Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Copyright Statement

Gravitational Microlensing: A Bayesian Modelling System Using MultiNest

Ashna Sharan

A thesis submitted in fulfillment
of the requirements for the degree of
Master of Philosophy
in
Physics
The University of Auckland
2019

Abstract

In this project, microlensing events are modelled and analysed using a new methodology employing the MultiNest algorithm. MultiNest is based on the principles of Bayesian inference, which allows us to solve the model selection and parameter estimation problems simultaneously. The focus is placed on the model selection problem since a Bayesian based algorithm such as MultiNest allows us to shift the approach to model selection from qualitative arguments to a quantitative quality factor.

The methodology is demonstrated by testing a finite-source point-lens model versus a finite-source binary-lens model as well as testing for the presence of parallax effects. This is done for a simulated synthetic event as proof of concept and for a real event, OGLE-2011-BLG-0251.

Nested Sampling and its variant algorithms such as MultiNest have been tried and tested in many fields of study. By demonstrating MultiNest on a real microlensing event, the aim of this project is to provide an impetus for said algorithms to find their place in the microlensing community as well.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A complete modelling and analysis of data involves parameter estimation as well as model selection. While there are many optimization methods that allow modellers to find parameter estimates in a direct straightforward manner, model selection is often left to ad hoc methods and qualitative arguments.

Microlensing modelling is a challenging problem due to the highly non-linear physics, the large, multidimensional parameter space and the rough chi-square landscape ridden with many hills and valleys. Furthermore, model selection is an important but difficult task because it is not as simple as choosing the model that best fits the data as this can lead to over-parameterised models. Occam's Razor must be applied.

In microlensing modelling, parameter estimation is usually achieved using maximum likelihood estimation or χ^2 minimization methods. Model selection is achieved by comparison of the χ^2 "goodness of fit", with a qualitative application of Occam's Razor, or experimental and theoretical arguments from prior knowledge to inform the choice.

Bayesian based data modelling methods like Nested Sampling (NS) offer a straightforward way of performing parameter estimation and model selection simultaneously. However, often parameter estimation is given prominence over model selection while selecting the model with parameter estimates that give the best fit to the data.

In this project, the aim is to develop a methodology that applies Bayesian model selection to microlensing modelling, using a version of the NS algorithm called MultiNest (Feroz et al., 2009, 2013, Feroz and Hobson, 2008). In developing the methodology, the MultiNest algorithm was integrated with an existing binary-lens microlensing modelling code whereby light curve computations, hence the likelihood computations, were GPU-accelerated (Ling, 2013).

In Chapter 2 the theory of microlensing is discussed whereas in Chapter 3, the theory of Bayesian data modelling as well the trends and gaps of data modelling in microlensing are discussed. The specifically developed methodology and modelling strategy employing the MultiNest algorithm and the GPU-accelerated code are described in Chapter 4. In Chapter 5, the analysis of a synthetic event is presented and in Chapter 6, the analysis of a real event, OGLE-2011-BLG-0251 (hereafter OB11-0251), is presented. Finally, a brief discussion and the conclusions are outlined in Chapter 7.

Chapter 2

Gravitational Microlensing

Microlensing is an astronomical phenomenon that can be observed when a massive object (the lens), a bright background object (the source star) and the observer become sufficiently aligned. The lens could be a single massive dark object such as a free-floating planet or a black-hole. It could also be multiple body systems, such as a star with a planet or a stellar binary.

When the light from a background source star is deflected by the gravity of a foreground lens system, multiple images are created due to the lensing effect. The multiple images have a larger total solid angle than the unlensed source star image, but are unresolved so that they appear as one image brighter than the unlensed source star image. Consequently, since surface brightness is conserved while the image areas evolve in time, the source star appears to brighten with time as the projected distance of closest approach between the source and lens star decreases. A good introduction to microlensing can be found in review articles by Bennett (2008), Dominik (2010), Gaudi (2010), Gould and Loeb (1992), Rattenbury (2006) and references therein.

First proposed as a means of detecting dark matter (Paczynski, 1986), it has now become a powerful technique for detecting exoplanets (Gould and Loeb, 1992, Mao and Paczynski, 1991). Microlensing is a powerful planet search technique, particularly well-suited, and more sensitive than other techniques, to finding cool low-mass planets and planets around distant or very dim stars. This is evidenced by the discovery of an exoplanet only 3.3 times the mass of Earth (Bennett et al., 2008). Hence, it plays a significant role in studying the statistical abundance of exoplanets in our Galaxy with properties similar to the planets in our own Solar System, as well as for studying planet formation and evolution. This makes gravitational microlensing an invaluable technique in the quest for finding an Earth-like planet.
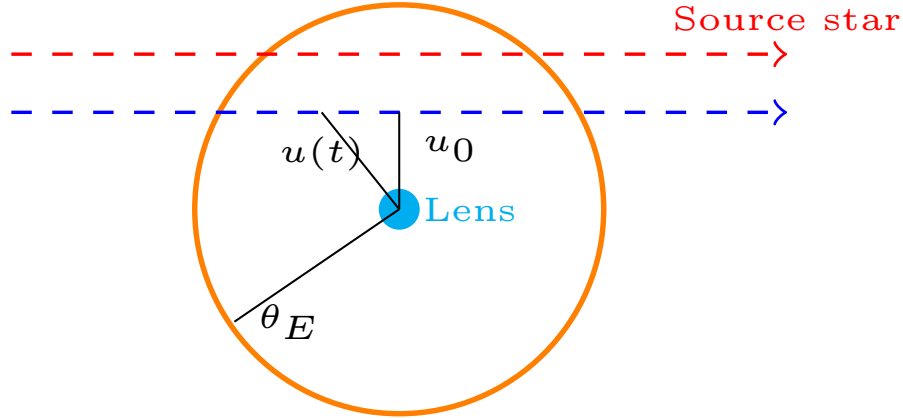
FIGURE 2.1: The Einstein ring represented in this figure with ring radius, θ_E, forms when the source, lens and observer align perfectly and the images formed due to the lensing effect merge together. The dashed lines represent the track of the source star as it moves relative to the lens star where the $u(t)$ denotes the projected distance between the source star and the lens star at time, t. Thus u_0 represents the projected distance of closest approach of the source to the lens.

In the case of perfect alignment of the source, lens and observer, the images form a ring-like structure known as the Einstein ring. This characteristic projected separation between the source and the lens inside which microlensing becomes significant is called the Einstein ring radius (represented in Figure 2.1). Given G, the gravitational constant, the speed of light, c and lens mass, M, the angular Einstein radius is:

$$\theta_E = \sqrt{\frac{4GM}{c^2}\left(\frac{D_S - D_L}{D_S D_L}\right)}, \tag{2.1}$$

where D_S is the distance from the observer to the source and D_L is the distance from the observer to the lens.

2.1 Single Lens Microlensing

When the foreground lens system is a single massive object, point-source point-lens (PSPL) approximation is assumed which considerably simplifies the geometrical derivation of single lens microlensing.
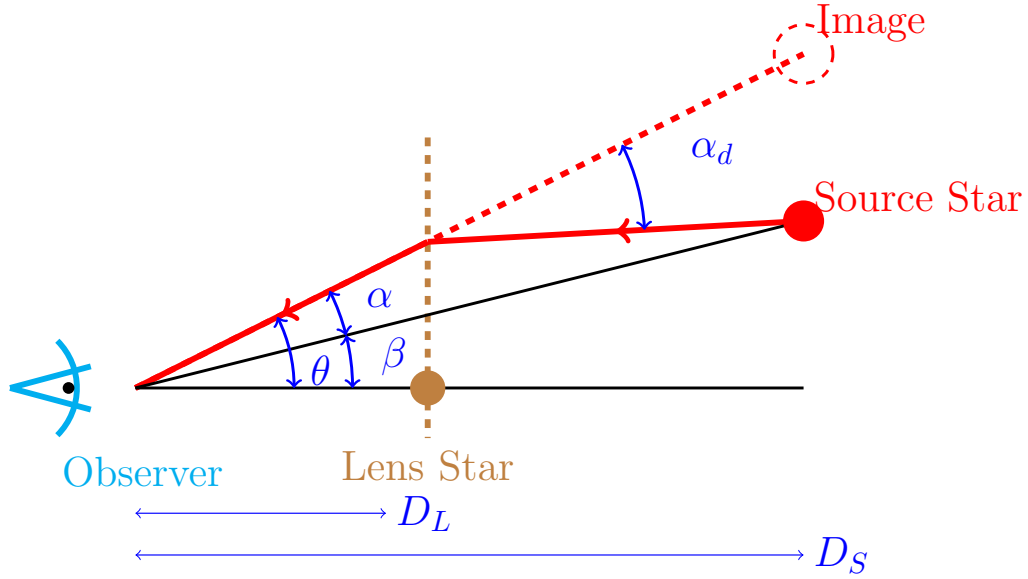
FIGURE 2.2: This figure shows the standard microlensing geometry, where D_S is the distance from the observer to the source star and D_L is the distance from the observer to the lens star. When the light from a source star passes a foreground lens star with mass, M, it is deflected by an angle derived by Einstein (1936) and represented in the figure by $\alpha_d = \frac{4GM}{c^2 r_E}$, where r_E is the Einstein radius. The angular positions of the source and the lens give the relationship, $\beta = \theta - \alpha$.

The basic geometry of a single lens microlensing system is presented in Figure 2.2 from which the lens equation may be derived, briefly shown here. More detailed derivations can be found in microlensing literature such as Bennett (2008) and Ling (2013).

From the relationship between the angular positions of the source and the lens stars, and using small angle approximations, an equation of the following form can be derived:

$$\beta = \theta - \theta_E^2/\theta. \tag{2.2}$$

Normalizing this equation with the Einstein angle, θ_E so that $\frac{\beta}{\theta_E} = u$ and $\frac{\theta}{\theta_E} = y$, the lens equation becomes:

$$u = y - y^{-1}. \tag{2.3}$$

Therefore, for a single lens microlensing event, a simple lens equation can be derived geometrically, which can be inverted to find the image positions given the source position. The inverted lens equation in this case is an easily solvable quadratic equation with two image solutions to be found.

Total magnification is then found by taking the ratio of the total area of the images to the area of the source. It is represented by the equation:

$$A(u) = \frac{u^2 + 2}{u\sqrt{u^2 + 4}}, \tag{2.4}$$

where $u(t)$ is the projected distance (in Einstein radius units, θ_E) between the lens and the source at time t and is given by:

$$u(t) = \sqrt{u_0^2 + (\frac{t - t_0}{t_E})^2}. \tag{2.5}$$

Therefore, the PSPL light curve can be parameterised with three parameters: the time at which the source passes closest to the centre of mass of the binary lens (the time of peak magnification), t_0, the Einstein radius crossing time (the duration of the microlensing event), t_E and the minimum impact parameter, u_0, which is the projected distance (in Einstein radius units, θ_E) of closest approach of the source to the lens.

Hence, for a single lens event, the source star appears to brighten, reach a peak and then fade away with respect to time giving it the characteristic bell-shaped light curve (a function of brightness versus time), as shown in Figure 2.3. The light curve for a typical microlensing event is well fit due to the PSPL approximation and the only significant physical parameter is the Einstein time, t_E.

The magnification (Equation 2.4) becomes infinite when the lens star, source star and the observer become perfectly aligned ($u(t) = 0$). In reality, the finite size of the source star breaks the infinity. The finite-source point-lens (FSPL) model can be described with an additional parameter, ρ, the angular size of the source star, in units of θ_E. The finite size of the source star affects the shape of the microlensing light curves by causing features of the light curves such as the spikes and dips to appear washed out.

2.2   Multiple Lens Microlensing

Events leading to the discovery of exoplanets are represented by multiple lens microlensing events. For such events different light curve phenomenologies can be observed with their features affected by higher order effects such as parallax, orbital motion of the binary lens system, the finite source size or even the limb darkening profile of the source star. For example, the usual indication for a star-planet lens system is a dip or spike near the peak of the light curve while finite source effects can cause this planetary signal to appear washed out.
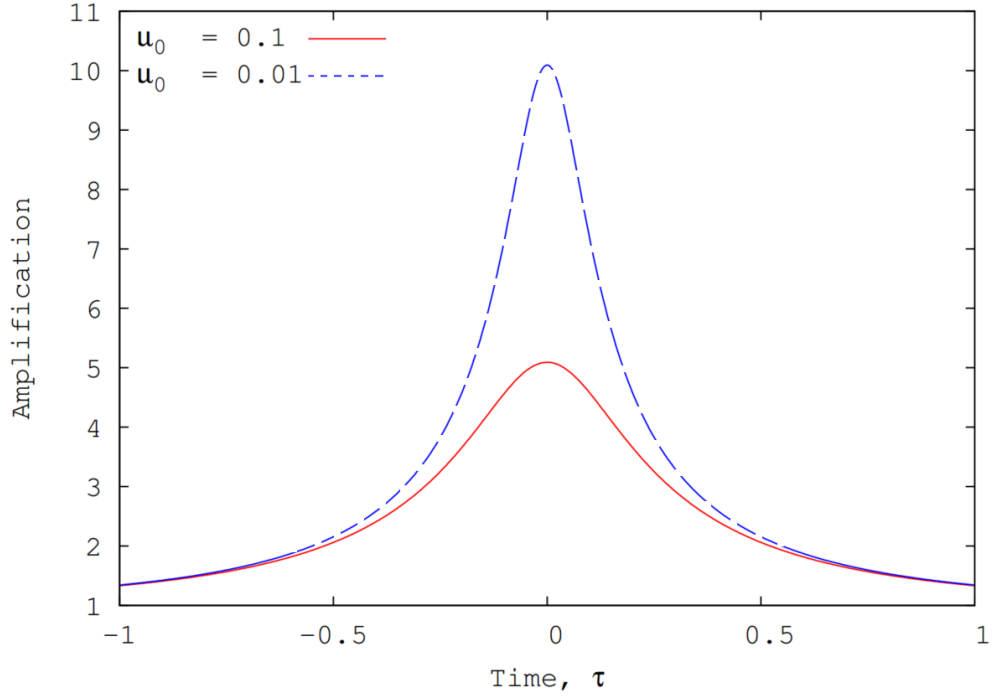
FIGURE 2.3: Bell-shaped light curves characteristic of single lens microlensing are shown here. Projected distances of closest approach of the source to the lens, $u_0 = 0.01$ and 0.1 were used in generating these light curves, with normalised time, $\tau = \frac{t-t_0}{t_E}$.

Consequently, additional parameters are required to model the multiple lens microlensing light curves. For example, a finite-source binary-lens (FSBL) model can be described by 7 free parameters. The free parameters include t_0, t_E, u_0, ρ, which are also used to describe FSPL light curves. Additional parameters are: α which is the source trajectory angle with respect to the lens components, d which is the projected separation between the two mass components and q which is their mass ratio.

Limb darkening effect of the source star, whereby the centre of the star appears brighter than the edge or limb of the star, can also affect the shape of microlensing light curves. This effect can be described by two coefficients which can be allowed to be free parameters or fixed based on the standard limb darkening model based on the colour of the source star (Gaudi et al., 2003).

Microlensing light curves can also exhibit the parallax effect due to the Earth's orbit around the Sun (Gould, 2000, 2004). Parallax can affect the alignment of the observer to the source during a microlensing event thereby affecting the microlensing light curves, inducing asymmetrical features. Hence, we also test for the presence or absence of parallax in modelling the microlensing events in this paper. Two parameters are used to describe parallax, $\pi_{E,E}$ and $\pi_{E,N}$, which are the parallax in the East and North directions respectively, in Einstein radius units. The magnitude of the parallax effect, π_E is then

given by,

$$\pi_{\mathrm{E}} = \sqrt{\pi_{\mathrm{E,E}}^2 + \pi_{\mathrm{E,N}}^2}. \tag{2.6}$$

With the introduction of additional parameters in multiple lens microlensing events, the PSPL approximation breaks, and the underlying physics becomes highly non-linear. For the purpose of modelling, the parameter space to be optimized is large, multidimensional and multi-modal with lots of local minima appearing in its "goodness-of-fit" χ^2 landscape.

For the single lens case, it is straightforward to invert the lens equation and find the magnification with respect to time for a specific set of parameters. Unlike the single lens event, there is no explicit expression for the amplification profile for the multiple lens case. The magnification has to be computed by finding the locations of all the images and summing their magnifications via non-analytical or semi-analytical techniques.

Generalisation of the lens equation for the single lens case gives the lens equation for the multiple lens case. In complex notation, it is as follows:

$$w = z - \sum_{j=1}^{N} \frac{\varepsilon_j}{\bar{z} - \bar{z}_{m,j}} \tag{2.7}$$

where, ε_j is defined by the mass of the jth lens divided by the combined mass of all lenses, $\bar{z}_{m,j}$ are the complex positions of the lens masses whereas w and z are the complex positions of the source and images respectively.

It is not as straightforward as the single-lens case to invert this equation and solve for the image positions. In general $N^2 + 1$ degree polynomial would need to be solved to arrive at an analytical solution for the amplification profile of an N-lens system. A binary lens system would require finding the solutions to a fifth degree polynomial while a triple lens system would require a tenth degree polynomial and so on, which gets increasingly computationally expensive.

It is possible to compute the magnification by solving the lens equation with its inverted complex polynomial but the inverted lens equation can not be used to compute the finite source radius. The lens equation is also undefined for caustic curves on the source plane. Caustic curves are closed regions of theoretically infinite source star magnification (in case of point-source approximation), which can be directly mapped from corresponding critical curves on the lens plane.

With such huge mathematical and computational demands, microlensing modelling and the determination of the lens system properties becomes a tremendously challenging task and requires the use of unorthodox methods to accomplish.

Numerical methods of modelling microlensing events addresses many of these problems. The modelling code used in this research project that addresses these challenges was developed by Ling (2013) and is discussed in Chapter 4.

Chapter 3

Data Modelling

3.1 Parameter Estimation Methods in Microlensing

Microlensing modelling (real-time and offline) is a non-trivial, difficult non-linear optimization problem ridden with many computational challenges. It requires a large multidimensional parameter space to be explored and gives rise to degeneracies and a highly multimodal posterior.

One of the most difficult optimization tasks is a non-linear modelling problem with a multimodal posterior, that is, a parameter search space ridden with many optima or in other words, a complex "goodness of fit", χ^2 topology with many hills and valleys.

In the case of a multimodal posterior that has an obvious single best estimate, finding the global optima in a large multidimensional parameter space is a serious computational problem since there are many local optima throughout the posterior that the optimization algorithm may get trapped in.

In the case of a multimodal posterior having many optima of comparable magnitude, the question of which optima represents the best estimate model arises and model comparison becomes a significant problem.

As discussed in Gregory (2005), Sivia and Skilling (2006), the different optimization methods have their merits but also meet with limitations when faced with the aforementioned challenges.

A brute force method of grid search, for instance, is conceptually simple and capable of dealing with multimodality but quickly becomes impractical for the multidimensional problem beyond two parameters. It becomes computationally expensive and time-consuming, even with a GPU-accelerated code and may even miss optima existing in narrow valleys in between the grid points.

A local optimization method such as the Nelder and Meade downhill simplex (Press, 2007) works well for a multidimensional problem in case of a unimodal posterior whereas it would be prone to getting trapped in the nearest local minima in case of a multimodal problem. It can not switch easily between global exploration and local optimization. The algorithm would need to perform jumps by providing several start values for one parameter, for example. In addition, the starting values would need to be specified carefully since it would not explore too far away from this point. Step sizes would also need to be chosen carefully.

On the other hand, a global optimization algorithm based on random search techniques such as genetic algorithms are capable of finding the global optimum in a multi-modal and multidimensional parameter space but they may not be capable of homing in on the peak of the optimum.

Although intelligent search strategies appear less frequently in microlensing literature, they have indeed been studied for their potential strengths. For instance, a classical genetic algorithm (Charbonneau, 1995) has been used in modelling binary lens events (Kubas et al., 2005, Kubas, 2005). The genetic algorithm in these cases was followed by gradient based techniques for final refinement of the parameter values.

Similarly, in their analysis of caustic crossing binary lens events, Kains et al. (2012, 2009) and Cassan et al. (2010) started their search scheme with a grid search, then used a GA followed by refinement of the model by an Markov chain Monte Carlo (MCMC) fit.

While the microlensing modelling community uses different combinations of optimization methods, MCMC methods remain ubiquitously used in finding best-fit models with numerical uncertainties.

However, even MCMC algorithms have their limitations, for example, simulated annealing, an MCMC method may be successful in finding a dominant optimum with a distinct underlying trend, it would not do a good job of finding dominant optima widely separated by insignificant optima.

Also, as mentioned in (Kains et al., 2009), a classical MCMC fitting scheme alone would be unlikely to explore and locate minima in regions of parameter space far away from the starting point.

However, while a combination of different optimization algorithms tailored specifically to the microlensing problem may succeed in exploring the parameter space efficiently, the above-mentioned methods offer no model comparison techniques except for the comparison of a "goodness of fit" measure combined with qualitative arguments to rule out improbable scenarios.

3.2   The Model Selection Gap in Microlensing Methods and Proposed Solution

A significant problem in microlensing modelling, in any modelling problem in fact, is choosing rationally between alternative models. Given two or more models, the model selection methodology should identify the model with the parameter estimates that explains the data best, while penalizing overly-complex, over-parameterised models.

However, optimization methods used in microlensing modelling, are primarily designed for finding the best-estimate parameters with numerical uncertainties. Parameter estimation is usually achieved using maximum likelihood estimation or χ^2 minimization methods. Model selection is achieved by comparison of the χ^2 "goodness of fit", with a qualitative application of Occam's Razor, or experimental and theoretical arguments from prior knowledge to inform the choice. This approach of making choices between alternative models is prevalent in microlensing modelling.

Some problems can be anticipated to arise from this approach. As Sivia and Skilling (2006) point out, the model selection problem is not as simple as choosing the best fitting model.

> "Naively we might think that a choice between proposed alternatives can be made on the basis of how well they fit the data. A little reflection soon reveals a potential difficulty in that more complicated models, defined by many parameters, will always be able to give better agreement with the experimental measurements."

There are cases whereby the microlensing community do not agree on the best-fit model; there are alternative models proposed for the same event by different modellers or even by the same modeller. Jaroszyñski et al. (2010) showcases many such alternative models

of similar fit quality to anomalous microlensing events in the OGLE-III EWS database, seasons 2006-2008.

There has also been at least one case of a published model retracted because a simpler model of similar χ^2 fit was found. Originally a model with a planetary mass ratio and orbital motion was published for the event OGLE-2008-BLG-513. Later on, a static binary star lens solution with a comparable χ^2 value was found and since the latter is the simpler of the two models, following Occam's Razor it has been accepted as likely to be the correct solution (Yee et al., 2011).

In (Kains et al., 2009) the global χ^2 minimum found for OGLE-2007-BLG-472 (OB07-472) identified a model with an extremely long timescale, $t_E \approx 2000$ days, which was rejected through a qualitative discussion. In Kains et al. (2012) the authors re-analysed OB07-472 with a Bayesian analysis including appropriate priors so that there was a natural shift to a favourable model with $t_E \sim 70$ days, a more typical event timescale.

While the χ^2 minimization method does not take into account the plausibility of physical parameters, it is easy to avoid implausible physical parameter estimations by constraining Bayesian priors using prior knowledge of typical parameter values. In microlensing, priors on physical parameters can be estimated from Galactic models or from empirical distributions of the physical parameters obtained from a large number of observed events (Cassan et al., 2010). In addition, calculation of the Bayesian evidence is a quantitative approach to the model selection problem (Sivia and Skilling, 2006).

Hence, the Bayesian approach offers a much more powerful way of comparing models by quantifying Occam's Razor and automatically penalizing more complicated models unless justified by the complexity of the data.

Furthermore, not only does the evidence value allow modellers to perform model selection between alternative current models, as Skilling (2006) points out, the evidence value allows models to be tested against any proposed alternatives, even future ones, without having to re-do the current calculation.

Evidently, in microlensing modelling, as in many other fields of study, comparing models is a serious matter whereby properties of exoplanet systems or stellar binary-lens systems or double source star systems, for instance, have to be inferred from the observational data.

The application of Bayesian data modelling is proposed as the solution to this gap in model selection methodology. The following sections describe the Bayesian data modelling theory and the optimization algorithm based on Bayesian theory, Nested

Sampling (NS) and its modified version, MultiNest, which was implemented in this research project.

3.3   Bayesian Data Modelling

A Bayesian approach to data modelling and analysis has been adopted in this project and for an illustrative introduction and tutorial on this data analysis approach, the reader is referred to Sivia and Skilling (2006).

Briefly stated, given a set of parameters for model M represented by parameter vector $\mathbf{X}$, measured data represented by data vector $\mathbf{D}$, Bayes' Theorem takes the form:

$$P(\mathbf{X}|\mathbf{D}, M) = \frac{P(\mathbf{D}|\mathbf{X}, M)P(\mathbf{X}, M)}{P(\mathbf{D})} \tag{3.1}$$

where the symbol '$P(x|y)$' denotes a conditional probability of x given y. In the cases of continuous rather than discrete set of probabilities, such as in this paper, it denotes the probability distribution function (hereafter pdf).

In this equation, $P(\mathbf{X}, M)$ is termed the prior and is the probability that the proposed set of parameter values is true before the data are observed. $P(\mathbf{D}|\mathbf{X}, M)$, the likelihood, is the probability of observing the data given that the model M and its parameters $\mathbf{X}$ are true.

The denominator, $P(\mathbf{D})$, is variously called the evidence, marginal likelihood or the normalization constant and is the probability of the data (irrespective of parameter values) given the model and is defined by:

$$P(\mathbf{D}) = \int P(\mathbf{D}|\mathbf{X}, M)P(\mathbf{X}, M)d\mathbf{X}. \tag{3.2}$$

Hereafter, the model evidence, $P(\mathbf{D})$ is denoted by Z.

$P(\mathbf{X}|\mathbf{D}, M)$, the posterior, is the probability of the model parameter values given the observed data. Thus,

$$posterior = \frac{likelihood \times prior}{evidence}. \tag{3.3}$$

Bayes' Theorem is a succinct statement of our state of knowledge before and after the data and available information are taken into account. It makes data analysis simple and straightforward by relating the posterior pdf to the likelihood. As Sivia and Skilling (2006) state,

> "The power of Bayes' theorem lies in the fact that it relates the quantity of interest, the probability that the hypothesis is true given the data, to the term we have a better chance of being able to assign, the probability that we would have observed the measured data if the hypothesis was true."

In the Bayesian approach to data analysis, there are two distinct requirements for a complete analysis, parameter estimation and model selection. The posterior pdf allows us to estimate the set of model parameters $\mathbf{X}$ whereas the evidence values allow us to perform model selection.

The evidence value is an integration over the entire parameter space of the model hence it can be thought of as a prior-weighted average of the likelihood. Therefore, it can be said that while parameter estimation requires us to look at the location of the maxima of the likelihood function, model selection requires the calculation of the average value of the likelihood.

Generally, more complicated models with greater numbers of parameters offer better agreement with data. Hence a complicated model might be favoured over a simpler model by a "goodness of fit" criterion, even though it may not necessarily be the best representative model. With the Bayesian approach to model selection, the gain in "goodness of fit" offered by a more complicated model is balanced against the cost of averaging the likelihood over the larger parameter space of the more complicated model. Hence, the Bayesian evidence penalizes the more complicated model. As a result, there is an inbuilt quantitative implementation of Occam's Razor in the Bayesian method.

For model comparison and selection, two approaches can be considered, as also outlined in (Buchner et al., 2014). In the first approach, Bayes' factor (K), the ratio of model evidences can be computed and Jeffreys' scale (Jeffreys, 1961) can then be used for the interpretation of K, the Bayes factor. $K > 100$ is taken as 'decisive' evidence that M_1 is favoured over M_2; $30 - 100$ is taken as 'very strong evidence', $10 - 30$ as 'strong evidence' and $3 - 10$ as 'substantial evidence'. These Bayes factors are equivalent to differences of 4.3, 3.4, 2.3 and 1.1 respectively, when dealing with log-evidences, $\ln Z$. When the difference between the log-evidences of two models is below 1.1, the two models are considered equally probable.

An alternative approach is that of comparing all the models simultaneously and selecting the model with the highest evidence or log-evidence value. This is the approach of model comparison and selection adopted in this research project, relying on Jeffreys' scale for pairwise interpretation where necessary.

3.4   Overview of Nested Sampling and MultiNest

With the tremendous increase in the volume of astrophysical data in recent years, Bayesian techniques have gained popularity with regards to making inferences on best-estimate parameters and model comparisons. Refer to (Trotta, 2008) and (Ford and Gregory, 2007) for a review of the applications of Bayesian parameter estimation and model selection in cosmology and extrasolar planet searches respectively.

In the Bayesian inference approach, there is a clear distinction between parameter estimation and model selection. With the computation of the Bayesian evidence value, there is a straightforward model comparison strategy to be found in the Bayesian approach.

However, the traditional MCMC methods which are based on the Bayesian approach, are designed primarily for the computation of their posterior distribution while ignoring the Bayesian by-product - the Bayesian evidence. MCMC methods also become inefficient in case of highly multimodal posteriors. In contrast, NS, a novel MCMC-type algorithm put forth by Skilling (2004, 2006) is targeted at computing the Bayesian evidence with posterior distributions as the by-product.

Unsurprisingly, NS has aroused a lot of attention resulting in a number of variant algorithms. For instance, Feroz and Hobson (2008) developed the multimodal NS as an efficient and robust alternative to MCMC for astronomical data analyses. Brewer et al. (2011) found their new algorithm, Diffusive NS to be four times the accuracy compared to the classic MCMC-based NS, for the same computational effort, "equivalent to a factor of 16 speedup", when run on a test problem.

NS has already found many successful applications. To name a few, it has been applied to cosmological model selection (Mukherjee et al., 2006), to radial velocity data in testing between models of multiplanet systems (Feroz et al., 2011) and to Kepler data for characterizing transiting planets (Placek et al., 2014).

Bayesian evidence does not affect the relative magnitudes of the posterior probabilities, therefore it may be ignored in the cases where parameter estimation is the only goal. Consequently, more often than not, data modellers using the Bayesian approach overlook the Bayesian evidence, which is often expensive to compute even though it provides a straightforward way of performing model selection. With numerical computations of log-evidence ($\ln Z$), the NS algorithm provides an affordable way to compute Bayesian evidence and perform model selection.

NS focuses on calculating the evidence value as its primary task and the computation of the posteriors as a by-product. The evidence value is the result of an integration and is computed numerically, as a summation, by the NS algorithm. NS essentially samples

N points from the prior pdf, sorting them according to their log-likelihood values. At each iteration, the lowest log-likelihood point is discarded and replaced by a new sample which is chosen from the prior pdf and accepted only if it has a higher log-likelihood than the point that was discarded. Once the change in the computed log-evidence values falls below a threshold value, the algorithm terminates.

In this project, the MultiNest algorithm is used, a reportedly more efficient and robust variation of the NS algorithm. The improvements of MultiNest over NS are detailed in (Feroz et al., 2009). MultiNest increases the efficiency of the NS algorithm by clustering the samples into ellipsoids, bounded by their worst log-likelihood values. The new samples are taken from the interior of these ellipsoids, which makes MultiNest more efficient. Since a large range of evidence values is being dealt with, MulNest computes and deals with log-evidence values instead, as a matter of mathematical convenience.

Feroz et al. (2009) note that MultiNest is controlled by two main parameters, N and e. The number of sampled points, N, must always exceed the dimensionality of the parameter space and must be large enough to allow adequate sampling of all regions of the parameter space, for an accurate calculation of the log-evidence. The efficiency parameter, e, controls the sampling volume and must be set to 1 for maximum efficiency in parameter estimation problems and to a lower value if one is interested in the evidence value as well. The MultiNest algorithm converges and terminates when the contribution from the current set of points to the log-evidence value falls below a user-specified threshold (ranging between 0 to 1), the tolerance (*tol*) value (Feroz et al., 2013).

Chapter 4

Methodology

4.1 The GPU-accelerated Binary-lens Modelling Code

The binary-lens microlensing modelling code utilized in this project has been developed by Ling (2013). There are two modelling approaches used in his code, the magnification map approach and the dynamic light curve engine.

Inverse ray shooting (IRS) (Wambsganss et al., 1992), a brute force numerical technique, is used in the magnification map approach. In this technique, billions of light rays are shot backwards from the lens plane onto the source plane to determine the magnification of each pixel on the source plane.

This bypasses the computational difficulty of inverting the $N^2 + 1$ degree polynomial for an N-lens system. Rather, the general lens equation is used in a straightforward manner to compute the light ray paths. The lens equation maps each position on the lens plane to a corresponding position (x, y) on the source plane.

A large number of rays are mapped from a uniform distribution on the lens plane onto a corresponding distribution on the source plane. Each ray is computed, weighted by the limb-darkening profile as described by Ling (2013) and then accumulated and stored as a two dimensional array of theoretical solutions called a magnification map, representing each pair of lens geometry parameters $\{q, d\}$. Figure 4.1 is a visual representation of a magnification map where the brighter regions depict higher magnification.

Once this magnification map has been created, lines representing source tracks can be drawn across it to extract theoretical light curves parameterised by $\{t_0, t_E, u_0, \alpha, \rho\}$.

The image-centred inverse ray shooting technique is used in the dynamic light curve engine approach. Here the rays are shot from just those regions of the lens plane that
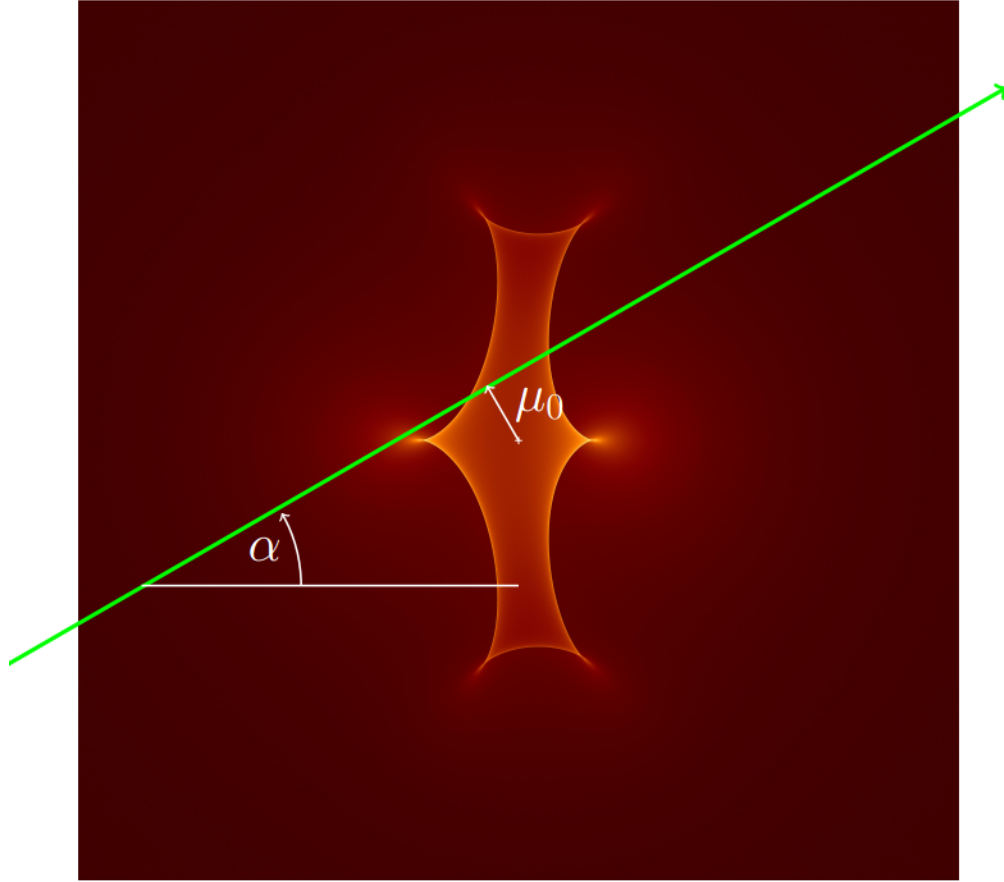
FIGURE 4.1: This figure shows a magnification map with a source star track drawn across it. α is the source trajectory angle with respect to the lens and u_0 is the projected distance (in Einstein radius units) of closest approach of the source to the lens. The magnification map is a two dimensional array of theoretical solutions of the lens equation for a pair of lens geometry parameters; q, the mass ratio and d, the separation of the lens components. Brighter regions on the magnification map depict higher magnification regions. The curves and cusps of the pattern in the centre of the magnification map are the caustics which represent regions of theoretically infinite magnification.

map to the source disk. The dynamic light curve engine is useful when modelling complex microlensing events with orbital motion effects, for instance. It is also appropriate when setting the parameters $\{q, d\}$ free during the optimization process, where a magnification map can not be reused. For each amplification point to be computed a new magnification map has to be generated.

The Graphics Processing Unit (GPU) is a massively parallel processor that can perform calculations on massive amounts of data simultaneously. Since the inverse ray shooting technique is highly parallel, researchers have begun accelerating their gravitational lensing and microlensing modelling codes using GPU architectures (Ling, 2013, McDougall, 2014, Thompson et al., 2010). Using GPU-architecture, Ling (2013) has achieved impressive improvements in the computational speed for the computationally expensive

task of magnification map generation as well as a high performance extraction of light curves from the magnification maps.

For a microlensing event without the orbital motion effect, the magnification map is easily reused to extract many light curves. Within one magnification map the downhill simplex method is used to optimize in the parameter space $\{t_0, t_E, u_0, \alpha, \rho\}$. Hundreds or thousands of magnification maps have to be searched in the parameter space $\{q, d\}$. It is possible to search thousands of magnification maps since the GPU-accelerated code takes only seconds to generate a magnification map. Therefore, the magnification maps can be generated on demand for a specific q and d parameter combination without storing them on disk.

A grid search coupled with the downhill simplex method finds candidate models optimised in the parameter space $\{t_0, t_E, u_0, \alpha, \rho\}$ with different q and d combinations. However, these models have not been optimized in the parameter space $\{q, d\}$. These initial models therefore act as an input to a modified Markov Chain Monte Carlo (MCMC) method for finding a more accurate model, optimized in the entire parameter space $\{q, d, t_0, t_E, u_0, \alpha, \rho\}$ for a binary lens microlensing model. The more accurate model is generated by the dynamic light curve engine which computes a light curve on demand without generating a magnification map.

This modified MCMC algorithm has mechanisms to avoid getting stuck in a local minima. It is part of the dynamic light curve engine which, in addition to optimizing the entire parameter space, is capable of dealing with complex microlensing events. For instance, in modelling complex microlensing events with orbital motion effects, a magnification map can not be reused and for each magnification point to be computed a new magnification map has to be generated. This makes the modelling of complex events even more time consuming and computationally expensive. The GPU-accelerated dynamic light curve engine which can compute light curves on demand has significant advantages in such modelling.

In this research, a specific methodology has been developed, particularly for model selection problems, by implementing the Bayesian-based MultiNest optimization algorithm. The MultiNest algorithm replaces the MCMC algorithm in finding more accurate parameter estimates and provides the additional functionality of straightforward Bayesian model selection. Since MultiNest requires thousands of likelihood evaluations for convergence, the advantages offered by a GPU-accelerated modelling code can be appreciated.

4.2 MultiNest Implementation

In this work, MultiNest v3.8 was used with $N = 200$ points and $tol = 0.5$ to achieve adequate accuracy without a prohibitively long computation time. Lower values of tolerance result in higher accuracy in the log-evidence estimates but require a much longer computation time. The quality of the log evidence estimates also increases by increasing the number of the N samples but once again this gain in accuracy occurs at the expense of speed.

PyMultiNest (Buchner et al., 2014), the Python interface to MultiNest, allowed the efficient integration of the MultiNest optimisation method with the binary lens modelling code by Ling (2013). PyMultiNest requires two major functions to be specified, the prior pdf and the log-likelihood function.

For defining the prior pdf, the MultiNest algorithm has a native parameter space whereby each parameter value varies from 0 to 1. This *"unit hypercube sampling space"* is explained in detail in (Feroz et al., 2009). Uniform priors were adopted for the parameters d, t_0, t_E, u_0, α and log-uniform priors for the parameters q and ρ since these two parameters range over several orders of magnitude. Hence, the MultiNest native space and our parameter space coincide, and it becomes a trivial matter of converting the unit interval $[0, 1]$ to the chosen parameter space interval. The determination of prior parameter space intervals is discussed in more detail in the Section 4.3

In the implementation of the MultiNest algorithm via PyMultiNest, a log-likelihood function has been used, as briefly derived below and given by the equation (4.5). The treatment of Sivia and Skilling (2006) has been followed in the choice of the log-likelihood function and so the reader is also referred to this textbook for full details on the derivation.

Assuming independent data, the likelihood is the product of that for each datum:

$$L = P(\mathbf{D}|\mathbf{X}, M) = \prod_{k=1}^{N} P(D_k|\mathbf{X}, M) \tag{4.1}$$

where D_k is the k-th datum. Assuming independent Gaussian measurement errors, the probability of an individual datum can be stated as:

$$P(D_k|\mathbf{X}, M) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left[\frac{-(F_k - D_k)^2}{2\sigma_k^2}\right], \tag{4.2}$$

where σ_k^2 is the error bar on D_k and F_k is the k-th ideal (noiseless) datum generated by an adequate model described by parameters $\mathbf{X}$. With equations (4.1) and (4.2) the

likelihood function can be approximated as follows:

$$P(\mathbf{D}|\mathbf{X}, M) \propto \exp\left[-\frac{\chi^2}{2}\right], \tag{4.3}$$

where χ^2 is the sum of squares of the normalized residuals,

$$\chi^2 = \sum_{k=1}^{N} \left(\frac{F_k - D_k}{\sigma_k}\right)^2. \tag{4.4}$$

The log-likelihood is then,

$$\ln L = \ln\left[P(\mathbf{D}|\mathbf{X}, M)\right] = C - \frac{\chi^2}{2}. \tag{4.5}$$

In addition to the log-evidence values, MultiNest outputs the maximum likelihood estimates (MLE), the maximum a-posteriori (MAP) parameter estimates and the mean, each with associated uncertainties. Given flat priors, the posterior pdf becomes directly proportional to the likelihood function. For this reason the MAP estimates are equivalent to the MLE estimates. The MAP parameter estimates are used to compute the best-fit model light curves. The log-evidence values, the MAP and the mean parameter estimates have been reported for the analyses of the synthetic and real event in the respective chapters.

4.3 The Modelling Strategy

The same strategy was used to model the synthetic event and the real event. In all the MultiNest modelling runs the value of the constant in the log-likelihood equation (4.5) was fixed at a value of 10^4, which is an arbitrarily chosen value for the constant, C, in Equation 4.5. This value affects the absolute evidence value but does not affect model comparison as long as it is fixed at one value for all the models being tested.

In microlensing modelling, large parameter spaces need to be explored for all the parameters and in the Bayesian framework, prior intervals would ideally reflect the entire range of parameter values considered viable. However, this approach of using the entire ranges of parameter values for prior intervals was found to be infeasible and was aborted due to it being extremely time consuming, while not returning comparable parameter estimate results.

For the events modelled in this research, a glance at the light curve data reveals that these are not single-lens events, since there are perturbations in the light curve rather than it

being a symmetrical bell curve (for a single-lens). However, an initial FSPL modelling is carried out to determine the starting points for t_0, t_E, u_0 and ρ to be specified for the downhill simplex optimization performed for each q, d during the grid search of the parameter space.

For the FSPL modelling, the magnification map technique was used to compute the light curves on a single magnification map, with mass fraction, $q = 0$, separation, $d = 0$ and angle, $\alpha = 0$. As a result, it usually takes on the order of minutes for the algorithm to converge, depending on the complexity of the event being modelled. Therefore, it was affordable to assign a wide range to the prior intervals for the initial FSPL modelling. The time span of the magnification event (roughly estimated from the light curve) was taken to be the prior interval for the parameter t_0. The prior intervals for the parameters t_E, u_0 and ρ were determined to be the range of values generally considered typical for these parameters. The parameter t_E is typically less than 200 days. For high amplification events, u_0 is roughly the inverse of the magnification at t_0. Furthermore, ρ is typically approximately 0.001-0.005 for main sequence stars and approximately 0.01-0.02 for red giant source stars.

A fixed grid-search was performed for the finite-source binary-lens (FSBL) model with 7 parameters. The grid search was performed over the range $-4 < \log 10(q) < 0.0$ and $-1.0 < \log 10(d) < 0.6$ with 40 uniform divisions for each parameter range. For each of the 1600 magnification maps created, α was explored over the range $0.0 < \alpha < 2\pi$ with 72 uniform divisions of the parameter search range of α. Each of these 72 values of α along with the user-specified starting points for t_0, t_E, u_0 and ρ are fed into the downhill simplex optimization algorithm. The grid search results aided in restricting the prior intervals for the 7 FSBL parameters.

For the FSBL with parallax effects (FSBLπ) model, the same constrained prior intervals for the 7 parameters are used as for the FSBL modelling. In addition, the prior intervals for parameters $\pi_{E,N}$ and $\pi_{E,E}$ were specified to be the range of values generally considered typical for parallax parameters. Another round of FSPL modelling was also carried out using the same prior intervals for t_0, t_E, u_0 and ρ, as used for the FSBL models. This is done so that the prior intervals for the all the models being compared would match and the comparison would be a fair one.

Grid-search results were used to constrain the prior intervals hence the prior volumes, for all the models being tested using the MultiNest algorithm. The prior intervals were also subdivided for the degenerate parameters d and u_0 into equally spaced subintervals and searched in parallel using MultiNest. In this approach, the evidence values had been artificially boosted by the same factor by which the prior volumes had been reduced. Rather than correcting the evidence values by that factor, by taking certain measures in

the modelling strategy, the need for correcting the evidence values was eliminated and ensured a fair model selection process.

It was specified whether the degenerate parameter d was optimised in the wide or close separation region by appending the letter W or C to the model acronyms respectively. To indicate whether u_0 was optimised in the positive or negative region for that particular model, positive or negative signs were appended respectively, as shown in Tables 5.2 and 6.2.

It was ensured that the prior volumes were reduced by the same factor for all the models being tested. This was done by using the same prior interval for each particular parameter across all the models being tested. It was also ensured that the subintervals were equally spaced. As a result, for all the models, evidence values were boosted by the same factor. Since this factor cancels out while taking the ratio of the evidence values of any two models being compared, ultimately the artificial boosting of the evidence values does not affect our model selection results. This is similar to artificially boosting the evidence values by increasing the value of the constant, "C", in the log-likelihood equation (4.5). In both cases, the artificial boosting does not affect the final results. Therefore, the comparison of models using this strategy remains fair and valid.

All the FSPL models and some of the FSBL and FSBLπ models converged according to the convergence criteria of the MultiNest algorithm as specified, within reasonable time limits, which are detailed in Tables 5.2 and 6.2.

After running the MultiNest algorithms till convergence or till the time limit was reached, parameter estimates and model evidences were extracted, as shown in Chapters 5 and 6, respectively.

For the real event, the most favourable model was used to derive the scaling coefficient in the normalization of data to obtain a χ^2 per degree of freedom (dof) equal to 1. This is common practice in model fitting to observed microlensing data and is done here to follow the same practice as employed in Kains et al. (2013), the paper with which results obtained in this research are compared. In re-scaling the error bars, the simple scaling relation $\sigma_i' = k * \sigma_i$ was used, where σ_i' is the re-scaled error bar of the ith data point and σ_i is the original error bar. The scaling coefficient, k, was determined such that $\chi^2/dof \approx 1$. MultiNest runs were repeated on the data with re-scaled error bars and then analysed as shown in Chapter 6. For the synthetic data, the objective was to recover the true model with the true parameter estimates that were used to generate the synthetic event. The rescaling procedure did not seem necessary for the synthetic event.

Chapter 5

Generation and Analysis of the Synthetic Event

5.1 Generation of the Synthetic Data

The objective of analysing synthetic data was to show that the true model with the true parameters could be recovered. Therefore, the synthetic data generated represents a theoretical light curve with error bars added and does not emulate a complex real microlensing event.

Following Skowron et al. (2009), the synthetic microlensing data was generated by taking the following steps:

1. A binary lens model was arbitrarily chosen with parameter values designated "actual values" as given in Table 5.3. Using these parameter values, a magnification, A_i, versus time, t_i, curve was generated using the dynamic light curve engine. A time range was specified and parameter values were chosen, to simulate a high magnification event as illustrated in Figure 5.1 and with caustic crossing features as illustrated in Figure 5.3.

2. The baseline magnitude in the I band was fixed at $I_0 = 15.0$.

3. The blending fraction (ratio of the lensed source flux to the total blend flux), f_s is sampled randomly from a uniform distribution between 0 and unity, to obtain a different f_s value for each point on the light curve.

4. Given I_0 and f_s, magnification data, A_i, at time, t_i, were converted to magnitudes, I_i, using,

$$I_i = I_0 - 2.5 \log[f_s A_i + (1 - f_s)]. \tag{5.1}$$

TABLE 5.1: Prior Distributions for Model Parameters for the Synthetic Event

Parameter Variable	Interval	Distribution
t_0 (JD)	[7045.0, 7055.0]	Uniform
t_E (day)	$[10^{1.74}, 10^{1.88}]$	Log-Uniform
u_0	[-0.03, -0.01] (Negative)	Uniform
	[0.01, 0.03] (Positive)	Uniform
q	$[10^{-2.3}, 10^{-1}]$	Log-Uniform
d (R_E)	$[10^{0.18}, 10^{0.4}]$ (Wide)	Log-Uniform
	$[10^{-0.3}, 10^{0.18}]$ (Close)	Log-Uniform
α (rad)	[2.5, 3.5]	Uniform
ρ	$[10^{-2.3}, 10^{-2.0}]$	Log-Uniform
$\pi_{E,N}$ (θ_E)	[-0.1, 0.1]	Uniform
$\pi_{E,E}$ (θ_E)	[-0.1, 0.1]	Uniform

This gives a theoretical light curve of magnitude versus time, without errors.

5. By sampling from a Gaussian distribution with mean, 0 and standard deviation, 0.1 and then taking their absolute values, different sized error bars for each magnitude, I_i, were obtained.

5.2 Analysis of the Synthetic Event

The synthetic event was modelled using the modelling strategy as described in Subsection 4.3. Grid search results for the finite-source binary lens model of the synthetic event are presented as a χ^2 map in Figure 5.2. The prior intervals for all the parameters in Table 5.1 were chosen to enclose regions with χ^2 values of less than 700.

The parameter estimates for the most favourable model (FSBLW+) that was selected are shown in Table 5.3. The MAP parameter estimates have been reported, which were used to generate the model light curve and the caustics diagrams as shown in Figure 5.1 and Figure 5.3 respectively. Also reported are the mean and respective uncertainty values. As can be seen in the table, all the parameter estimates agree with the actual values that were used to generate the synthetic event. In Figure 5.4, the probability distribution for each of the parameters of our most favourable model are presented.

Comparing all the models, including the ones that were aborted after a certain time, the most favourable model selected would be the FSBLπW- model which does not correspond to the true model. On the other hand, simultaneously comparing only the models that
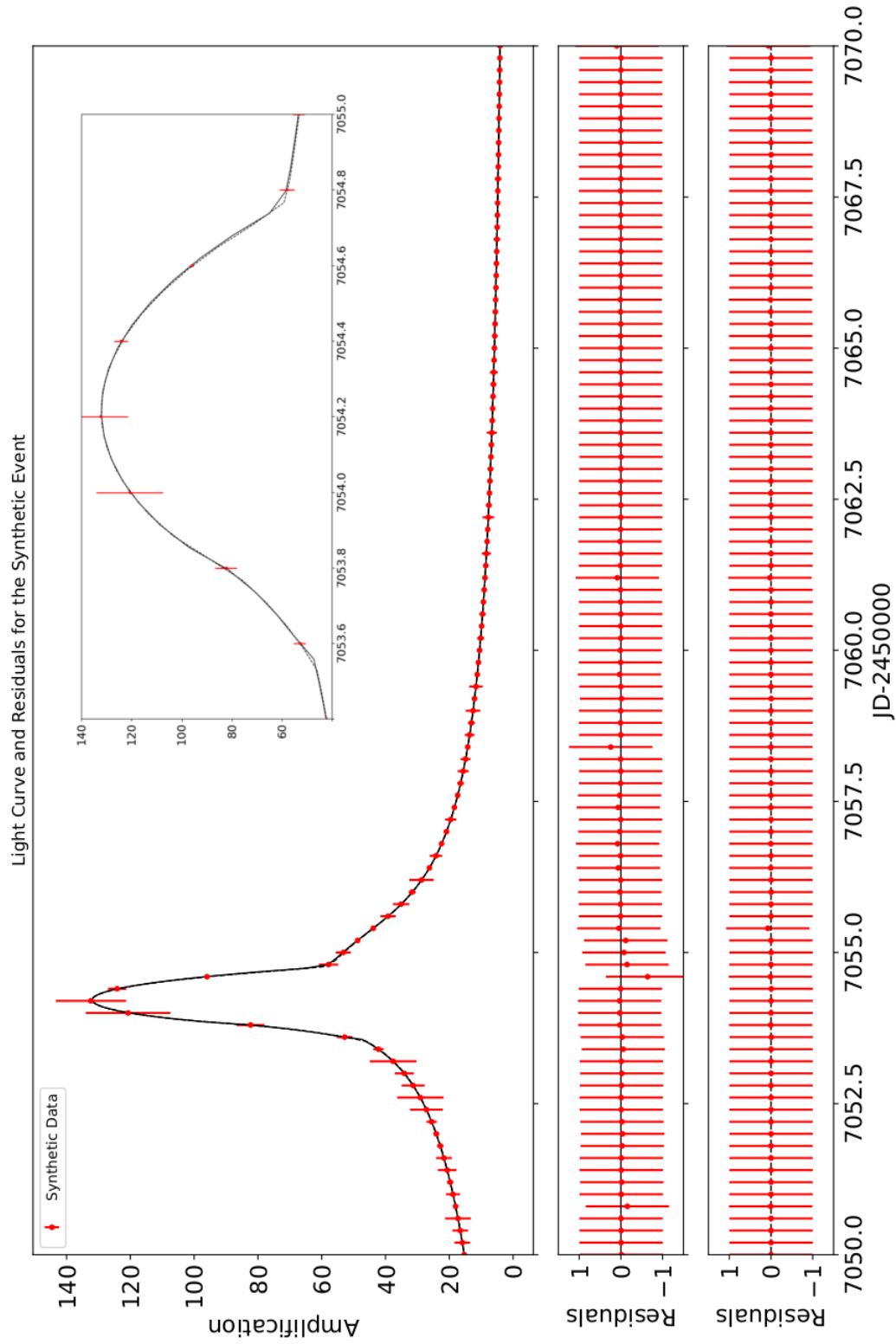
FIGURE 5.1: Light curve for the most favourable model (FSBLW+) for the synthetic event is shown by the solid line. MAP parameter estimates were used to generate this light curve. Light curve representing the true model is shown by the dashed line. The top residual plot is for the most favourable model as found by MultiNest whereas the bottom panel is for the true model.
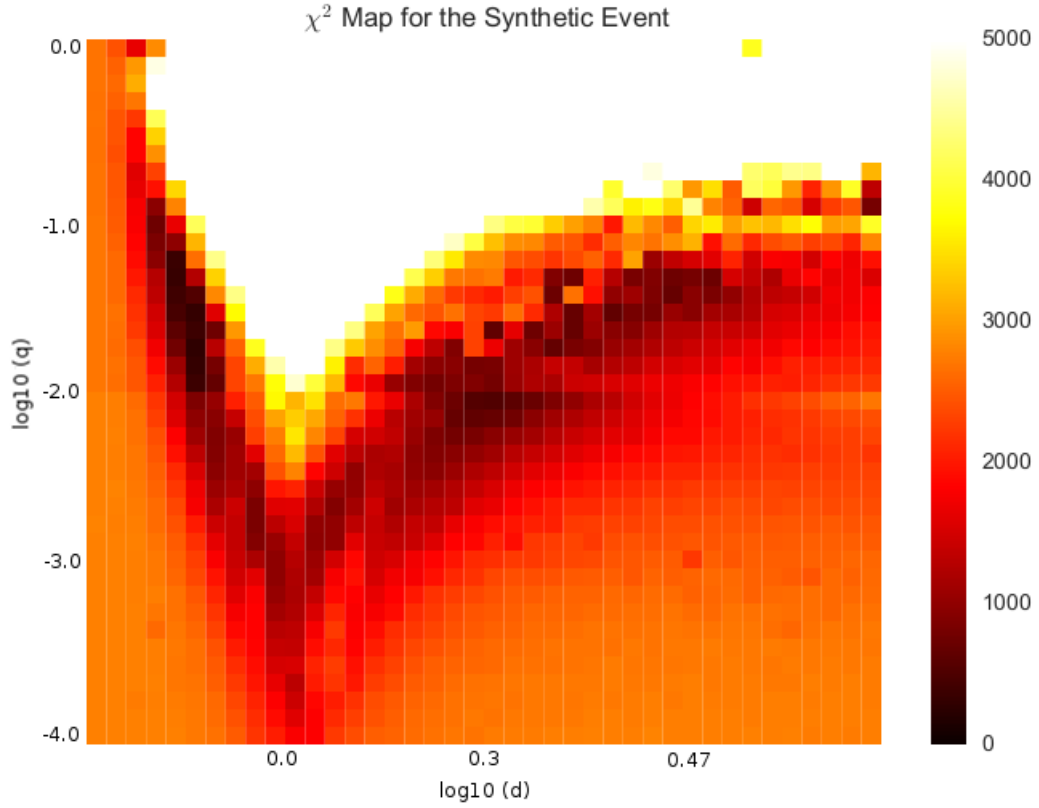
FIGURE 5.2: χ^2 values are indicated by the colour bar, brighter colours indicating a
higher χ^2 value. Log values of the mass fraction extend from -4.0 to 0.0 whereas log
values of separation, extend from -1.0 to 0.6.

had fully converged and choosing the model with the highest log-evidence value, the most
favourable model (FSBLW+) corresponds to the true model that was used to generate
the synthetic event.

Furthermore, the difference in the log-evidence values, $\Delta \ln Z$, between the fully con-
verged model solutions (FSBLW+ , FSBLπW+, FSBLW-) are less than 1.1 and using
Jeffreys' scale, these models would be considered equally probable as well. However,
ultimately the model with the highest log-evidence value out of the fully converged
solutions was selected and this corresponds to the true model as mentioned above.

The log-evidence values and the run time for all the models are shown in Table 5.2.
There is a large difference in the log-evidence values between the binary-lens and single-
lens models, therefore, there is 'decisive' evidence favouring the binary-lens models over
the single-lens models, using Jeffreys' scale.

TABLE 5.2: MultiNest log-evidences ($\ln Z$) for all the models tested for the synthetic event.

Model	$\ln Z$	Convergence Status (Time)
FSBLW+	9970.966 ± 0.355	Fully Converged (14.8 hrs)
FSBLπW+	9970.917 ± 0.356	Fully Converged (18.4 hrs)
FSBLW-	9970.382 ± 0.359	Fully Converged (14.3 hrs)
FSBLπC-	9926.981 ± 0.377	Fully Converged (20.6 hrs)
FSBLπW-	9971.508 ± 0.349	Aborted (24.5 hrs)
FSBLC-	9849.876 ± 0.260	Aborted (24.5 hrs)
FSBLπC+	9823.031 ± 0.264	Aborted (24.5 hrs)
FSBLC+	9806.577 ± 0.261	Aborted (24.5 hrs)
FSPL-	-584611.645 ± 0.375	Fully Converged (11.6 s)
FSPL+	-594080.768 ± 0.376	Fully Converged (74.4 s)

TABLE 5.3: MultiNest Parameter Estimates for the Synthetic Event for the Most Favourable Model (FSBLW+).

Parameter	MAP	Mean $\pm$ Uncertainty	Actual
t_0 (JD)	7050.081	7050.012 ± 0.168	7050.0
t_E (days)	65.511	65.321 ± 0.660	65.4
u_0	0.020	0.020 ± 0.001	0.02
q	0.032	0.033 ± 0.001	0.033
d	2.392	2.399 ± 0.017	2.4
α (rad)	3.000	3.000 ± 0.003	3.0
ρ	0.007	0.007 ± 0.0002	0.007
$\pi_{E,N}$	0.0	0.0 ± 0.0	0.0
$\pi_{E,E}$	0.0	0.0 ± 0.0	0.0
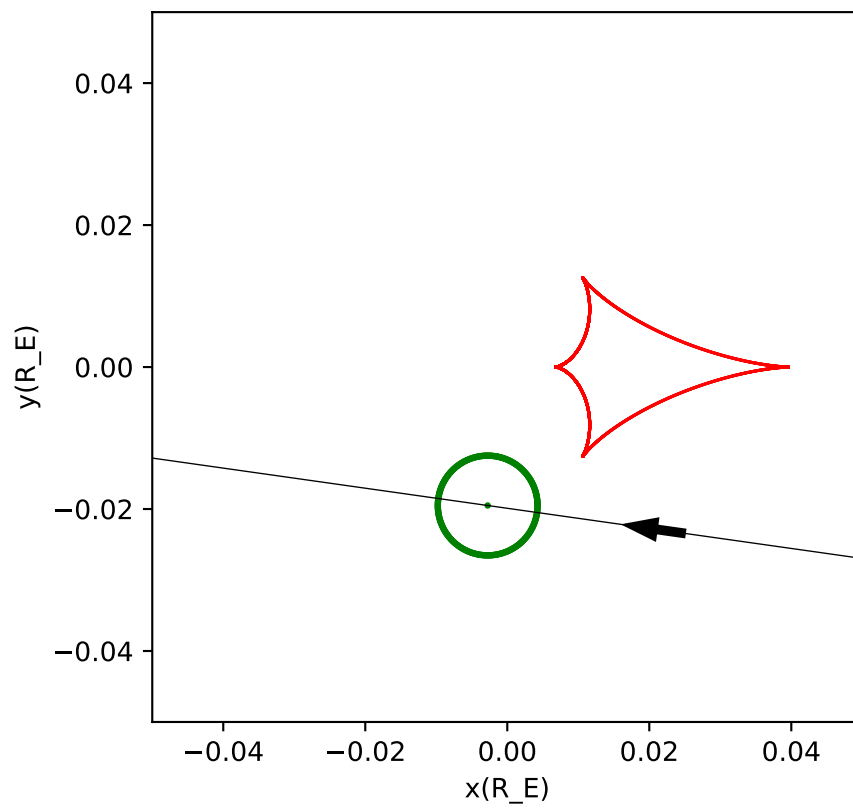
FIGURE 5.3: Caustic curves for the most favourable model (FSBLW+) for the synthetic event, with the source trajectory in the direction indicated by the arrow and the source star size indicated by the solid circle on the trajectory.
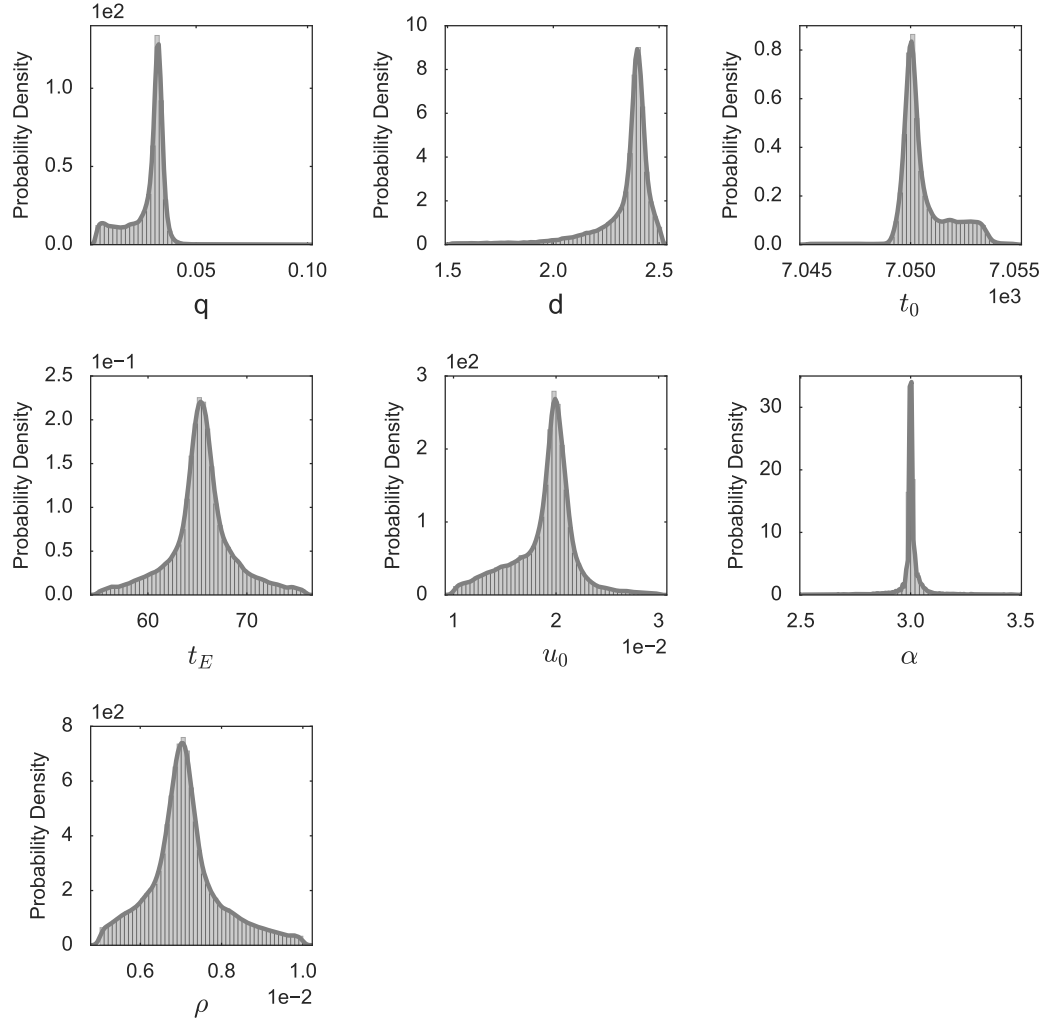
FIGURE 5.4: Grey bars show the probability distribution of all the parameters of the most favourable model (FSBLW+) . The histograms have been normalized so that the area under the bars and the gray curve representing the kernel density estimate totals to one.

Chapter 6

Analysis of the Real Event OB11-0251

In demonstrating the methodology developed in this research, on a real event, OB11-0251, the results obtained using this methodology were compared with the results published by Kains et al. (2013). The published results show OB11-0251 to be a giant planet beyond the snow line with a mass ratio of $q = 1.92 \pm 0.12$ and separation of $d = 1.408 \pm 0.019$.

The modelling strategy as described in Subsection 4.3 was used to model the real event. Grid search results for OB11-0251 are presented as a χ^2 map in Figure 6.1. The prior intervals for all the parameters in Table 6.1 were chosen to enclose regions with χ^2 values of less than 6×10^4.

TABLE 6.1: Prior Distributions for Model Parameters for OB11-0251

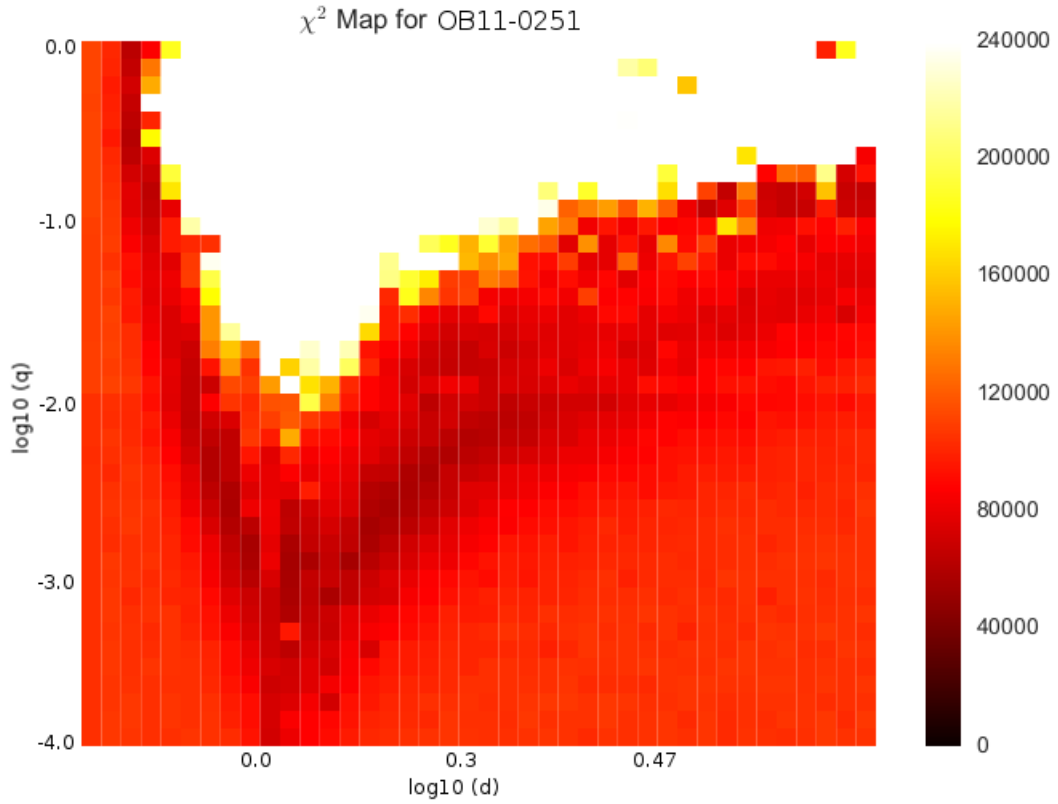| Parameter	Interval	Distribution
t_0 (JD)	[5781.0, 5782.0]	Uniform
t_E (day)	$[10^{1.77}, 10^{1.85}]$	Log-Uniform
u_0	[-0.07, -0.03] (Negative)	Uniform
	[0.03, 0.07] (Positive)	Uniform
q	$[10^{-3}, 10^{-2}]$	Log-uniform
d (R_E)	$[10^{0.18}, 10^{0.4}]$ (Wide)	Log-Uniform
	$[10^{-0.3}, 10^{0.18}]$ (Close)	Log-Uniform
α (rad)	[1.6, 4.6]	Uniform
ρ	$[10^{-2}, 10^{-1.5}]$	Log-uniform
$\pi_{E,N}$ (θ_E)	[-0.1, 0.1]	Uniform
$\pi_{E,E}$ (θ_E)	[-0.1, 0.1]	Uniform

FIGURE 6.1:   This χ^2 map is a visual representation of the grid search results for the event OB11-0251. χ^2 values are indicated by the colour gradient, brighter colours indicating a higher χ^2 value. Log values of the mass fraction extend from -4.0 to 0.0 whereas log values of separation, extend from -1.0 to 0.6.

MultiNest was run on OB11-0251 raw data and then on rescaled data with a time limit of 10 days. Some of the model solutions obtained from the run on the rescaled data achieved full convergence.

Table 6.2 shows all the models tested along with their log-evidence and $\chi^2/$ dof values. As mentioned in the introduction to Bayesian inference, all the models tested can be simply compared simultaneously and the model with the highest log-evidence value can be chosen as the most favourable model or the Jeffreys' scale can also be referred to compare models.

Comparing all the models, including the ones that were aborted due to the time limit, FSBLC- is obtained as the most favourable model, which does not correspond to the published results by Kains et al. (2013). Comparing only the models that had fully converged and choosing the model with the highest log-evidence value, FSBLW+ is obtained as the most favourable model which does not correspond to results by Kains et al. (2013) either.

TABLE 6.2: MultiNest log-evidences ($\ln Z$) for the models tested for OB11-0251. In brackets it is specified whether the degenerate parameter d was searched in the wide or close separation region and whether u_0 was searched in the positive or negative region for that particular model.

Model	$\ln Z$	Convergence Status (Time)
FSBLW+	8059.292 ± 0.363	Fully Converged (7.8 days)
FSBLπW+	8058.900 ± 0.374	Fully Converged (8.6 days)
FSBLW-	8058.600 ± 0.367	Fully Converged (7.5 days)
FSBLC-	8099.526 ± 0.350	Aborted (10 days)
FSBLπC-	8099.155 ± 0.376	Aborted (10 days)
FSBLπW-	8059.082 ± 0.371	Aborted (10 days)
FSBLπC+	8007.104 ± 0.281	Aborted (10 days)
FSBLC+	8001.962 ± 0.243	Aborted (10 days)
FSPL-	-1703713.603 ± 0.375	Fully Converged (173 s)
FSPL+	-1704157.617 ± 0.377	Fully Converged (286 s)

However, when a pairwise comparison is made between the models using Jeffreys' scale, the difference between FSBLW+ and FSBLπW+ and between FSBLC- and FSBLπC- is found to be less than 1.1, and it can be inferred that the two pairs of models are equally probable. None of these models correspond directly to the model solution published by Kains et al. (2013), which is a FSBLπW- model.

The most favourable models obtained from the results are FSBLW+ and FSBLπW+ which are equally probable. However, the model with parallax effects, FSBLπW+ was selected as the representative most favourable model, to compare the parameter estimates of this model solution with those of Kains et al. (2013).

Table 6.3 shows the parameter estimates for the selected model for OB11-0251. The MAP parameter estimates were used to generate the model light curve and the caustics diagram as shown in Figure 6.2 and Figure 6.3 respectively. The table also shows the mean parameter estimates and the respective uncertainty values.

In comparing the parameter estimates with those reported by Kains et al. (2013), the parameter differences in sigma units were calculated. As represented in the table 6.3, the parameter α was converted to a negative angle equivalent to the positive value which was originally obtained.

In Figure 6.4, the probability distribution for each of the parameters of the selected model is presented.
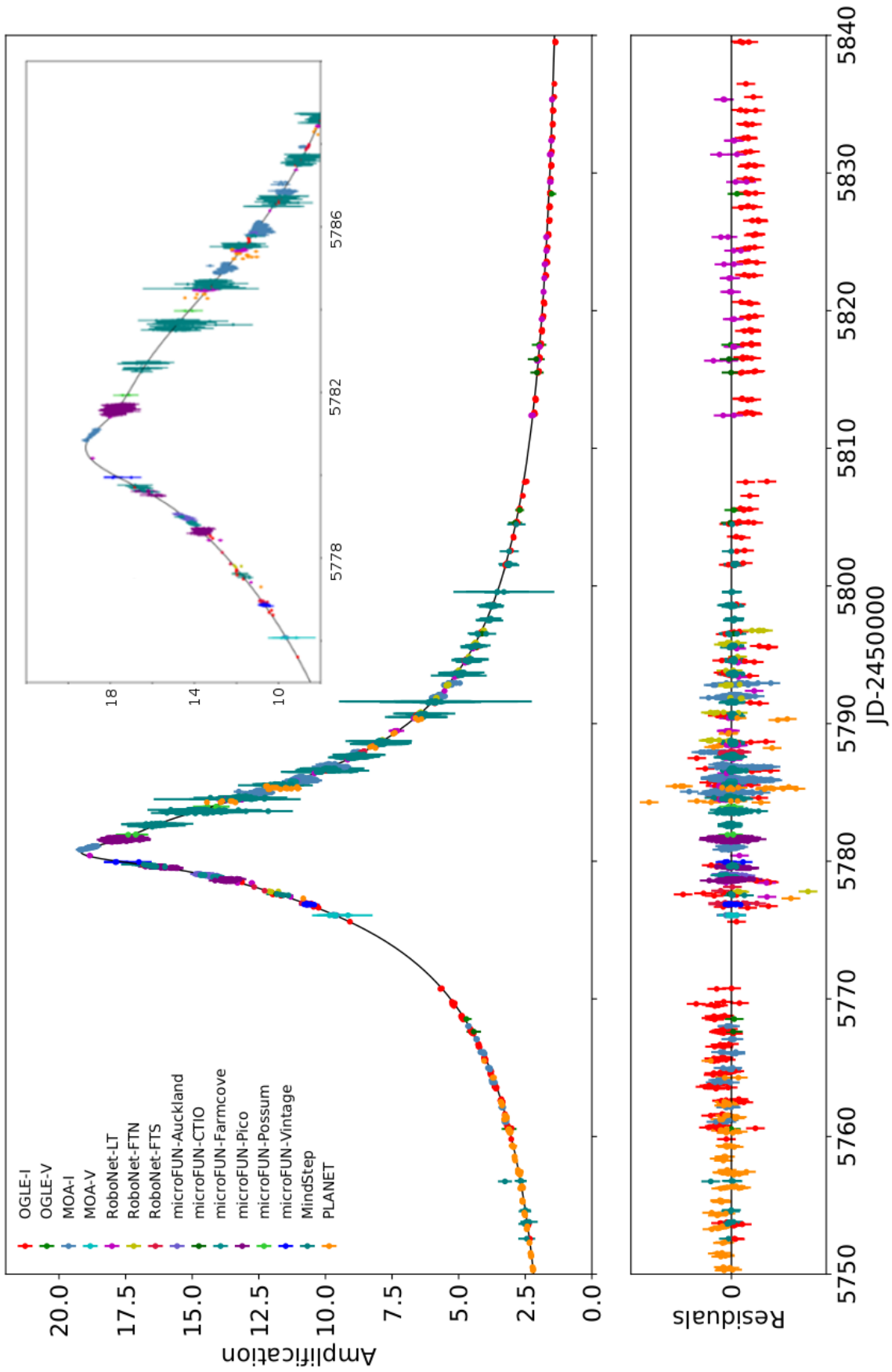
FIGURE 6.2: Light curve obtained for the most favourable model found for the event OB11-0251 – the wide separation finite-source binary-lens model with parallax. Inset zooms in on the perturbation region of the light curve.

TABLE 6.3: MultiNest Parameter Estimates for OB11-0251

| Parameter	MAP	Mean $\pm$ Uncertainties	Parameter Estimates (Kains et al., 2013)	Difference (σ)	Ratio of Error Bars
t_0 (MHJD)	5781.128	5781.374 ± 0.002	5781.503 ± 0.004	28.84	0.5
t_E (days)	63.47	63.41 ± 0.26	63.88 ± 0.46	0.89	0.56
$u_0/10^{-2}$	5.44	5.44 ± 0.03	-5.63 ± 0.04	3.80	0.75
$q/10^{-3}$	2.50	2.54 ± 0.04	1.92 ± 0.12	4.90	0.33
d	1.51	1.52 ± 0.003	1.408 ± 0.019	5.82	0.16
α (rad)	4.44	4.45 ± 0.01	-1.855 ± 0.002	4.41	5.0
$\rho/10^{-2}$	1.160	1.131 ± 0.07	1.39 ± 0.05	3.01	1.4
$\pi_{E,E}$	-0.09	-0.04 ± 0.05	0.09 ± 0.01	2.54	5.0
$\pi_{E,N}$	0.007	0.02 ± 0.01	-0.34 ± 0.05	7.06	0.2
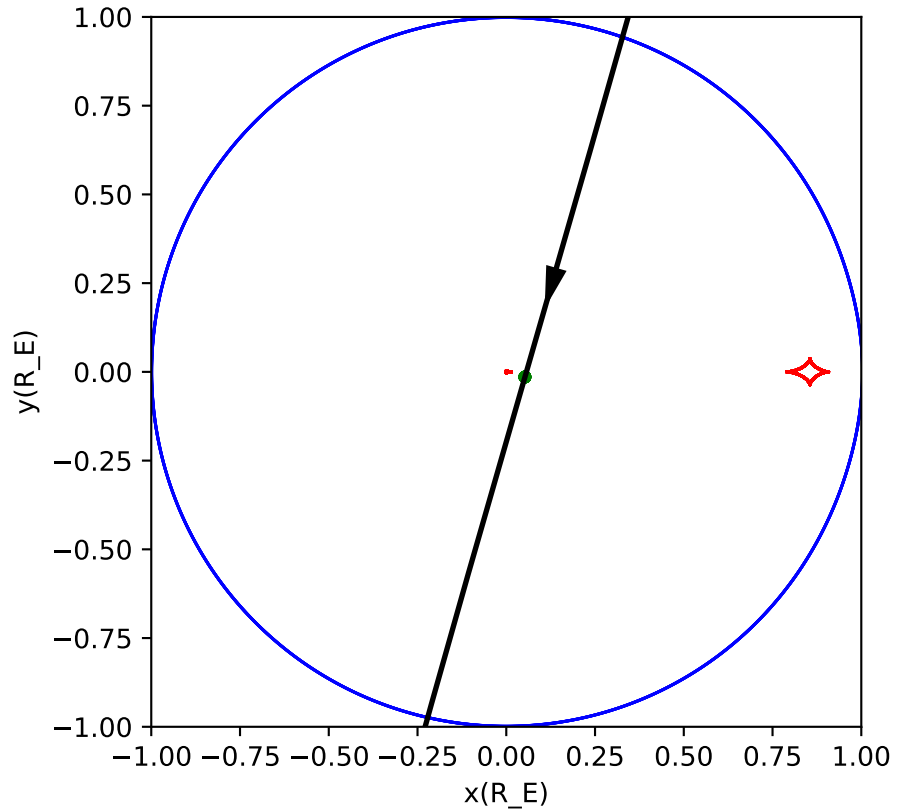
FIGURE 6.3: Caustics pattern for the event OB11-0251 shown as the diamond shaped curves and the critical curve is represented by the circle. The source star trajectory is drawn across the plane with solid black line, the arrow indicating its direction and the small circle drawn upon it indicating the finite size of the source star.
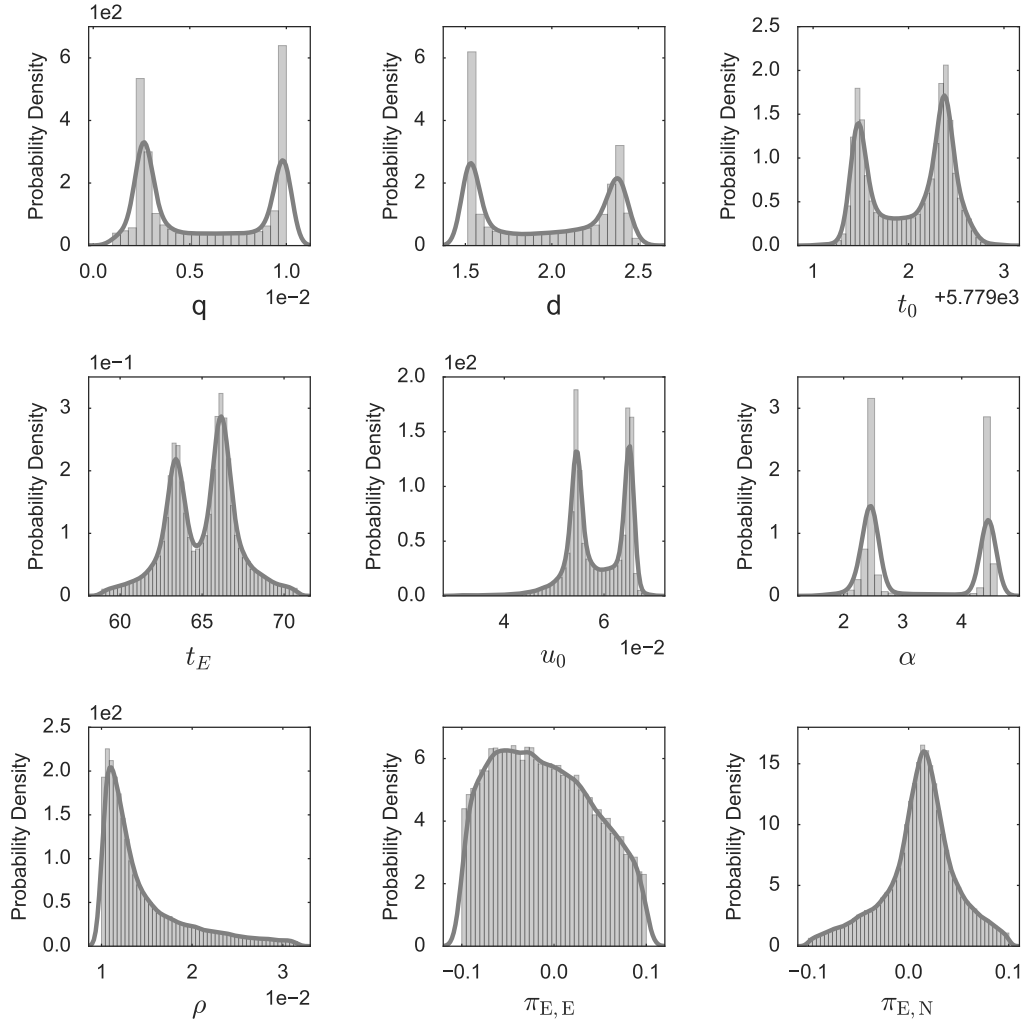
FIGURE 6.4: Grey bars show the probability distribution of all the parameters of the most favourable model. The histograms have been normalized so that the area under the bars and the grey kernel density estimate curve totals to one.

Chapter 7

Discussion and Conclusion

A new modelling strategy for microlensing modelling has been developed in this project, which offers a different approach to the model selection problem in particular. This methodology is based on Bayesian principles and allows us to compare alternative models against the same data set and to make a quantitative preference for one model out of two or more competing models. This methodology can also prevent implausible values for the physical parameters entering the scenario by informing the priors with knowledge obtained from Galactic models, for example.

During the research, prior intervals that reflected the entire range of possible parameter values were used during the preliminary MultiNest runs but they had to be aborted since the run-time for these runs were longer than affordable, on the order of several days to weeks, while the data obtained during the run suggested that there was difficulty obtaining convergence or viable solutions. These preliminary runs were done on several real events and proved to be challenging especially for the caustic crossing events due to the aforementioned reasons.

The exception to this were the FSPL models, for which it was feasible to assign the entire range of possible parameter values and still achieve convergence with a runtime on the order of minutes, as explained in Section 4.3. This was possible for the FSPL models since the magnification map technique as described in Section 4.1 could be used to generate light curves for these simpler models. For each FSPL model only one magnification map with $q = 0$ and $d = 0$ would have to be created and then thousands of light curves could be generated from this magnification map within seconds to minutes, enabling a fast convergence.

However, for the binary-lens models the magnification map technique could not be used since q and d were to be optimised as free parameters. Therefore, the dynamic light curve

engine technique had to be used. While it can generate a light curve without having to generate a magnification map, it is considerably slower in generating thousands of light curves.

Narrower prior intervals had to be carefully chosen for the binary lens models to keep computer time within affordable limits. The non-Bayesian method of a grid search coupled with downhill simplex optimization were included in our modelling process to achieve this. It provided a bird's-eye view of the χ^2 landscape to locate the regions of the parameter space to be explored further with MultiNest, essentially narrowing down the prior intervals.

In fact, despite narrowing the prior intervals, the MultiNest runs for some of the models tested still had to be aborted as shown in Tables 5.2 and 6.2, due to the relatively longer run-times compared to the other models being tested in parallel.

The methodology was demonstrated using synthetic data as well as data for a real microlensing event, OB11-0251. To demonstrate the capability of the methodology to perform model selection, various models with different numbers of parameters, including FSPL and FSBL models, were tested. The presence or absence of the parallax effect in the binary-lens models were also tested. There is usually a degeneracy between the wide and close separation solutions, and between positive and negative impact parameter, u_0 solutions. Therefore, models with various degenerate scenarios were tested as well.

For the synthetic event, the correct model with accurate parameter estimates, was recovered. However, for OB11-0251, the model preference resulting from the methodology used in this project did not agree with that of the benchmark, (Kains et al., 2013). Furthermore, the parameter estimates recovered by the methodology used in this research project, roughly matched the previously published results but did not match them precisely.

Therefore, it is inferred that for complex, real microlensing events, especially caustic crossing events, the computation resources available may not be sufficient to successfully use the methodology that was tested in this project. However there is potential for further refining this methodology and using it as a definitive model selection methodology, demonstrated as a proof of concept by the synthetic event modelling results.

An improvement would be to use Bayesian priors specifically derived for the caustic crossing binary-lens parameters as discussed in Cassan et al. (2010) instead of using uniform priors as was done in this project.

Given computational resources, another improvement would be to choose the entire range of possible parameter values as the prior intervals. In other words, to essentially

remove the need for a grid search coupled with the downhill simplex method or similar non-Bayesian optimization method as an initial step to narrow down the prior intervals to be searched using MultiNest. This would considerably simplify the strategy and make it a stand-alone Bayesian optimisation methodology rather than a hybrid of Bayesian and non-Bayesian optimisation methods.

Bibliography

Bennett, D.P., Bond, I.A., Udalski, A., Sumi, T., Abe, F., Fukui, A., Furusawa, K., Hearnshaw, J.B., et al., 2008, The Astrophysical Journal, 684, 1, 663, URL `http://iopscience.iop.org/0004-637X/684/1/663`

Bennett, D.P., 2008, in Exoplanets, Springer, 47–88, URL `http://link.springer.com/chapter/10.1007/978-3-540-74008-7_3`

Brewer, B.J., Pártay, L.B., Csányi, G., 2011, Statistics and Computing, 21, 4, 649, URL `http://link.springer.com/article/10.1007/s11222-010-9198-8`

Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., Merloni, A., Salvato, M., et al., 2014, Astronomy & Astrophysics, 564, A125, URL `http://www.aanda.org/10.1051/0004-6361/201322971`

Cassan, A., Horne, K., Kains, N., Tsapras, Y., Browne, P., 2010, Astronomy and Astrophysics, 515, A52, URL `http://adsabs.harvard.edu/abs/2010A%26A...515A..52C`

Charbonneau, P., 1995, The Astrophysical Journal Supplement Series, 101, 309, URL `http://adsabs.harvard.edu/abs/1995ApJS..101..309C`

Dominik, M., 2010, General Relativity and Gravitation, 42, 9, 2075, URL `http://link.springer.com/article/10.1007/s10714-010-0930-7`

Einstein, A., 1936, Science, 84, 2188, 506, URL `http://science.sciencemag.org/content/84/2188/506`

Feroz, F., Balan, S.T., Hobson, M.P., 2011, Monthly Notices of the Royal Astronomical Society, 415, 4, 3462, URL `http://mnras.oxfordjournals.org/content/415/4/3462`

Feroz, F., Hobson, M.P., Bridges, M., 2009, Monthly Notices of the Royal Astronomical Society, 398, 4, 1601, URL `http://mnras.oxfordjournals.org/content/398/4/1601`

Feroz, F., Hobson, M.P., Cameron, E., Pettitt, A.N., 2013, arXiv preprint arXiv:1306.2144, URL `http://arxiv.org/abs/1306.2144`

41

Feroz, F., Hobson, M.P., 2008, Monthly Notices of the Royal Astronomical Society, 384, 2, 449, URL `http://mnras.oxfordjournals.org/content/384/2/449.short`

Ford, E.B., Gregory, P.C., 2007, volume 371, eprint: arXiv:astro-ph/0608328, 189, URL `http://adsabs.harvard.edu/abs/2007ASPC..371..189F`

Gaudi, B.S., 2010, arXiv:1002.0332 [astro-ph], URL `http://arxiv.org/abs/1002.0332`, arXiv: 1002.0332

Gaudi, B.S., Chang, H.Y., Han, C., 2003, The Astrophysical Journal, 586, 1, 527, URL `http://iopscience.iop.org/article/10.1086/367539/meta`

Gould, A., 2000, The Astrophysical Journal, 542, 2, 785, URL `http://stacks.iop.org/0004-637X/542/i=2/a=785`

Gould, A., 2004, The Astrophysical Journal, 606, 1, 319, URL `http://stacks.iop.org/0004-637X/606/i=1/a=319`

Gould, A., Loeb, A., 1992, The Astrophysical Journal, 396, 1, 104, URL `http://cat.inist.fr/?aModele=afficheN&cpsidt=5493357`

Gregory, P.C., 2005, Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support, volume -1, URL `http://adsabs.harvard.edu/abs/2005blda.book.....G`

Jaroszyñski, M., Skowron, J., Udalski, A., Kubiak, M., Szymañski, M.K., Pietrzyñski, G., Soszyñski, I., Wyrzykowski, L., et al., 2010, Acta Astronomica, 60, 197, URL `http://adsabs.harvard.edu/abs/2010AcA....60..197J`

Jeffreys, H., 1961, The Theory of Probability, OUP Oxford, google-Books-ID: vh9Act9rtzQC

Kains, N., Browne, P., Horne, K., Hundertmark, M., Cassan, A., 2012, Monthly Notices of the Royal Astronomical Society, 426, 2228, URL `http://adsabs.harvard.edu/abs/2012MNRAS.426.2228K`

Kains, N., Cassan, A., Horne, K., Albrow, M.D., Dieters, S., Fouqué, P., Greenhill, J., Udalski, A., et al., 2009, Monthly Notices of the Royal Astronomical Society, 395, 2, 787, URL `http://mnras.oxfordjournals.org/content/395/2/787`

Kains, N., Street, R.A., Choi, J.Y., Han, C., Udalski, A., Almeida, L.A., Jablonski, F., Tristram, P.J., et al., 2013, Astronomy & Astrophysics, 552, A70, URL `http://www.aanda.org/10.1051/0004-6361/201220626`

Kubas, D., Cassan, A., Beaulieu, J.P., Coutures, C., Dominik, M., Albrow, M.D., Brillant, S., Caldwell, J.A.R., et al., 2005, Astronomy and Astrophysics, 435, 3, 941, URL `http://arxiv.org/abs/astro-ph/0502018`, arXiv: astro-ph/0502018

Kubas, D., 2005, Applications of Galactic Microlensing, Ph.D. thesis, University of Potsdam, URL `https://www.deutsche-digitale-bibliothek.de/binary/5CWRJERPXNYUIXCM4Q2YNOW336HYRTP4/full/1.pdf`

Ling, C., 2013, Simulation and modelling of gravitational microlensing events using graphical processing units, Ph.D. thesis, URL `http://muir.massey.ac.nz/handle/10179/5148`

Mao, S., Paczynski, B., 1991, The Astrophysical Journal Letters, 374, L37, URL `http://adsabs.harvard.edu/abs/1991ApJ...374L..37M`

McDougall, 2014, Gravitational Microlensing: An automated high-performance modelling system, Ph.D. thesis, URL `http://ir.canterbury.ac.nz/handle/10092/9973`

Mukherjee, P., Parkinson, D., Liddle, A.R., 2006, The Astrophysical Journal, 638, 2, L51, URL `http://arxiv.org/abs/astro-ph/0508461`, arXiv: astro-ph/0508461

Paczynski, B., 1986, The Astrophysical Journal, 304, 1, URL `http://adsabs.harvard.edu/abs/1986ApJ...304....1P`

Placek, B., Knuth, K.H., Angerhausen, D., 2014, The Astrophysical Journal, 795, 2, 112, URL `http://stacks.iop.org/0004-637X/795/i=2/a=112?key=crossref.0b6043a1fd369f93678369045384987d`

Press, W.H., 2007, Numerical Recipes 3rd Edition: The Art of Scientific Computing, Cambridge University Press, google-Books-ID: 1aAOdzK3FegC

Rattenbury, N.J., 2006, Modern Physics Letters A, 21, 12, 919, URL `http://www.worldscientific.com/doi/abs/10.1142/S0217732306020470`

Sivia, D.S., Skilling, J., 2006, Data Analysis : A Bayesian Tutorial, Oxford University Press, UK, Oxford, GBR, URL `http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10229918`

Skilling, J., 2004, in AIP Conference Proceedings, volume 735, AIP Publishing, 395–405, URL `http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.1835238`

Skilling, J., 2006, Bayesian Analysis, 1, 4, 833, URL `https://projecteuclid.org/euclid.ba/1340370944`

Skowron, J., Wyrzykowski, L., Mao, S., Jaroszyński, M., 2009, Monthly Notices of the Royal Astronomical Society, 393, 3, 999, URL `http://mnras.oxfordjournals.org/content/393/3/999`

Thompson, A.C., Fluke, C.J., Barnes, D.G., Barsdell, B.R., 2010, New Astronomy, 15, 1, 16, URL `http://arxiv.org/abs/0905.2453`, arXiv: 0905.2453

Trotta, R., 2008, Contemporary Physics, 49, 2, 71, URL `http://www.tandfonline.com/doi/abs/10.1080/00107510802066753`

Wambsganss, J., Witt, H.J., Schneider, P., 1992, Astronomy and Astrophysics, 258, 591, URL `http://adsabs.harvard.edu/abs/1992A%26A...258..591W`

Yee, J.C., Udalski, A., Dong, S., Greenhill, J., Tsapras, Y., Bond, I.A., Gould, A., Kozlowski, S., et al., 2011, OGLE-2008-BLG-513lb: The Orbital Solution for a Microlensing Planet, URL `http://adsabs.harvard.edu/abs/2011arXiv1106.4013Y`