center', ApJ **640**(2), L159–L162.

Green, P. J. (1995), 'Reversible jump markov chain monte carlo computation and bayesian model determination', *Biometrika* **82**(4), 711–732.
URL: *https://doi.org/10.1093/biomet/82.4.711*

Handley, W. (2019), 'anesthetic: nested sampling visualisation', *The Journal of Open Source Software* **4**(37), 1414.
URL: *http://dx.doi.org/10.21105/joss.01414*

Handley, W. J., Hobson, M. P. & Lasenby, A. N. (2015), 'POLYCHORD: next-generation nested sampling', MNRAS **453**(4), 4384–4398.

Hastings, W. K. (1970), 'Monte carlo sampling methods using markov chains and their applications', *Biometrika* **57**(1), 97–109.
URL: *https://doi.org/10.1093/biomet/57.1.97*

Healey, S. E. et al. (2008), 'Cgrabs: An all-sky survey of gamma-ray blazar candidates', *Astrophys. J. Suppl.* **175**, 97.

Heavens, A., Fantaye, Y., Mootoovaloo, A., Eggers, H., Hosenie, Z., Kroon, S. & Sellentin, E. (2017), 'Marginal Likelihoods from Monte Carlo Markov Chains', *arXiv e-prints* p. arXiv:1704.03472.

Hee, S., Handley, W. J., Hobson, M. P. & Lasenby, A. N. (2016), 'Bayesian model selection without evidences: application to the dark energy equation-of-state', MNRAS **455**(3), 2461–2473.

Hergt, L. T., Handley, W. J., Hobson, M. P. & Lasenby, A. N. (2021), 'Bayesian evidence for the tensor-to-scalar ratio r and neutrino masses m_ν : Effects of uniform versus logarithmic priors', Phys. Rev. D **103**(12), 123511.

Higson, E., Handley, W., Hobson, M. & Lasenby, A. (2019), 'Bayesian sparse reconstruction: a brute-force approach to astronomical imaging and machine learning', MNRAS **483**(4), 4828–4846.

Hovatta, T., Valtaoja, E., Tornikoski, M. & Lähteenmäki, A. (2009), 'Doppler factors, Lorentz factors and viewing angles for quasars, BL Lacertae objects and radio galaxies', A&A **494**(2), 527–537.

Jeffreys, H. (1961), *The Theory of Probability*, Oxford University Press.

Kass, R. E., Carlin, B. P., Gelman, A. & M., N. R. (1998), 'Markov chain monte carlo in practice: A roundtable discussion', *The American Statistician* **52**(2), 93–100.
URL: *https://www.tandfonline.com/doi/abs/10.1080/00031305.1998.10480547*

Kellermann, K. I. et al. (2003), Superluminal motion and relativistic beaming in blazar jets, *in* L. O. Takalo & E. Valtaoja, eds, 'High Energy Blazar Astronomy', Vol. 299 of *Astronomical Society of the Pacific Conference Series*, p. 117.

Lartillot, N. & Philippe, H. (2006), 'Computing Bayes Factors Using Thermodynamic Integration', *Systematic Biology* **55**(2), 195–207.
URL: *https://doi.org/10.1080/10635150500433722*

MacKay, D. (2003), *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

Marin, J. & Robert, C. P. (2009), 'Importance sampling methods for Bayesian discrimination between embedded models', *arXiv e-prints* p. arXiv:0910.2325.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**(6), 1087–1092.
URL: *https://doi.org/10.1063/1.1699114*

Neal, R. M. (2000), 'Slice Sampling', *arXiv e-prints* p. physics/0009028.

Obreschkow, D., Klöckner, H. R., Heywood, I., Levrier, F. & Rawlings, S. (2009), 'A Virtual Sky with Extragalactic H I and CO Lines for the Square Kilometre Array and the Atacama Large Millimeter/Submillimeter Array', ApJ **703**(2), 1890–1903.

Orr, M. J. L. & Browne, I. W. A. (1982), 'Relativistic beaming and quasar statistics.', MNRAS **200**, 1067–1080.

Osterbrock, D. E. (1978), 'Observational Model of the Ionized Gas in Seyfert and Radio-Galaxy Nuclei', *Proceedings of the National Academy of Science* **75**(2), 540–544.

Readhead, A. C. S. (2021), personal communication.

Readhead, A. C. S. et al. (1978), 'Bent beams and the overall size of extragalactic radio sources', Nature **276**, 768–771.

Richards, J. L. (2012), The Radio Variability of Gamma-Ray Blazars, PhD thesis, California Institute of Technology.

Richards, J. L. et al. (2011), 'Blazars in the fermi era: The ovro 40m telescope monitoring program', *The Astrophysical Journal Supplement Series* **194**(2), 29.

Roy, V. (2020), 'Convergence diagnostics for markov chain monte carlo', *Annual Review of Statistics and Its Application* **7**(1), 387–412.
  URL: *https://doi.org/10.1146/annurev-statistics-031219-041300*

Schmidt, M. (1965), 3C 273: A Star-like Object with Large Red-Shift, *in* I. Robinson, A. Schild & E. L. Schucking, eds, 'Quasi-Stellar Sources and Gravitational Collapse', p. 455.

Schneider, P., Ehlers, J. & Falco, E. E. (1992), *Gravitational Lenses*, 2 edn, Springer Publishing.

Seyfert, C. K. (1943), 'Nuclear Emission in Spiral Nebulae.', ApJ **97**, 28.

Sivia, D. S. & Skilling, J. (2006), *Data analysis: A Bayesian Tutorial*, Oxford University Press.

Skilling, J. (2006), 'Nested sampling for general Bayesian computation', *Bayesian Analysis* **1**(4), 833 – 859.
  URL: *https://doi.org/10.1214/06-BA127*

Slipher, V. M. (1917), 'The spectrum and velocity of the nebula N.G.C. 1068 (M 77)', *Lowell Observatory Bulletin* **3**, 59–62.

Trotta, R. (2008), 'Bayes in the sky: Bayesian inference and model selection in cosmology', *Contemporary Physics* **49**(2), 71–104.

Vedantham, H. K. et al. (2017*a*), 'Symmetric Achromatic Variability in Active Galaxies: A Powerful New Gravitational Lensing Probe?', ApJ **845**(2), 89.

Vedantham, H. K. et al. (2017*b*), 'The Peculiar Light Curve of J1415+1320: A Case Study in Extreme Scattering Events', ApJ **845**(2), 90.

Walker, M. & Wardle, M. (1998), 'Extreme Scattering Events and Galactic Dark Matter', ApJ **498**(2), L125–L128.

Wilson, A. S. & Colbert, E. J. M. (1995), 'The Difference between Radio-loud and Radio-quiet Active Galaxies', ApJ **438**, 62.